IMT Institute for Advanced Studies, Lucca

Lucca, Italy

Essays in Applied Environmental and Innovation Economics

PhD Program in Economics, Markets, Institutions

XXIV Cycle

By

Giovanni Marin

2012

The dissertation of Giovanni Marin is approved.

Program Coordinator: Prof. Fabio Pammolli, IMT Advanced Studies Lucca

Supervisor: Dott. Francesca Lotti, Banca d'Italia

Supervisor: Prof. Massimiliano Mazzanti, Università di Ferrara

Tutor: Dott. Andrea Giannaccari, IMT Advanced Studies Lucca

The dissertation of Giovanni Marin has been reviewed by:

Prof. Massimo Riccaboni, IMT Advanced Studies Lucca

Prof. Carlo Cambini, Politecnico di Torino

IMT Institute for Advanced Studies, Lucca

2012

Contents

List of Figures

List of Tables

Acknowledgements

Chapter 2 is an extended version of the book chapter titled "Closing the Gap? Dynamic Analyses of Emission Efficiency and Sector Productivity in Europe" which is included in the book "The Dynamics of Economic and Environmental Efficiency" edited by Massimiliano Mazzanti and Valeria Costantini (Springer, 2012)

Chapter 3 is the reproduction of the article published on *Ecological Economics* (2012, Vol. 74, pp. 71-84) with the title "Linking NAMEA and Input Output for 'Consumption vs. Production Perspective' Analyses - Evidence on Emission Efficiency and Aggregation Biases using the Italian and Spanish Environmental Accounts" and co-authored with Massimiliano Mazzanti and Anna Montini.

First of all, I want to thank my two great supervisors: Francesca Lotti (Banca d'Italia) and Massimiliano Mazzanti (Università di Ferrara). They have been (and are) two endless sources of ideas, suggestions, stimuli and support.

I am also indebted to Roberto Zoboli (Ceris-CNR and Università Cattolica del Sacro Cuore di Milano) and Anna Montini (Università di Bologna) for their general support and comments and their crucial contribution to the article reported in chapter 3 (especially Anna).

I am really grateful to Bronwyn Huges Hall (University of California, Berkeley) who gave me the possibility to spend some months of my life in the beautiful city of Berkeley and in the great campus of the Univerisity of California at Berkeley.

Many thanks also to all other colleagues and friends who helped, supported, criticized and stimulated my work: Emmanuele Somma (Banca d'Italia), Simone Borghesi (Università di Siena), Susanna Mancinelli (Università di Ferrara), Valeria Costantini (Università di Roma Tre), Alessio d'Amato (Università di Roma Tor Vergata), Antonio Dalla Malva (Katholieke Universiteit Leuven), Ivan Faiella (Banca d'Italia).

Finally, many thanks to all the IMT community for the good time I had in Lucca.

Vita

13 February 1984	Born, Ferrara (FE), Italy
2006	Bachelor in Economics (Laurea triennale) Final mark: 110/110 cum laude Università di Ferrara, Ferrara (Italy)
2008	MA in Economics (Laurea magistrale) Final mark: 110/110 cum laude Università di Ferrara, Ferrara (Italy)
2011	Visiting Research Scholar (February-July) Department of Economics University of California, Berkeley invited by Professor Bronwyn Huges Hall
2011	Consultant at CERIS-CNR (Milano) European Consortium 'European Topic Centre on Sustainable Production and Consumption'

Publications

1. G. Marin, M. Mazzanti, “The dynamics of delinking in industrial emissions: The role of productivity, trade and R&D” *Journal of Innovation Economics*, 0(1):91-117, 2009
2. G. Marin, M. Mazzanti, “Emissions Trends and Labour Productivity Dynamics”, in (eds Mazzanti M., Montini A.) Environmental Efficiency, Innovation and Economic Performances, Routledge, 2010
3. G. Marin, M. Mazzanti, “The Evolution of Environmental and Labor Productivity Dynamics - Sector analyses and Decoupling / Recoupling trends on a 1990-2007 Italian NAMEA” *Journal of Evolutionary Economics*, in press (accepted on 20 October 2010)
4. G. Marin, M. Mazzanti, A. Montini “Aggregation Bias in ‘Consumption vs Production Perspective’ Comparisons - Evidence Using the Italian and Spanish NAMEAs” (with Massimiliano Mazzanti and Anna Montini), in (eds V. Costantini, M. Mazzanti and A. Montini) Hybrid Economic and Environmental Accounts, Routledge, 2012
5. G. Marin, M. Mazzanti, A. Montini “Linking NAMEA and Input Output for ‘Consumption vs. Production Perspective’ Analyses - Evidence on Emission Efficiency and Aggregation Biases using the Italian and Spanish Environmental Accounts” *Ecological Economics*, 74:71-84, 2012
6. G. Marin “Closing the Gap? Dynamic Analyses of Emission Efficiency and Sector Productivity in Europe”, forthcoming on (eds V. Costantini, M. Mazzanti) The Dynamics of Economic and Environmental Efficiency, Springer, 2012

Presentations

1. G. Marin (with M. Mazzanti) “The Evolution of Environmental and Labor Productivity Dynamics - Sector analyses and Decoupling / Recoupling trends on a 1990-2007 Italian NAMEA” at:
   - EA-SDI EMAN Conference 2009 (Prague, April 2009)
   - Druid Summer Conference 2009 (Copenhagen, June 2009)
   - 50th Riunione Annuale della Societ Italiana degli Economisti (Roma, October 2009)
   - 3rd International Conference on Eco-Efficiency (Egmond aan Zee, The Netherlands, June 2010)
   - 11th Biennial Conference of the International Society of Ecological Economics (Oldenburg and Bremen, Germany, August 2010)
2. G. Marin “Sector CO2 and SOx Emissions Efficiency and Investment: Homogeneous vs Heterogeneous Estimates using the Italian NAMEA” at the 11th Biennial Conference of the International Society of Ecological Economics (Oldenburg and Bremen, Germany, August 2010)
3. G. Marin (with G. Cainelli and M. Mazzanti) “ICT investments, Eco-Innovations and Environmental Efficiency - Micro and sector studies from Italy” at the Workshop ‘The Economics of Green IT’, Centre for European Economic Research (Mannheim, Germany, 22 November 2010)
4. G. Marin (with M. Mazzanti and A. Montini) “Linking NAMEA and Input Output for ‘Consumption vs. Production Perspective’ Analyses” at the International Scientific Conference ‘Sustainable Consumption - Towards Action and Impact’ (Hamburg, Germany, 6-8 November 2011)
5. G. Marin “Do Eco-Innovations Harm Productivity Growth through Crowding Out? Results of an Extended CDM Model for Italy” at:
   - European Seminar EuroLIO 2012 ’Geography of Innovation’ (Saint-Etienne, France, 26-28 January 2012)
   - Workshop ‘Presente e Futuro dell’Economia dell’Ambiente e delle Risorse Naturali in Italia’ (Università di Roma Tor Vergata, February 2012)
   - Brown Bag Seminar IMT Lucca (Lucca, March 2012)
   - Lunch Seminar Banca d’Italia (Roma, March 2012)
   - Lunch seminar, Facoltà di Economia, Università di Ferrara (Ferrara, May 2012)

Abstract

Changes in consumption patterns, technological change and environmental innovations are essential to achieve the reduction of environmental externalities, to slow down the exploitation of natural resources and to reduce the likelihood of environmental disasters (e.g. climate change). The current dissertation touches some of these aspects by performing empirical applications in the fields of innovation and environmental economics.

The dissertation is composed by two main blocks. The first block (chapters 2 and 3) employs a sector-level approach to investigate the patterns of emission efficiency in EU countries (chapter 2) and the extent to which aggregation bias affects the estimate of the amount of emissions induced by final consumption activities by means of environmentally extended input-output models (chapter 3). The second block (chapter 4) relies on firm-level data with the aim of investigating the drivers of environmental innovation activities of firms and their effect on firm-level productivity. Finally, appendix A describes in detail the methodology I used to match firms in AIDA to patent applicants in the PATSTAT database. These data have been used in chapter 4.

Chapter 2

Chapter 2 investigates the patterns of emission efficiency (value added per emission) growth of 23 manufacturing sectors in 12 European countries with a focus on five emissions (CO2, NOx, NMVOC, SOx and CO). Emission efficiency growth is expected to be triggered by improvements in the

efficiency of frontier countries through the diffusion of better technologies to laggard countries. This effect is likely to differ according to the distance from the frontier country. Finally, the role of productivity patterns (Total Factor Productivity) and energy prices dynamics is assessed.

Results based on the European NAMEA (National Accounting Matrix including Environmental Accounts) further merged with sector accounts highlight significant spillovers from leaders in emission efficiency and a general tendency to converge for laggard countries and sectors (except for NMVOC emission efficiency). Energy prices weakly induce improvements in emission efficiency, with the effect being generally stronger for sectors and countries farther away from the emission efficiency frontier. Finally, total factor productivity (TFP) is strongly correlated with emission efficiency while the distance from TFP frontier significantly harms emission efficiency growth.

Chapter 3 (co-authored with Massimiliano Mazzanti and Anna Montini)

We integrate input-output and NAMEA (National Accounting Matrix including Environmental Accounts) tables for Spain and Italy in 1995, 2000 and 2005, in order to address the hot policy issue of sustainable consumption and production. A comparison of production and consumption perspectives may have relevant policy implications. We deal with the domestic technology assumption and primarily the aggregation bias that may result when calculating indirect emission using different sector aggregations in the analyses (e.g. 16, 30, 50). Environmentally Extended Input-Output Analysis (EE-IOA) provides analyses of the emissions embodied in domestic consumption and domestic production by considering the structure of intermediate inputs and environmental efficiency in each production sector.

Our empirical findings show that different sectoral aggregation significantly biases the amount of emissions for the consumption perspective, though differently in the two countries. Italy surprisingly shows consumption/production ratios around or lower than one, but in line with some major work at EU level. Our results thus suggest that special attention must be paid when interpreting the EE-IOA of country estimated amounts of embodied emissions, both in domestic final demand and those directly associated with the production sectors when the sectoral aggregation level has a low definition as considered in some recent similar studies.

Chapter 4

This chapter discusses the results for Italy of an extension of the standard CDM model (Crepon et al, 1998) used to evaluate innovation, with a focus on environmental innovations, and productivity patterns. The particular nature of environmental innovations, especially as regards the need of government intervention to create market opportunities, is likely to affect the way through which they are pursued (innovation equation within the CDM model) and their effect on productivity (productivity equation).

Here I test two main hypothesis: (i) to what extent polluting firms rely on own innovations to improve their environmental performance? (ii) do the pursue of environmental innovations reduce the likelihood of obtaining other profitable innovations (crowding out)? Results show that innovation efforts of polluting firms and sectors is significantly biased towards environmental innovations and that environmental innovations tend to crowd out other more profitable (at least in the short run) innovations.

I employ administrative data, namely the AIDA database by Bureau Van Dijk with balance sheet and income statement information further extended with patent applications from PATSTAT and information on emissions contained into the European Pollutant Release and Transfer Register. Administrative data allow to increase substantially the size of the sample and to reduce the selection bias of innovation surveys. However, administrative data are characterized by serious limitations in terms of definition of the variable of interest and measurement errors. The empirical approach used in the current chapter aims at reducing those biases.

Appendix A

This chapter is a brief methodological note on the matching of Italian firms in the AIDA (Bureau van Dijk) database with patent applicants to the European Patent Office (EPO) from the PATSTAT database.

The challenging task of matching data on patent applications with databases with balance sheet information is motivated by the relevance of patent statistics as source of information on the innovative performance of firms. In addition to simple count, patent data allow to investigate the technological specialization of innovating firms (through the IPC classification and search strategies on abstracts and titles), possible knowledge flows embodied in patent citations and the technological relevance of patented innovations through citation counts.

Using as starting point recent efforts in the matching of applicants in PATSTAT with firms in *Bureau van Dijk* databases (ORBIS, AMADEUS, FAME), I combined an improved cleaning routine to maximize exact matches followed by an approximate matching based on multiple combination of similarity scores.

Starting from about 270k AIDA firms, I matched about 50k EPO applications for the period 1977-2009. The matching covers, on average, 68 percent of EPO applications by Italian firms, which increases to 89 percent for 2000-2009. Finally, I describe the temporal, sector, size, geographic and technological distribution of the matched patent applications.

Chapter 1

Introduction

Changes in consumption patterns, technological change and environmental innovations are essential to achieve the reduction of environmental externalities, to slow down the exploitation of natural resources and to reduce the likelihood of environmental disasters (e.g. climate change). The current dissertation touches some of these aspects with an empirical perspective.

Content of the dissertation

Chapter 2 investigates the extent to which manufacturing sectors in European countries converge in terms of emission efficiency and assesses the role played by economic TFP and energy prices as drivers of convergence (if any).

The focus of chapter 3[1] is, instead, on consumption patterns. The chapter reports estimates of air emissions induced by domestic final consumption for Italy and Spain. Moreover, original data are aggregated in order to estimate the aggregation bias which is likely to affect most of the studies based on Multi Regional Input–Output Models.

[1]Chapter 3 is the reproduction of the article "Linking NAMEA and Input Output for 'Consumption vs. Production Perspective' Analyses - Evidence on Emission Efficiency and Aggregation Biases using the Italian and Spanish Environmental Accounts" (with Massimiliano Mazzanti and Anna Montini), 2012, *Ecological Economics*, 74:71-84.

Finally, chapter 4 moves to the micro (firm) level perspective to investigate the drivers of environmental innovations and their effect on productivity for Italian manufacturing firms. I use a modified version of the so-called CDM model (Crepon et al, 1998), which is a structural empirical model to investigate innovation patterns at the firm level, and I estimate it by using administrative data (balance sheet and patent data[2]).

Research questions

To summarize, the current dissertation aims at dealing with the following research questions:

1. Is it possible to observe any convergence pattern in environmental efficiency of manufacturing sectors across European countries? (Chapter 2)

2. Which are the drivers of convergence (if any) and diffusion of more environmentally efficient technologies across European countries? (Chapter 2)

3. Does the divergence between consumption and production structures of Italy and Spain lead to a changes in the balance between emissions produced domestically and emissions induced worldwide? (Chapter 3)

4. Which is the role played by aggregation (bias) within Environmentally Extended Input–Output (EEIO) models when estimating the amount of emissions induced worldwide? (Chapter 3)

5. Which are the drivers of environmental innovation activities of firms? To what extent do they differ from the drivers of other innovations? (Chapter 4)

[2] Appendix A describes in detail the matching between the AIDA database (Bureau van Dijk, with balance sheet information) and the PATSTAT database (OECD, with patent data).

6. Does the effect on productivity of environmental innovations differ from the effect of other innovations? Is there any evidence of (indirect) crowding out? (Chapter 4)

Main results

The empirical evidence related to the previous set of research questions of each chapter is summarized as follows.

Chapter 2

Emission efficiency growth of manufacturing sectors in European countries tends to follow a convergence pattern, with countries far from the frontier having a greater emission efficiency growth. However, technological improvements at the frontier tend to favour the growth of emission efficiency of the sector in all countries. Energy prices dynamics weakly drives emission efficiency growth, the effect changing with the distance from the frontier. Finally, there is a very strong relationship between emission efficiency growth and productivity (total factor productivity - TFP) growth while laggard countries (in terms of TFP) are generally characterized by a slower growth of emission efficiency.

Chapter 3

The difference between air emissions induced by domestic final consumption (consumption perspective) and domestic direct emissions arising from production (production perspective) is generally positive (emission leakage) for Spain and negligible for Italy. However, when starting from more aggregated data, estimates change substantially (up to 40 percent) relative to the benchmark. The relevance of the aggregation bias raises some concern when commenting on estimates of the consumption perspective based on models with low sector resolution.

Chapter 4

Environmental innovations tend to be more sensitive to firm size and R&D than other innovations. Moreover, polluting firms and sectors are generally biased towards environmental innovations as opposed to other firms and sectors. Looking at the productivity effect of innovations, the ones related to environmental technologies tend to have a positive effect which is lower than that found for other technologies and, in some cases, even a negative effect, especially for polluting firms. These results might be an evidence of crowding out of environmental innovations at the expenses of other more profitable innovations.

Contribution to the literature

This last section of this introductory chapter aims at stressing the most important innovative contributions to the economic literature of the current dissertation chapter-by-chapter.

Chapter 2

The economic empirical literature on the cross-country convergence of sectors generally focused on the dynamics of total factor productivity while no analysis has been performed on the dynamics of the environmental performance. Chapter 2 is, to my knowledge, the first attempt to fill this gap. Moreover, the analysis is based on recently delivered high-quality environmental data (environmental accounts NAMEA[3]) merged with other more standard data sources on economic variables (OECD Stan and Eurostat).

Chapter 3

The main innovative contributions of chapter 3 are the use of high-quality data sources (environmental accounts) for quite standard analyses of 'consumption perspective' estimates and, more importantly, the

[3]National Accounting Matrix including Environmental Accounts.

empirical assessment of the aggregation bias in environmentally extended input-output (EEIO) models. Aggregation bias has been generally ignored in the empirical literature using EEIO models with very few exceptions.

Chapter 4

The innovative contribution to the economic literature of chapter 4 regards both the model and the data I employed. To my knowledge, chapter 4 is the first attempt to consider at the same time the drivers of environmental innovations and their effect on firm's productivity, dealing with endogeneity issues. Moreover, while the literature on the identification of the drivers of environmental innovations is quite rich (although often based on ad hoc surveys), few empirical works dealt with the issue of the effect of environmental innovations on productivity and on possible crowding out effects. Finally, the use of administrative data (balance sheet and patent information) instead of ad hoc survey, which is another innovative contribution to the literature, reduces substantially measurement errors arising from self reporting and allows to rely on much larger samples as opposed to survey data.

Chapter 2

Closing the Gap? Dynamic Analyses of Emission Efficiency and Sector Productivity in Europe

2.1 Introduction[1]

A key factor in the attainment of environmental sustainability is the improvement of environmental efficiency of production and consumption activities. Environmental efficiency improvements at the aggregate (country) level are a combination of structural change, with a shift of production and consumption toward more environmentally friendly sectors and products, and improvements in environmental efficiency within sectors and product categories determined by technological change[2]. In this framework, technological change directed at reducing environmental pressures is characterized by a double externality problem, with im-

[1] A preliminary version of the current chapter has been published as book chapter in the book "The Dynamics of Economic and Environmental Efficiency", (eds V. Costantini, M. Mazzanti), Springer, 2012.

[2] For an extensive review of the literature on the role of technological change in environmental issues, refer to Popp et al (2009).

provements in environmental efficiency (reductions in negative externalities) not valued by the markets in absence of specific regulations and with the usual knowledge spillovers (positive externality) that reduce the incentives to innovate (Jaffe et al, 2005).

The correction of the double externality requires a combination of both environmental and innovation policies to stimulate the introduction and diffusion of more efficient technologies and products. During the last decades, European institutions promoted the convergence to a common EU-wide framework for environmental policies. Among other reasons, highly heterogeneous environmental policies across European countries may induce distortions to competition and strategic uses of environmental policies to favour domestic economic actors. Strategic use of environmental policies could have led to a 'race to the bottom' to the less stringent standard. Moreover, the achievement of environmental sustainability has been identified by the Lisbon Strategy in 2000 both as an objective per se and as a mean of transforming the EU into 'the most competitive and dynamic knowledge-based economy in the world'[3].

In order to reduce the burden of environmental regulations for producers and consumers and exploit the potential early-mover advantage in environmental technologies, international diffusion of environmental innovations and technologies should be favoured. A harmonized and stable regulatory framework favours more radical (environmental) innovations and the transition to more environmentally efficient production technologies through the adoption of environmental innovations. Lanjouw and Mody (1996) investigate the diffusion of environmental innovations using data on environmental patents and on trade flows in pollution control equipment. They emphasize the importance of both embodied (in pollution control equipment) and disembodied (through international patenting) diffusion of environmental innovations and the relevance of regulatory stringency as driver of diffusion. Popp (2006) investigates the extent to which the rate of patenting in pollution abatement technologies was triggered by the introduction of NOx and SO2 regulations in the US, Japan and Germany, the world's technological leaders.

[3]http://europa.eu/scadplus/glossary/lisbon_strategy_en.htm

Environmental innovations in these countries respond to both domestic and foreign environmental regulations. An interesting result in Popp (2006) is the need for 'domestic' knowledge even when domestic regulations follow regulations and innovation efforts in other countries. Foreign environmental innovations introduced to reduce compliance costs in early regulator countries, once adopted by 'followers', are not enough and follower countries need to introduce complementary innovations.

Another channel through which environmental efficiency in the technological leader countries and the distance from the leader affect domestic environmental efficiency is related to the diffusion of environmental policies. Lovely and Popp (2011) use data on patented innovations for SO2 and NOx emissions abatement in coal-fired power plants to show the extent to which innovations in countries on the technological frontier induce the introduction of more stringent pollution control policies in other countries. Improvements in the abatement technology obtained in leader countries reduce the abatement costs in other countries thus favouring the diffusion of more stringent environmental standards.

The diffusion of technologies to improve environmental efficiency may also occur within a country through inter-sectoral flows of knowledge (Corradini et al, 2011). Knowledge flows may occur both by embodiment of more efficient environmental technologies in intermediate goods or capital goods and by pure 'immaterial' knowledge flows.

A final consideration relates to domestic drivers of emission efficiency. Environmental regulation is expected to be a crucial factor in spurring environmental efficiency, especially due to the (pure or impure) public good nature of environmental efficiency improvements. Even though different kinds of environmental regulation are characterized by heterogeneous levels of efficiency in meeting their environmental targets[4], the effect of environmental policies is in the direction of improv-

[4]Environmental regulations can be classified according to various criteria. The most common distinction is between command-and-control regulations, with no reward for over-compliance, and market-based regulations, according to which environmental externalities are priced. A second classification which is relevant in the context of this chapter is related to the environmental scope of regulations, that is, the variety of environmental issues targeted by the regulation. Regulations with a wide scope are likely to reduce overall compliance costs for single policy instruments because they exploit the complementarities

ing environmental efficiency by definition[5]. Another important 'domestic' driver of emission efficiency is the domestic stock of knowledge in environmental technologies (Carrión-Flores and Innes, 2010). Domestic actors may strategically invest in environmental innovations to exploit early mover advantages in the world markets for environmental technologies. These strategies could be partly independent of the incentives to reduce compliance costs for domestic environmental policies (Porter and van der Linde, 1995). The 'side effects' of these innovation strategies may be an autonomous (from environmental policies) improvement of domestic environmental efficiency and the tightening of domestic environmental policies as a consequence of reduced compliance costs. Environmental policies and environmental innovation strategies are generally targeted to very narrow environmental issues, which could limit their effects on specific economic sectors or to specific environmental problems. Moreover, market-based environmental policies such as environmental taxes and emission trading schemes are generally characterized by low values for external costs (taxes) and polluting rights (emission trading schemes), leading to weak inducement effects. This weak inducement has been substantially compensated by the dynamics of energy prices. Due to their pervasiveness (Costantini and Mazzanti, 2010), with effects on the whole supply chain and on consumers, energy prices have been identified as a crucial driver of energy efficiency (Newell et al, 1999; Popp, 2002), which is one of the most important components of emission efficiency strategies[6]. The channel through which energy prices are likely to improve energy (and thus emission) efficiency is the classical idea of Hicksian induced innovation, according to which an increase in the relative price of an input triggers innovation aimed at reducing the use (i.e. increasing the efficiency) of that input. Energy price shocks,

between the abatement of distinct environmental externalities in a more efficient way.

[5]Policies aimed at targeting specific environmental issues may, however, generate negative effects on other environmental issues.

[6]The link between energy efficiency and emission efficiency is very strict for CO2 emissions because, differently from other air pollutant, they cannot be easily abated by means of filters or, more generally, end-of-pipe equipment. Moreover, in addition to aggregate energy price indexes, the relative price of different fossil fuels is likely to substantially affect the environmental effect of energy price patterns due to changes in the fuel mix.

such as oil shocks in 1973 and 1980, were sources of very significant structural changes in carbon dioxide emissions (Mazzanti and Musolesi, 2010; Moomaw and Unruh, 1997) while regulatory efforts such as the ratification of the Kyoto protocol did not generate significant breaks (Marin and Mazzanti (in press) for Italy). The pervasiveness of energy prices as a driver of emission efficiency also regards the great variety of air emissions affected by changes in energy prices and induced improvements in energy efficiency. On the one hand, high overall prices induce end-use improvements in energy efficiency, with a reduction (or a slow down) of energy production and beneficial effects on the abatement of all types of air emissions. On the other hand, shocks affecting the price of specific fuels will also induce changes in the energy mix, with differentiated effects on different types of emissions.

To sum up, this chapter aims to find evidence for the following research questions:

- what are the drivers of sectoral emission efficiency growth in Europe?
- to what extent do improvements in emission efficiency in the technological frontier spread to laggard countries? What is the role of the emission efficiency gap?
- do energy prices dynamics affect emission efficiency growth? Does this inducement change according to the distance from the emission efficiency frontier?
- do productivity (total factor productivity) growth and gap affect the pattern of emission efficiency?
- are there systematic differences between different types of emissions?

The chapter is organized as follows. Section 2 discusses the empirical model used to investigate the drivers of sectoral emission efficiency, section 3 describes data sources, section 4 discusses the most relevant results and section 5 concludes.

2.2 Model

In order to investigate the drivers of emission efficiency improvements and the patterns of emission efficiency diffusion I use an adapted version of a quite standard empirical framework to account for productivity growth at the industry level. The general idea[7] is that productivity level (total factor productivity - TFP - in early applications of the model) is an ARDL(1,1)[8] process which is cointegrated with the level of TFP of the technological frontier. Under the assumption of long run homogeneity, TFP growth is described by the following equation:

$$\begin{aligned} \triangle \log(\mathrm{TFP}_{c,s,t}) = \beta_1 \triangle \log(\mathrm{TFP}_{F,s,t}) + \\ + \beta_2 \left[\log(\mathrm{TFP}_{F,s,t-1}) - \log(\mathrm{TFP}_{c,s,t-1})\right] + \epsilon_{c,s,t} \end{aligned} \tag{2.1}$$

Productivity growth in country *c*, sector *s* and year *t* is positively related to the growth in the technological frontier country *F* and to the distance from the technological frontier. The rationale is that improvements in productivity in the most productive countries (technological frontier) enlarge the production possibility set (Nicoletti and Scarpetta, 2003) allowing laggard countries to improve their own productivity. Moreover, conditional on that effect, the distance from the technological frontier (technological gap) is expected to positively affect productivity growth. The idea is that the greater the distance from the frontier, the greater the marginal returns of adopting new technologies. A positive β_2 will result in a decreasing speed of convergence the closer a sector is to the frontier.

This basic model was employed in several OECD studies to investigate the effect of innovation, labour market institutions (Scarpetta and Tressel, 2002), product market competition and anticompetitive regulations (Nicoletti and Scarpetta, 2003) on productivity growth.

In the current chapter I adapt this model to estimate improvements (if any) of sectoral emission efficiency. Emission efficiency growth (expressed in terms of value added per unit of emission) is a function of

[7] I briefly describe the model used by Scarpetta and Tressel (2002), and Nicoletti and Scarpetta (2003).

[8] Auto regressive distributed lag of order 1.

emission efficiency growth in the frontier country and of the gap in emission efficiency from the frontier country. Growth of emission efficiency 'at the frontier' is expected to induce improvements in all countries due to the (partial) international diffusion of new, more efficient technologies. Diffusion may take place through various channels: embodiment in capital goods, imitation or disembodied transfer (e.g. patent licensing).

Moreover, I expect overall (economic) production technology to play a role in emission efficiency growth. The idea is that a technology that improve 'economic' productivity (i.e. greater value added for the same amount of inputs) will also result in a (either intended or not) increase in 'environmental' efficiency. To account for this effect I add TFP growth (both in the country and in the frontier) and the technological gap in terms of TFP as covariates. I expect domestic TFP growth to positively positively emission efficiency. Both Mazzanti and Zoboli (2009) and Marin and Mazzanti (in press) consider the relationship between labour productivity and emission efficiency for Italian sectors, testing for non-linearities. Depending on the indicator for emission efficiency (emission per value added in Mazzanti and Zoboli (2009) and emission per labour unit in Marin and Mazzanti (in press)[9]), they find either weak (emission per labour) or moderate (emission per value added) complementarity between emission efficiency and labour productivity, with the magnitude being specific to both emission type and macro-sector. Cole et al (2005) use a more structured empirical model to assess the role of industrial characteristics and environmental regulation in determining the level of sectoral air pollution for the UK. Among other regressors, they consider the effect of total factor productivity on air emissions, finding a negative (increased emission efficiency) significant effect in most of the specifications. These results highlight the potential complementarities between economic (productivity) and environmental (efficiency) performance, at

[9] In a log-linear setting, it is possible to evaluate the relationship between estimates using emission per labour (E/L) and estimates using emission per value added (E/VA). The log-linear relationship between emission per value added and labour productivity (VA/L) is given by $E/VA = (VA/L)^{\beta}$. By multiplying both sides by VA/L and rearranging, the relationship becomes $E/L = (VA/L)^{\beta+1}$, which means that, by construction, the coefficient in a log-linear setting using E/L as emission efficiency indicator is exactly equal to the coefficient when using E/VA as emission efficiency indicator plus one.

least at the sector level.

In addition to this direct effect, being distant from the technological leader could be an indication of general technological laggardness of the sector, with potential negative effects on both economic and environmental performance. Finally, TFP growth in the frontier country is included in order to account for the dynamics of the state of the technology of a sector.

To conclude, I investigate the effect of country-wide industry energy prices dynamics on emission efficiency. Following the approach of Scarpetta and Tressel (2002) and Nicoletti and Scarpetta (2003), whose focus is on product market regulations, I assume the inducement effect of energy prices on emission efficiency to change with the distance from the emission efficiency frontier. The idea is that very inefficient countries suffer more in terms of additional production costs than efficient countries because of a given increase in energy prices due to their greater energy (and thus emission) intensity of production. This potential higher costs is likely to amplify the inducement effect of energy prices on laggard countries.

The empirical model used here is described by the following equation:

$$\begin{aligned}
\triangle \log(\mathrm{VA}_{c,s,t}/\mathrm{E}_{c,s,t}) = \beta_0 + \beta_1 \triangle \log(\mathrm{VA}_{F,s,t}/\mathrm{E}_{F,s,t}) + \\
+\beta_2 \mathrm{gap_log}(\mathrm{VA}_{c,s,t-1}/\mathrm{E}_{c,s,t-1}) + \beta_3 \triangle \log(\mathrm{TFP}_{c,s,t}) + \\
+\beta_4 \triangle \log(\mathrm{TFP}_{F,s,t}) + \beta_5 \mathrm{gap_log}(\mathrm{TFP}_{c,s,t-1}) + \\
+\beta_6 \triangle \mathrm{ener_prices}_{c,t-1} + \\
+\beta_7 \triangle \mathrm{ener_prices}_{c,t-1} \times \mathrm{gap_log}(\mathrm{VA}_{c,s,t-1}/\mathrm{E}_{c,s,t-1}) + \\
+\eta_c + \gamma_s + \delta_t + \epsilon_{c,s,t}
\end{aligned} \tag{2.2}$$

where $\triangle \log(\mathrm{VA}_{c,s,t}/\mathrm{E}_{c,s,t})$ represents the relative change in sectoral emission efficiency, $\triangle \log(\mathrm{VA}_{F,s,t}/\mathrm{E}_{F,s,t})$ is the relative change in sectoral emission efficiency in the frontier country, $\mathrm{gap\_log}(\mathrm{VA}_{c,s,t-1}/\mathrm{E}_{c,s,t-1})$ is the distance of sector s in country c from the emission efficiency frontier, $\triangle \log(\mathrm{TFP}_{c,s,t})$ is TFP growth, $\triangle \log(\mathrm{TFP}_{F,s,t})$ is TFP growth in the frontier country, $\mathrm{gap\_log}(\mathrm{TFP}_{c,s,t-1})$ is the gap from the TFP frontier, $\triangle\mathrm{ener\_prices}_{c,t-1}$ is the relative change in industrial energy prices and η_c, γ_s and δ_t are, respectively, country, sector and year dummies.

All estimates have been performed using OLS regressions, with standard errors clustered by sector and country.

2.3 Data

I use sectoral data at the 2-digit NACE level covering 23 manufacturing sectors in 13 European countries (Austria, Belgium, Czech Republic, Germany, Denmark, Spain, Finland, France, Italy, the Netherlands, Norway, Sweden and the UK) over 12 years (1996-2007). The selection of countries is based on the availability of relevant data and by trying to include all large countries which are likely to be among the technological leaders of Europe. Some EU15 country have been excluded due to the very limited data coverage (Luxemburg, Portugal, Greece and Ireland). The choice to include Norway (which is not part of the European Union) is motivated by the fact that it is likely to belong to the group of technological leaders both in productivity and emissions efficiency and by the fact that Norway, through its membership of the European Environment Agency, partly shares the environmental regulatory framework of EU countries[10]. Moreover, the only country I included among those which joined the EU in 2004 is Czech Republic[11] because no other country had a satisfactory data coverage. A final consideration is needed concerning the focus on Europe only. Although many European countries are included in the group of technological leaders (both in terms of productivity and environmental efficiency), in many fields, the European technological frontier does not always coincide with the global technological frontier. In addition to Western European countries, the US, Canada, Japan, Australia and South Korea were found to be among the technological leaders (at

[10] Another potential technological leader in Europe not belonging to the EU27 is Switzerland. However, due to a very high proportion of missing observations in relevant variables, its inclusion in the sample was not possible.

[11] Results excluding Czech Republic do not change substantially from those reported in this chapter.

least third in the ranking) by Scarpetta and Tressel (2002) based on TFP. The absence of these countries is likely to downward bias the relative gap from the frontier (either technological or for emission efficiency) and reduce the reliability of estimated improvements of the TFP and emission efficiency frontiers.

Data on value added, employment and gross fixed capital formation come from Eurostat and the OECD STAN (Structural Analysis) database. Missing values in the OECD STAN database were filled with data from EUROSTAT. Value added (in Euro) was deflated to 2000 prices according to country-specific deflators for manufacturing[12]. In the version of the results reported in the current chapter, no PPP (Purchasing Power Parity) adjustment was performed[13].

The capital stock variable, needed to obtain TFP estimates, was built by using the perpetual inventory method. Data on capital stock in OECD STAN has several missing values as well as the variable 'gross fixed capital formation' in constant prices. I use gross fixed capital formation (GFCF) in current prices, deflated with country-specific manufacturing deflators. The initial (1980, when available, or the first year of the series of sectoral gross fixed capital formation) fixed capital stock (K) for sector s and country c was set to:

$$\mathrm{K}_{c,s,0} = \mathrm{GFCF}_{c,s,0}/(\delta + g) \tag{2.3}$$

where g is the average growth rate (set to zero when negative) of GFCF in the first 5 years of the series and δ is the depreciation rate (set to 0.04). For $t > 0$, the fixed capital stock was computed according to the following equation:

[12] When using sector-specific deflators for value added and aggregate deflators for gross fixed capital formation, production function estimates are not plausible, with negative elasticity for capital.

[13] Estimates excluding Norway were performed using time-invariant PPP (sector-specific or aggregate for manufacturing goods) adjustments obtained from EU KLEMS (www.euklems.eu). Results for the emission efficiency growth equation did not change substantially while the estimates of the labour and capital shares in the production function were quite unstable. However, sector-level PPP coupled with aggregate price deflators is likely to give rise to substantial measurement errors.

$$\mathrm{K}_{c,s,t} = (1-\delta) \times \mathrm{K}_{c,s,t-1} + \mathrm{GFCF}_{c,s,t} \tag{2.4}$$

Data on labour input refers to simple employees count (OECD STAN). This is an imperfect measure of labour input because there is no adjustment for full-time / part-time employees and for the actual number of hours worked. However, country coverage and reliability of employees count was much greater than measures of total hours worked or full-time equivalent estimates. Robustness checks were performed on a sub-sample with information on hours worked and full time equivalent estimates: no relevant difference was found[14].

Data on sectoral air emission come from the Eurostat NAMEA (National Accounting Matrix including Environmental Accounts) database. By construction, environmental pressures reported in NAMEA are consistent with the full set of national economic accounts because they use the same definitions and classifications as national accounts. The main advantage of NAMEA relative to standard environmental statistics is the direct link between environmental externalities and economic aggregates, based on the residential principle (environmental pressures by resident units only) and on the consideration of anthropogenic sources only (emissions from natural sources such as volcanos are excluded). Moreover, the European NAMEA currently covers a remarkable variety of air emissions. Here I focus on air emissions of carbon dioxide (CO2), sulphur oxides (SOx), nitrogen oxides (NOx), non-methane volatile organic compounds (NMVOC) and carbon monoxide (CO). The main source of all emissions is the combustion of fossil fuels[15]. For additional information on the features of these emissions refer to the final section (Air emission features) of the current chapter.

Finally, data on energy price come from IEA and they describe yearly relative changes in the price index of energy inputs for the industrial sector.

[14]Pairwise correlation among employees count, hours worked and full-time equivalent estimates is slightly above 99.5 percent.

[15]Other relevant sources of NMVOC emissions are paintings, solvents and coatings.

In order to obtain a rough estimate of the level of the production technology, I compute an approximate measure of total factor productivity (TFP henceforth). TFP has been estimated as the residual of a constant returns to scale Cobb-Douglas production function, with value added as output measure and capital stock and labour (employees count) as inputs. The sum of the labour and capital coefficients was constrained to be 1 (constant returns to scale) and year and sector dummies were included in order to control for sector-specific technologies and Europe-wide shocks. The estimated labour share, corresponding to the elasticity of value added with respect to labour under the assumption of perfect competition, is 61.5 percent. Alternative measures of TFP[16] were employed with very small changes in the results.

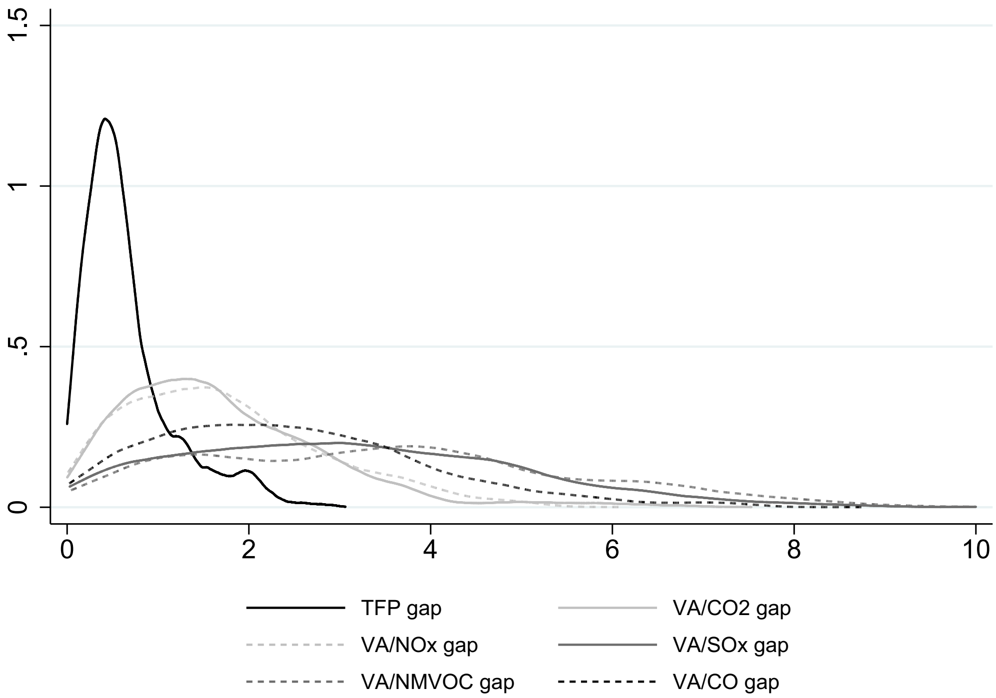
This data potentially relies 3588 observations. Despite ad hoc adjustment, some missing values remain[17]. Moreover, I excluded both outlier observations (labour productivity growth or reduction greater than 50 percent) and small sectors (the first percentile of sectors in terms of manufacturing value added or employment) to avoid potentially great measurement errors in sector representing a negligible share of an economy. Measurement errors may depend on the fact that a very small sector could include secondary activities only, with little or misleading information on the true state of the technology and on emission efficiency of the sector in a specific country.

Figure 1 shows the distribution of percentage gaps in both TFP and environmental efficiency (value added per unit of emissions). Interestingly, environmentally efficiency is much more dispersed than productivity with still great potentials for laggard countries and sectors to converge towards more environmentally efficient technologies. The lack of convergence depends on the 'external' nature of the benefits arising from

[16] Alternative measures consisted in TFP estimated as the residual of a translog production function and a Cobb-Douglas with no CRS assumption. Moreover, estimates on smaller samples with value added and gross fixed capital formation deflated with sector-specific deflators gave rise to very similar results in terms of labour share and TFP estimates.

[17] Spain for 1996, France for 1996-1999 (except sectors 20, 26 and 29, for a total of 80 missing values), Netherlands 1996-2001 (except sectors 20-29, for a total of 78 missing values) and other more scattered missing data.

Figure 1: Distribution of productivity and environmental efficiency relative gaps



environmental efficiency improvements as opposed to standard TFP improvements.

The gap is relatively small for CO2 and NOx emission efficiency while it is relevant for CO, SOx and NMVOC emission efficiency. This may seem a quite surprising result, given that local pollutants are regulated more strictly than CO2 emissions at European level, with potential greater homogeneity. However, pollutants are generally reduced with end-of-pipe technology which represent a pure cost for polluting firms while carbon dioxide emissions are very strongly correlated to energy use. The generally lower gap in CO2 efficiency could be the result of its strict correlation with energy use which is characterized by a substantial component of private benefit relative to pollutant emissions.

Descriptive statistics for relevant variables are reported in table 1.

Table 1: Descriptive statistics

Variable	Mean	25th pct	Median	75th pct	Min	Max	Coeff var
log(VA)	21.39	20.37	21.39	22.54	16.18	25.11	.07281
log(K)	22.36	21.26	22.43	23.53	16.96	26.33	.07524
log(L)	10.5	9.62	10.49	11.64	5.298	13.93	.147
△ energy prices	.02744	-.004866	.02531	.05842	-.151	.1593	1.917
TFP	8.57e-09	-.1415	.06369	.2662	-2.041	1.652	-
log(VA/L)	10.89	10.69	10.93	11.19	8.191	13.16	.05993
log(VA/CO2)	15.63	14.62	15.96	16.76	10.43	23.2	.1109
log(VA/NOx)	14.61	13.52	14.75	15.72	9.943	20.75	.1088
log(VA/SOx)	16.09	14.21	16.19	17.95	9.095	25.58	.1642
log(VA/NMVOC)	14.74	13.17	14.39	16.21	7.653	23.32	.1434
log(VA/CO)	14.46	13.57	14.65	15.58	8.322	22.3	.1276
TFP gap	.6159	.264	.4996	.8006	0	3.106	.8482
log(VA/L) gap	.5434	.1728	.3646	.6871	0	2.986	1.043
log(VA/CO2) gap	1.584	.7163	1.412	2.251	0	7.359	.7591
log(VA/NOx) gap	1.604	.713	1.473	2.29	0	5.936	.7184
log(VA/SOx) gap	2.9	1.358	2.779	4.267	0	10.52	.6777
log(VA/NMVOC) gap	3.217	1.378	3.156	4.625	0	11.87	.6849
log(VA/CO) gap	2.281	1.06	2.138	3.284	0	8.496	.6969
TFP frontier	.6159	.4383	.5547	.7284	.2201	2.071	.4224
log(VA/L) frontier	11.44	11.1	11.34	11.66	10.64	13.16	.04423
log(VA/CO2) frontier	17.21	16.06	17.38	18.39	12.4	23.2	.1175
log(VA/NOx) frontier	16.21	14.92	16.41	17.3	12.1	20.75	.1107
log(VA/SOx) frontier	18.99	17.31	19.13	20.68	10.79	25.58	.1419
log(VA/NMVOC) frontier	17.96	16.72	17.76	18.94	12.5	23.32	.1083
log(VA/CO) frontier	16.74	15.76	16.45	17.63	12.43	22.3	.09733

2.4 Results

For all emissions, I report results for various versions of the baseline model, from the simplest version with no role for energy prices and TFP to the most complete version including energy prices and TFP. Results for the full sample of manufacturing sectors are reported in tables 2-6. A first remarkable result is the positive effect of emission efficiency improvements in the frontier country on domestic sectoral emission efficiency growth. This result is robust in all specifications and for all emissions, its magnitude ranging from an elasticity of 0.03-0.04 for CO2 emissions to an elasticity of 0.09-0.1 for SOx emissions. As expected, improvements in environmental efficiency at the frontier spill over to laggard countries with a beneficial effect on their emission efficiency growth. These positive spillovers may occur as a consequence of the diffusion of more environmental efficient technologies from 'frontier' countries and sectors to laggard countries and sectors.

The distance from the frontier country in terms of emission efficiency affects[18] domestic emission efficiency growth positively and significantly

[18] I refer here to the direct effect assuming no energy price change ($\triangle \text{ener\_prices}_{c,t-1} =$

Table 2: Estimates for CO2 emission efficiency

$\triangle$ log(VA/CO2)	(1)	(2)	(3)	(4)	(5)	(6)
$\triangle$ VA/CO2 frontier	0.0319*	0.0384**	0.0342*	0.0407**	0.0379**	0.0404**
	(0.0192)	(0.0189)	(0.0189)	(0.0187)	(0.0170)	(0.0170)
VA/CO2 gap (t-1)	0.00530	0.0196***	0.00160	0.0152***	0.0130***	0.0184***
	(0.00515)	(0.00591)	(0.00533)	(0.00589)	(0.00497)	(0.00537)
$\triangle$ energy prices			0.0522	0.0317	0.0611	0.0540
			(0.144)	(0.145)	(0.132)	(0.133)
$\triangle$ energy prices $\times$ VA/CO2 gap (t-1)			0.142*	0.152*	0.0913	0.110
			(0.0791)	(0.0781)	(0.0718)	(0.0715)
$\triangle$ TFP					0.981***	0.960***
					(0.0396)	(0.0399)
$\triangle$ TFP frontier					-0.0446*	-0.0519**
					(0.0247)	(0.0248)
TFP gap (t-1)					-0.0431***	-0.0704***
					(0.00786)	(0.0152)
Constant	-0.0240	-0.0493**	-0.0244	-0.0494**	-0.0107	-0.0314
	(0.0213)	(0.0227)	(0.0219)	(0.0231)	(0.0212)	(0.0237)
F	3.008***	5.970***	3.229***	5.874***	28.81***	26.02***
R squared	0.0290	0.0703	0.0337	0.0751	0.268	0.288
Year dummies (F)	5.182***	5.338***	6.198***	6.365***	8.716***	8.891***
Sector dummies (F)	0.780	0.851	0.762	0.833	0.996	1.136
Country dummies (F)		12.08***		12.19***		5.214***
Ramsey o.v. test (F)	0.662	3.590**	0.831	4.813***	0.469	1.657
N	3115	3115	3115	3115	3115	3115

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

for all emissions except NMVOC. This generally positive effect is a clear evidence of (beta) convergence in emission efficiency of laggard countries towards the emission efficiency frontier, with the speed of convergence being greater for countries and sectors with the biggest gap. It is evident from figure 1 that there are huge potentials of convergence in emission efficiency performance. However, it is clear that to accelerate the rate of convergence there is a need for further harmonization of environmental policies across countries and additional effort made to promote the diffusion of efficient technologies. The negative effect of the efficiency gap for NMVOC is small in magnitude and insignificant when including either country fixed effects or TFP growth (domestic and frontier country) and gap. Unlike other types of emission, NMVOC emission efficiency is not characterized by convergence patterns.

The coefficient for the change in energy prices (β_6) describes the effect of prices on emission efficiency growth as if the sector was the technological leader whereas the actual effect of prices is given by $\beta_6 + \beta_7 \times$ gap_log$(\mathrm{VA}_{c,s,t-1}/\mathrm{E}_{c,s,t-1})$. The effect on frontier sectors is always positive although it is significant for NMVOC and CO emissions only. The

0) in columns 3-6.

Table 3: Estimates for NOx emission efficiency

$\triangle$ log(VA/NOx)	(1)	(2)	(3)	(4)	(5)	(6)
$\triangle$ VA/NOx frontier	0.0562***	0.0637***	0.0580***	0.0654***	0.0593***	0.0604***
	(0.0202)	(0.0203)	(0.0198)	(0.0200)	(0.0186)	(0.0187)
VA/NOx gap (t-1)	0.0148***	0.0328***	0.0100**	0.0268***	0.0233***	0.0264***
	(0.00476)	(0.00667)	(0.00492)	(0.00653)	(0.00507)	(0.00618)
$\triangle$ energy prices			0.156	0.113	0.133	0.124
			(0.187)	(0.189)	(0.180)	(0.184)
$\triangle$ energy prices $\times$ VA/NOx (t-1) gap			0.185**	0.202**	0.146*	0.169**
			(0.0864)	(0.0848)	(0.0806)	(0.0803)
$\triangle$ TFP					1.016***	0.991***
					(0.0438)	(0.0436)
$\triangle$ TFP frontier					-0.0652**	-0.0665**
					(0.0319)	(0.0327)
TFP gap (t-1)					-0.0522***	-0.0657***
					(0.00933)	(0.0162)
Constant	0.00508	-0.0470**	0.00264	-0.0484*	0.0159	-0.0342
	(0.0216)	(0.0232)	(0.0232)	(0.0247)	(0.0229)	(0.0260)
F	2.335***	5.782***	2.838***	5.850***	21.98***	22.65***
R squared	0.0248	0.0630	0.0326	0.0704	0.210	0.230
Year dummies (F)	3.670***	3.786***	5.034***	5.154***	6.507***	6.536***
Sector dummies (F)	0.982	1.275	0.954	1.247	0.691	0.746
Country dummies (F)		15.35***		15.02***		11.73***
Ramsey o.v. test (F)	6.632***	3.234**	3.986***	8.240***	0.178	0.227
N	3115	3115	3115	3115	3115	3115

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

interaction term, on the other hand, is positive for CO2, NOx and SOx (weakly significant for CO2, significant for NOx and not significant for SOx) and negative for NMVOC (though not significant) and CO (significant). A positive effect means that the effect of energy price changes on emission efficiency growth is increasing in the gap in emission efficiency from the frontier country, making laggards countries more sensitive to price changes than frontier countries. When computing marginal effects, the effect of energy prices for CO2, NOx and SOx increases with distance from the frontier. The overall effect of energy prices turns out to be positive and significant (10 percent of significnce already at the first quartile of emission efficiency gap). For these emissions, energy prices trigger significant improvement in laggard countries while the emission efficiency frontier is not significantly affected.

On the contrary, the marginal effect of energy prices decreases in the emission efficiency gap for NMVOC and CO emissions even though it is still strongly significant at the 90 percentile of the emission efficiency gap. In these cases, energy prices dynamics generates a stronger incentive for sectors that are close to the emission efficiency frontier than for laggard sectors. A possible explanation for the opposite results relative to

Table 4: Estimates for NMVOC emission efficiency

$\triangle$ log(VA/NMVOC)	(1)	(2)	(3)	(4)	(5)	(6)
$\triangle$ VA/NMVOC frontier	0.0421*	0.0484**	0.0416*	0.0474**	0.0425*	0.0468**
	(0.0240)	(0.0243)	(0.0237)	(0.0239)	(0.0234)	(0.0236)
VA/NMVOC gap (t-1)	-0.0122***	0.00156	-0.00803**	0.00424	-0.00281	0.00503
	(0.00353)	(0.00451)	(0.00343)	(0.00496)	(0.00351)	(0.00488)
$\triangle$ energy prices			1.105**	1.099**	1.073**	1.054**
			(0.446)	(0.455)	(0.441)	(0.453)
$\triangle$ energy prices $\times$ VA/NMVOC gap (t-1)			-0.123	-0.115	-0.125	-0.113
			(0.0828)	(0.0825)	(0.0819)	(0.0822)
$\triangle$ TFP					0.939***	0.908***
					(0.0568)	(0.0584)
$\triangle$ TFP frontier					-0.0523	-0.0518
					(0.0385)	(0.0382)
TFP gap (t-1)					-0.0319***	-0.0532**
					(0.0109)	(0.0210)
Constant	0.0549**	-0.0119	0.0181	-0.0452	0.0287	-0.0335
	(0.0246)	(0.0265)	(0.0295)	(0.0330)	(0.0285)	(0.0321)
F	3.261***	4.919***	3.196***	4.754***	14.81***	16.03***
R squared	0.0373	0.0676	0.0479	0.0777	0.139	0.159
Year dummies (F)	2.119**	2.035**	2.967***	2.928***	3.289***	3.246***
Sector dummies (F)	3.155***	2.098***	3.032***	2.112***	3.462***	2.778***
Country dummies (F)		7.374***		7.223***		6.438***
Ramsey o.v. test (F)	15.64***	172.9***	148.9***	253.8***	35.27***	93.40***
N	3115	3115	3115	3115	3115	3115

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

CO2, NOx and SOx regarding the effect of energy prices may be related to opposite patterns of environmental technological change for laggards and frontier sectors. On the one hand, laggard sectors seem to focus on the improvement of energy efficiency (strongly correlated with CO2 efficiency) and on the abatement of more 'classical' pollutants such as SOx and NOx. On the other hand, sectors lying close to the emission efficiency frontier seem to be characterized by fewer energy inefficiencies (and, consequently, higher marginal costs to improve energy efficiency) and by higher marginal costs for the abatement of classical pollutants due to the long tradition of stringent environmental standards.

The inclusion of productivity measures (total factor productivity TFP growth in the sector and in the frontier country and TFP gap from the frontier) in the last two columns does not affect substantially the estimates of other parameters. However, considering TFP has the consequence of improving substantially the goodness of fit (R squared)[19]. As expected, the relationship between sectoral TFP growth and emission efficiency growth is positive and strongly significant, with coefficients

[19] No relevant improvements in the R squared is found for SOx estimates where the gain is of about 2-3 percent of explained variance.

Table 5: Estimates for SOx emission efficiency

$\triangle$ log(VA/SOx)	(1)	(2)	(3)	(4)	(5)	(6)
$\triangle$ VA/SOx frontier	0.0901***	0.0965***	0.0902***	0.0963***	0.0926***	0.0979***
	(0.0229)	(0.0232)	(0.0228)	(0.0231)	(0.0229)	(0.0230)
VA/SOx gap (t-1)	0.0467***	0.0605***	0.0452***	0.0582***	0.0521***	0.0608***
	(0.00740)	(0.00882)	(0.00785)	(0.00912)	(0.00894)	(0.00935)
$\triangle$ energy prices			0.475	0.380	0.398	0.282
			(0.365)	(0.362)	(0.363)	(0.360)
$\triangle$ energy prices $\times$ VA/SOx gap (t-1)			0.0753	0.0864	0.0684	0.105
			(0.0949)	(0.0944)	(0.0947)	(0.0932)
$\triangle$ TFP					0.861***	0.907***
					(0.0917)	(0.0969)
$\triangle$ TFP frontier					0.000683	-0.0368
					(0.0759)	(0.0752)
TFP gap (t-1)					-0.0697***	-0.151***
					(0.0247)	(0.0451)
Constant	0.0380	-0.00240	0.0251	-0.0118	0.0552	0.0485
	(0.0371)	(0.0472)	(0.0393)	(0.0471)	(0.0393)	(0.0496)
F	4.888***	5.329***	4.926***	5.316***	7.387***	7.263***
R squared	0.0580	0.0660	0.0610	0.0684	0.0882	0.0981
Year dummies (F)	7.833***	7.928***	8.220***	8.342***	8.384***	8.028***
Sector dummies (F)	2.778***	3.087***	2.758***	3.045***	3.354***	3.862***
Country dummies (F)		2.716***		2.458***		2.644***
Ramsey o.v. test (F)	47.20***	48.42***	47.88***	49.15***	11.55***	9.017***
N	3115	3115	3115	3115	3115	3115

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

varying from a minimum of 0.86 (SOx without country fixed effects) to 1.04 (CO without country fixed effects). This means that an increase in TFP translates into a very similar increase in emission efficiency conditional on other covariates. This very robust result highlights the strong complementarity between economic productivity and environmental efficiency. The effect of TFP growth in the frontier country has a generally negative effect on emission efficiency growth, with the coefficient being statistically significant just for CO2 (5 percent), NOx (5 percent) and CO (10 percent only when including country fixed effects, insignificant otherwise). The insignificant or negative effect of TFP growth in the frontier country may suggest that frontier technological change is not explicitly directed to improve emission efficiency and, in some cases, there is a weak evidence of 'emission-intensive' technical change. Finally, the gap in TFP from the frontier country negatively and significantly affects emission efficiency growth in all cases. The existence of a negative effect of TFP gap further stresses the complementarity links between economic and environmental performance, especially since differences in emission efficiency were already accounted for. As stated in the previous section, results employing alternative measures of TFP or using labour produc-

Table 6: Estimates for CO emission efficiency

$\triangle$ log(VA/CO)	(1)	(2)	(3)	(4)	(5)	(6)
$\triangle$ VA/CO frontier	0.0758***	0.0789***	0.0734***	0.0769***	0.0760***	0.0771***
	(0.0258)	(0.0259)	(0.0251)	(0.0253)	(0.0245)	(0.0247)
VA/CO gap (t-1)	0.0152***	0.0240***	0.0194***	0.0280***	0.0257***	0.0276***
	(0.00476)	(0.00622)	(0.00505)	(0.00666)	(0.00534)	(0.00646)
$\triangle$ energy prices			1.438***	1.335***	1.450***	1.381***
			(0.406)	(0.393)	(0.398)	(0.392)
$\triangle$ energy prices $\times$ VA/CO gap (t-1)			-0.225**	-0.196*	-0.273**	-0.237**
			(0.107)	(0.102)	(0.107)	(0.102)
$\triangle$ TFP					1.042***	1.013***
					(0.0753)	(0.0768)
$\triangle$ TFP frontier					-0.0688	-0.0720*
					(0.0428)	(0.0431)
TFP gap (t-1)					-0.0575***	-0.0789***
					(0.0129)	(0.0251)
Constant	0.0191	-0.00121	-0.0255	-0.0425	-0.00601	-0.0236
	(0.0314)	(0.0340)	(0.0355)	(0.0386)	(0.0343)	(0.0381)
F	2.541***	6.218***	2.626***	5.936***	10.71***	13.99***
R squared	0.0380	0.0756	0.0533	0.0883	0.143	0.168
Year dummies (F)	1.685*	1.733*	2.889***	2.832***	3.445***	3.113***
Sector dummies (F)	2.418***	2.415***	2.419***	2.414***	2.367***	2.150***
Country dummies (F)		16.21***		14.58***		11.39***
Ramsey o.v. test (F)	3.265**	96.58***	47.25***	166.7***	53.05***	116.5***
N	3115	3115	3115	3115	3115	3115

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

tivity give rise to qualitatively very similar estimates.

Some considerations on year, sector and country fixed effects are needed. Year and country dummies are jointly strongly significant in all specifications and for all emissions. Significant Europe-wide time dummies possibly highlight the relevance of regulatory efforts at the European level affecting all countries.

Sector dummies, on the contrary, are not jointly significant for both for CO2 and NOx estimates, highlighting quite uniform efficiency patterns among sectors within countries for these types of emissions. On the contrary, they are jointly strongly significant for SOx, NMVOC and CO, highlighting heterogeneous patterns of emission efficiency potentially driven by sector-specific environmental regulations.

Country dummies are jointly strongly significant in all cases, stressing the great heterogeneity of environmental efficiency and highlighting the relevance of systematic differences among countries in emission efficiency dynamics even after controlling for the gap in environmental efficiency and productivity.

Results reported in the chapter do not change substantially when performing some simple robustness checks. The inclusion of outliers or small sectors does not influence either the magnitude or the significance of estimated coefficients. The use of more aggregate sector information, for example at the level of subsection NACE with 14 manufacturing sectors, reduces the significance of many coefficients but the magnitude does not change[20].

When removing specific countries or sectors (one by one) the magnitude of estimated coefficients does not change substantially even if significance is generally lower. Finally, tests on the presence of structural breaks in estimated coefficients were performed[21]. No significant structural break was found for CO2 and NMVOC emissions. Statistically significant breaks were found for NOx (1998 and 2000), SOx (2005) and CO (1999, 2001, 2002 and 2005) even though just three of them were significant at the 1 percent level (NOx 2000, SOx 2005 and CO 1999).

Tables 7-10 report estimates for sub-samples of sectors: sectors covered by the ETS[22] (Emission Trading System for carbon dioxide emissions, introduced in 2005) or not, medium-high technology and medium-low technology sectors[23]. Results for ETS sectors (table 7) tend to be quite volatile, with very low significance for most covariates except TFP growth. The reduced significance may depend on the small size of the sample (less than one fourth of the full sample). Comparing the magni-

[20] The level of aggregation of sectoral data is always a relevant issues when dealing with indicators of environmental efficiency. An improvement in emission efficiency could simply be the result of the changing composition of sectors within the considered macro-sector (e.g. 2-digit) towards more emission efficient sub-sectors (e.g. 4-digit). The observed improvement may thus occur even in absence of any change in the production or abatement technology of the sectors. The reader should always consider this caveat when interpreting the results.

[21] I performed a Chow test by interacting a dummy variable identifying a specific time period with all covariates in the model described by equation 2. The test (a simple F test) is performed by assuming, under the null hypothesis, that the parameters of all interaction terms are jointly equal to zero, thus indicating no structural break.

[22] The European ETS for carbon dioxide emissions covers plants operating in the following 2-digit NACE Rev 1.1 sectors: 21 (pulp, paper and paper products), 23 (coke, refined petroleum products and nuclear fuel), 26 (other non-metallic mineral products), 27 (basic metals) and 28 (fabricated metal products, except machinery and equipment).

[23] According to the OECD, medium-high technology manufacturing sectors include the following 2-digit NACE Rev 1.1 sectors: 24 (chemicals and chemical products), 29 (machinery and equipment n.e.c.), 30-33 (electrical and optical equipment) and 34-35 (transport equipment) while the remaining manufacturing sectors are considered as medium-low technology sectors.

Table 7: Estimates for ETS sectors

	CO2	NOx	NMVOC	SOx	CO
△ VA/E frontier	0.0473*	0.0841***	0.0696	0.0989*	0.0716*
	(0.0284)	(0.0294)	(0.0484)	(0.0550)	(0.0428)
VA/E gap (t-1)	0.00557	0.00738	-0.00358	0.0444	-0.000285
	(0.00684)	(0.00804)	(0.00678)	(0.0309)	(0.00927)
△ energy prices	-0.247	-0.249	-0.462	0.892	0.565
	(0.186)	(0.306)	(0.854)	(0.738)	(0.910)
△ energy prices × VA/E gap (t-1)	0.101	0.240*	0.142	-0.191	-0.0765
	(0.0979)	(0.145)	(0.181)	(0.191)	(0.207)
△ TFP	0.924***	1.016***	1.025***	0.854***	0.873***
	(0.0329)	(0.0488)	(0.0831)	(0.101)	(0.108)
△ Frontier TFP	-0.0133	-0.0317	-0.0346	-0.0352	0.0834
	(0.0370)	(0.0541)	(0.106)	(0.117)	(0.122)
TFP gap (t-1)	-0.0149	-0.00692	-0.000658	-0.0383	0.0337
	(0.0130)	(0.0186)	(0.0316)	(0.0477)	(0.0418)
Constant	-0.0294	-0.0171	0.0439	-0.0645	0.00252
	(0.0199)	(0.0285)	(0.0418)	(0.103)	(0.0552)
F	37.22***	26.86***	12.97***	8.352***	5.820***
R squared	0.589	0.450	0.249	0.153	0.166
Year dummies (F)	2.672***	1.997**	1.325	2.428***	0.734
Sector dummies (F)	0.575	0.734	1.907	4.155***	4.112***
Country dummies (F)	2.047**	2.760***	2.466***	1.558*	1.519
Ramesey o.v. test (F)	1.175	1.144	0.504	8.084***	3.851***
N	712	712	712	712	712

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

tude of estimated coefficients with baseline estimates, the effect of emission efficiency growth at the frontier is similar to the full sample while the gap in terms of emission efficiency has a systematically lower effect, with very small and always insignificant coefficients. This results underlines a weak tendency to converge of ETS sectors. With the only exception of SOx estimates, the effect of energy prices (both direct and conditional on the emission efficiency gap) is much lower than the effect for the full sample. This weak responsiveness to price signals is very relevant in the choice of an effective and efficient policy tool to limit air emissions. A cap and trade system such as the European ETS seems more appropriate than a tax on emissions to effectively reduce air emissions of these sectors. By setting quantitative aggregate targets, cap and trade systems ensure the effectiveness of the policy leaving some uncertainty on the overall cost of compliance. On the contrary, non-ETS sectors (table 8) are characterized by more robust results. Sectors characterized with a relevant gap from the emission efficiency frontier grow significantly faster for all emissions while being far from the productivity frontier affects negatively and significantly emission efficiency growth. Moreover, non-ETS sectors tend to be more responsive to energy prices than ETS sectors.

Table 8: Estimates for non-ETS sectors

	CO2	NOx	NMVOC	SOx	CO
△ VA/E frontier	0.0421**	0.0614***	0.0396	0.0932***	0.0795***
	(0.0177)	(0.0199)	(0.0254)	(0.0252)	(0.0281)
VA/E gap (t-1)	0.0250***	0.0351***	0.00827	0.0646***	0.0396***
	(0.00669)	(0.00819)	(0.00613)	(0.0100)	(0.00856)
△ energy prices	0.152	0.223	1.603***	0.119	1.561***
	(0.164)	(0.226)	(0.525)	(0.429)	(0.437)
△ energy prices ×	0.102	0.155*	-0.196**	0.179	-0.278**
VA/E gap (t-1)	(0.0838)	(0.0928)	(0.0931)	(0.110)	(0.125)
△ TFP	0.979***	0.982***	0.870***	0.906***	1.056***
	(0.0521)	(0.0565)	(0.0744)	(0.126)	(0.0954)
△ Frontier TFP	-0.0591**	-0.0809**	-0.0637	-0.0655	-0.0874*
	(0.0295)	(0.0391)	(0.0418)	(0.0894)	(0.0482)
TFP gap (t-1)	-0.0899***	-0.0844***	-0.0661***	-0.178***	-0.113***
	(0.0192)	(0.0201)	(0.0246)	(0.0540)	(0.0297)
Constant	-0.0298	-0.0373	-0.0562	0.0574	-0.0345
	(0.0296)	(0.0320)	(0.0401)	(0.0592)	(0.0454)
F	18.98***	17.12***	12.34***	5.479***	12.75***
R squared	0.259	0.208	0.156	0.0963	0.180
Year dummies (F)	7.303***	5.175***	2.649***	6.375***	3.209***
Sector dummies (F)	1.038	0.758	2.562***	1.401	0.848
Country dummies (F)	4.773***	9.982***	5.400***	2.115**	9.571***
Ramesey o.v. test (F)	0.829	0.210	168.2***	9.533***	123.9***
N	2403	2403	2403	2403	2403

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The comparison of medium-high technology sectors (table 9) with medium-low technology sectors (table 10) allows to underline systematic differences in the determinants of emission efficiency growth. Growth at the frontier is more relevant for medium-low than for medium-high technology sectors (with the only exception of NMVOC emissions). Medium-low technology sectors seem to rely on environmentally efficient technologies developed abroad to a greater extent than medium-high technology sectors, for which the development of domestic technologies seems to prevail. Despite that, convergence is faster in medium-high technology sectors due to a stronger positive effect of the emission efficiency gap. On the contrary, productivity gap (in terms of TFP) is more detrimental for emission efficiency growth in medium-high sectors than in medium-low technology sectors, highlighting a stricter link between economic and environmental performance in medium-high technology sectors. Finally, evidence for energy prices is more mixed, with emission-specific differences (in magnitude and significance but not in signs) between medium-high and medium-low technology sectors.

To conclude, I report estimates including the lag of emission efficiency growth to account for dynamic adjustments (table 11). Results

Table 9: Estimates for medium-high technology sectors

	CO2	NOx	NMVOC	SOx	CO
△ VA/E frontier	0.0324	0.0295	0.0416	0.118***	0.0916**
	(0.0249)	(0.0284)	(0.0309)	(0.0445)	(0.0439)
VA/E gap (t-1)	0.0320**	0.0439***	0.00339	0.101***	0.0413***
	(0.0148)	(0.0158)	(0.00827)	(0.0216)	(0.0121)
△ energy prices	0.193	0.00279	1.388**	0.413	1.548***
	(0.241)	(0.264)	(0.639)	(0.547)	(0.546)
△ energy prices ×	0.0909	0.262*	-0.117	0.150	-0.193
VA/E gap (t-1)	(0.129)	(0.137)	(0.115)	(0.150)	(0.173)
△ TFP	1.094***	1.037***	0.879***	0.998***	1.175***
	(0.0741)	(0.0830)	(0.118)	(0.159)	(0.159)
△ Frontier TFP	-0.107	-0.0646	-0.0769	-0.237	-0.190
	(0.0776)	(0.0885)	(0.0954)	(0.190)	(0.116)
TFP gap (t-1)	-0.118***	-0.128***	-0.0989***	-0.296***	-0.164***
	(0.0330)	(0.0326)	(0.0364)	(0.0800)	(0.0451)
Constant	0.0211	0.0541	0.0682	0.184*	-0.00357
	(0.0570)	(0.0627)	(0.0604)	(0.0977)	(0.0832)
F	14.02***	11.62***	9.986***	5.633***	8.226***
R squared	0.271	0.226	0.169	0.161	0.191
Year dummies (F)	3.064***	2.296***	1.522	7.473***	2.448***
Sector dummies (F)	1.343	0.856	0.598	1.221	0.783
Country dummies (F)	2.728***	4.978***	4.226***	1.661*	4.328***
Ramesey o.v. test (F)	0.313	3.663**	11.97***	15.67***	12.54***
N	1091	1091	1091	1091	1091

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

were obtained by applying the system GMM estimator (Blundell and Bond, 1998). Results were very similar in magnitude and significance to baseline estimates. The coefficient for the lagged dependent variable is significantly negative for CO2, NOx and SOx and insignificant for NMVOC (positive) and CO (negative). This means that the dynamic adjustment of emission efficiency growth is not smooth and, on average, occurs by means of accelerations followed by slowdowns. While most covariates show very similar effect to baseline static estimates, a remarkable systematic difference regards the estimated coefficients for the emission efficiency gap that increase substantially for all emissions. Finally, it is worth discussing some specification test on system GMM estimates. The Arellano-Bond test for second order autocorrelation of residuals accepts the null hypothesis of absence of second order autocorrelation for CO2, NMVOC and CO while the null hypothesis cannot be accepted for NOx and SOx. The Hansen test of joint validity of instruments rejects the null hypothesis of weak instruments at the 1% level of significance for SOx emissions only (the null hypothesis is rejected at the 10% level of significance for NOx and CO), with both exogenous independent variables (IV) and lags of the dependent variable (GMM)

Table 10: Estimates for medium-low technology sectors

	CO2	NOx	NMVOC	SOx	CO
△ VA/E frontier	0.0557***	0.0921***	0.0491*	0.0822***	0.0725**
	(0.0212)	(0.0259)	(0.0290)	(0.0282)	(0.0302)
VA/E gap (t-1)	0.0158***	0.0230***	0.00757	0.0518***	0.0240***
	(0.00465)	(0.00662)	(0.00604)	(0.0105)	(0.00811)
△ energy prices	0.0170	0.174	0.894	0.236	1.277**
	(0.170)	(0.246)	(0.601)	(0.490)	(0.534)
△ energy prices ×	0.0981	0.124	-0.113	0.0784	-0.252*
VA/E gap (t-1)	(0.0765)	(0.0959)	(0.111)	(0.125)	(0.131)
△ TFP	0.867***	0.951***	0.902***	0.766***	0.881***
	(0.0432)	(0.0465)	(0.0576)	(0.125)	(0.0683)
△ Frontier TFP	-0.0156	-0.0425	-0.0348	0.0562	-0.0102
	(0.0228)	(0.0336)	(0.0418)	(0.0842)	(0.0464)
TFP gap (t-1)	-0.0345***	-0.0210	-0.00934	-0.0408	-0.00652
	(0.0126)	(0.0169)	(0.0228)	(0.0481)	(0.0242)
Constant	-0.0568***	-0.0628***	-0.0579*	-0.0122	-0.0562
	(0.0208)	(0.0224)	(0.0326)	(0.0547)	(0.0358)
F	22.70***	22.68***	15.20***	5.031***	11.40***
R squared	0.325	0.251	0.160	0.0856	0.169
Year dummies (F)	8.481***	5.514***	2.638***	3.409***	2.177**
Sector dummies (F)	0.804	0.696	2.738***	4.054***	2.492***
Country dummies (F)	4.320***	8.686***	3.831***	2.053**	8.523***
Ramesey o.v. test (F)	2.726**	0.135	83.11***	5.441***	185.3***
N	2024	2024	2024	2024	2024

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

being invalid instruments. Exogenous independent variables are invalid instruments also for NOx emissions. Finally, the difference-in-Hansen test generally shows that instruments were exogenous, the only remarkable exception being exogenous independent variables for SOx and CO emissions (the null hypothesis of exogeneity is rejected at the 5% level of significance).

2.5 Concluding remarks

This chapter investigates the dynamics of sectoral emission efficiency in a selection of European countries. International diffusion of more efficient environmental technologies, distance from the technological frontier, energy prices and economic productivity patterns are found to be important drivers of emission efficiency growth in manufacturing sectors.

Results highlight the importance of the diffusion of more environmentally efficient production technologies from leader countries to laggards. However, the channels through which the diffusion occurs are not investigated directly. The convergence of emission efficiency towards

Table 11: Estimates including lagged dependent variable (system GMM)

	CO2	NOx	NMVOC	SOx	CO
△ VA/E (t-1)	-0.0805***	-0.177***	0.0408	-0.184***	-0.0342
	(0.0252)	(0.0305)	(0.0467)	(0.0391)	(0.0315)
△ VA/E frontier	0.0398**	0.0674***	0.0483**	0.106***	0.0838***
	(0.0168)	(0.0181)	(0.0217)	(0.0290)	(0.0219)
VA/E gap (t-1)	0.0249***	0.0403***	0.0168***	0.102***	0.0548***
	(0.00504)	(0.00564)	(0.00567)	(0.0124)	(0.00810)
△ energy prices	0.0462	0.157	1.513***	0.376	1.380***
	(0.145)	(0.199)	(0.512)	(0.392)	(0.396)
△ energy prices x	0.102	0.173**	-0.167*	0.107	-0.234**
VA/E gap (t-1)	(0.0657)	(0.0751)	(0.0904)	(0.0988)	(0.100)
△ TFP	0.940***	0.971***	0.911***	0.872***	0.978***
	(0.0402)	(0.0405)	(0.0602)	(0.0992)	(0.0726)
△ TFP frontier	-0.0398	-0.0725**	-0.00378	-0.0375	-0.0576
	(0.0252)	(0.0361)	(0.0427)	(0.0793)	(0.0506)
TFP gap (t-1)	-0.0575***	-0.0747***	-0.0380***	-0.139***	-0.0836***
	(0.00823)	(0.00965)	(0.0101)	(0.0337)	(0.0147)
Constant	0.0672***	0.0636***	0.0393*	-0.121***	0.00209
	(0.0136)	(0.0146)	(0.0238)	(0.0425)	(0.0205)
Chi sq	1006.5***	811.8***	342.6***	312.3***	265.6***
AR(1)	-5.431***	-9.102***	-5.347***	-7.789***	-7.222***
AR(2)	1.300	-2.228**	1.064	-2.421**	-0.860
# instruments	62	55	73	38	38
Sargan test	232.9***	474.1***	479.2***	200.6***	174.9***
Hansen test	43.25	54.92*	65.34	39.61***	32.37*
Hansen test (GMM)	38.84	44.03**	50.65	22.81***	14.93*
Hansen test (diff - GMM)	4.41	10.89	14.69	16.8*	17.44*
Hansen test (IV)	27.94	37.37***	45.47	11.13***	3.93
Hansen test (diff - IV)	15.3	17.55	19.87	28.48**	28.44**
N	2848	2848	2848	2848	2848

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

the frontier is faster for countries and sectors with a greater efficiency gap, probably showing evidence of increasing marginal costs of abatement. Energy prices dynamics has a positive effect on emission efficiency and the effect is decreasing in the emission efficiency gap for CO and NMVOC emission efficiency growth while it is significant only for laggard sectors (and increasing in the emission efficiency gap) for CO2, NOx and SOx emission efficiency growth. Moreover, there is a very robust evidence of complementarity between emission efficiency and economic productivity (here measured with TFP). Finally, the homogeneity of estimates across different types of air emissions is quite surprising, especially in the presence of moderate pairwise correlation between emission efficiency growth rates[24].

Based on the evidence discussed in this chapter concerning the international diffusion of emission efficiency, further research is needed

[24] Pairwise correlation between emission efficiency growth rates is greater than 50 percent in just three cases (CO2-NOx, 70 percent; CO-NMVOC, 60.21 percent; NOx-CO, 60.16 percent) and is lower than 20 percent in one case (19.59 percent for NMVOC-SOx).

to investigate the way through which sectors in laggard countries take advantage of emission efficiency improvements occurring in the frontier countries. As discussed in the introduction, the diffusion of environmental technologies leading to improvement in emission efficiency may be triggered by a variety of factors. The assessment of the contribution to the diffusion of environmental technologies of these factors is crucial to identifying the optimal policy mix. Finally, it is worth combining patterns of international diffusion with patterns of cross-sectoral diffusion within the same country (Corradini et al, 2011) in a comprehensive framework to obtain a more complete representation of the diffusion of emission efficient technologies.

Air emission features

Emissions differ substantially as regards the 'external cost' they produce. Carbon dioxide emissions have no direct effect on health and on local communities whereas they contribute to the greenhouse effect and global climate change. On the contrary, other emissions (NOx, SOx, NMVOC and CO) have serious effects on health and damage the environment at the local level through acidification (NOx and SOx), ozone depletion (NOx), eutrophication (NOx) and tropospheric ozone formation (CO and NOx).

These difference resulted in different timing and characteristics of national or supra-national regulations. Pollutant emissions have been regulated at the European level since the mid 80s through a series of Directives which have eased the harmonization of national policies. Among others, consider the following directives aimed at regulating pollution. The Sulphur Dioxide Air Pollution Directive, approved in 1980 (1980/779/EEC), aimed at reducing SOx emissions while the Nitrogen Dioxide Air Pollution Directive approved in 1985 (1985/203/EEC) focused on the reduction of NOx emissions. They were replaced by the First Daughter Directive 'Sulphur Dioxide, Nitrogen Dioxide and Oxides of Nitrogen, Particulate Matter and Lead in Ambient Air' in 1999 (99/30/EC) broadening the scope of pollutant reductions to SOx and other local pollutants. The Fuel Quality Directive introduced in 1998 (98/70/EC), revised in 2003 (2003/17/EC) and in 2009 (2009/30/EC), sets specific requirements for the quality of fuels in order to reduce emissions of pollutant substances. The NEC (National Emission Ceilings) Directive (2001/81/EC), approved by the European Commission in 2001, sets legally binding limits to national emissions of NOx, SO2, NMVOC and ammonia. Finally, a broader programme to consider air pollution emissions in a comprehensive way was launched by the European Commission in 2005 (Clean Air For Europe programme CAFE).

On the contrary, regulatory efforts explicitly aimed at reducing carbon dioxide emissions were less effective. No relevant policy was introduced before the approval of the Kyoto Protocol (1997) and, even after the protocol started being legally binding (2001), no real action was taken before the introduction of the Emission Trading Scheme (in its pilot phase) in 2005 and the '20-20-20' strategy proposed in 2007.

Chapter 3

Linking NAMEA and I-O for 'Consumption vs. Production Perspective' Analyses

3.1 Introduction[1]

3.1.1 The Background and the Rationale within an Economic-Policy Perspective

The integration of the National Accounting Matrix including Environmental Accounts (NAMEA) and input-output (I-O) tables (usually referred to as Environmental Extended Input-Output Analysis - EEIOA based on National Accounting Matrices including Environmental Accounts - NAMEA) is a challenging but promising way to analyse the factors behind income-environment relationships in international settings, with sound overlapping with research fields such as IPAT (Impact Pop-

[1]This chapter is the reproduction of the article "Linking NAMEA and Input Output for 'Consumption vs. Production Perspective' Analyses - Evidence on Emission Efficiency and Aggregation Biases using the Italian and Spanish Environmental Accounts" (with Massimiliano Mazzanti and Anna Montini), 2012, *Ecological Economics*, 74:71-84.

ulation Affluence Technology) based analysis, environmental Kuznets curves (EKC), trade-related and globalisation-dependant environmental impacts and 'sustainable growth and resource productivity' analysis (Bleischwitz et al, 2009; Cole, 2004; Copeland and Taylor, 2004; Frankel and Rose, 2005; Marin and Mazzanti, in press). More specifically, it can be used to disentangle production and consumption perspectives on sustainability through the detailed sector-based information provided by the two frameworks.

The sector based perspective is crucial in the current analyses of economic-environmental dynamics since it may shed light on structural phenomena that neither macro nor microeconomic settings can provide due to opposite limitations (too large, too narrow focus). The meso level is capable of unveiling what the changing composition (e.g. industry mix, increasing share of services in advanced economies) and new specializations of our economies mean in economic and environmental terms. New sources of competitiveness and their environmental impacts are possibly analysed in a way that also provides relevant food for thought to environmental and industrial policies, that are in this perspective necessarily integrated.

National and international sources of environmental effects can be ascertained in strict connection with streams of literature such as the 'ecological footprint' kind of analysis and decomposition analyses, that are probably the closer fields. The production and exploitation of EE-IOA and NAMEA are also heavily embedded in the wide research and policy realm that deals with 'sustainable consumption and production (SCP)' issues (Eurostat, 2001), a key pillar of current and future EU policy efforts. The analysis of sector specificities, direct and indirect emissions, the role of international trade are ways to make concrete and operational the discussion on the Green economy. EE-IOA links to another quite concrete issue: economic and resource productivity dynamics (ETC/SCP (2011a) and OECD (2011b), which among other findings highlights the increasing role of trade and that resource productivity has improved less than labour productivity, a signal of potential un-sustainability; Mazzanti and

Zoboli (2009))[2], insofar the changing industry mix (will a service based economy be associated with higher resource efficiency? The persistence of manufacturing in some countries as Germany, is a key issue[3].) and the environmental impacts embodied in trade contribute to the overall resource productivity performance of our economies, which is a first signal of sustainability (complementary to capital based view of sustainability such as the Genuine saving approach). Resource productivity and its sub-themes are manageable from both analytical and political points of view. A 'Resource Efficiency Roadmap' is currently under development by the European Commission. SCP is the main operational framework, where EE-IOA plays its role.

It is worth noting in the discussion of the various EU strategies, that EUROSTAT is aimed at releasing a full 2000-2006 NAMEA for EU27 that will support EU SCP policy efforts, and for the first time released in April 2011 an indirect emission dataset that should take into account the 'consumption perspective', as a complement to the production view offered by original NAMEA. The ongoing status of the project, which is a key pillar of EU data production, is summarized in Eurostat (2011).

A comparison of the production vs. consumption perspective can have important policy implications. Substantially, the production perspective takes the view of a country producer responsibility considering direct emissions in a country due to domestic production processes that generate pressures and impacts within the country. On the other hand, the consumption perspective (or country final user responsibility as appropriately suggested by Serrano and Dietzenbacher (2010)) investigates

[2] We also refer the reader to the web site of the EU topic centre on Sustainable Consumption and production, http://scp.eionet.europa.eu/.

[3] Another key issue that gives value to EE-IOA analyses is the observed increasing interdependency in production: the intensity of intermediate inputs in the production of total output has increased following service-manufacturing stricter inter relationships. I-O based analyses and sector specific investigations are motivated by those facts (European Commission, 2009). Recent evidence seems to suggest that resource efficiency trends are driven more by technology than composition effects. This is in part dependant on the gloomier performance of services when indirect emissions are accounted for, that this chapter also discusses, and in part relates to the fact that increasing inter-industry linkages are part of the technological dynamic (e.g. outsourcing of production, vertical disintegration, etc.).

the impacts due to domestic consumption (all domestic final demand and not exclusively from household consumption) regardless where they have been produced. So the two perspectives take into different consideration the direct effects of the needs of society when producing the products needed in a particular domestic territory (regardless if domestic consumers or consumers abroad - exports - caused the emissions) and the country's responsibility for emissions generated globally (including the embodied emissions in imports) in order to satisfy its domestic final demand.

Traditionally, environmental policy has focused mainly on production activities as sources of impacts and the actor to be targeted by legislation and regulation (examples are carbon taxes, emission trading). Looking at the role of final consumption for vertically integrated domestic and international impacts can push policy attention towards the possible role of the consumer as an actor of environmental policies, together with the international responsibility for spillover of impacts abroad. In that direction, policies on the supply side may find complements in environmental policies that target consumption (labelling, but also green consumption taxes, taxes that correlate with the embodied emissions or materials in the production of the good). The revenue accruing from ecological taxes can also find a possible use in the funding of 'product innovation' aimed at resource efficiency.

A key issue is the modelling of the technology associated with imported goods (produced abroad by the stimulus of domestic consumption), which is tricky in practice given the scarcity of data at that level of detail and at sector level. Given the technology, (net, accounting for export and import, see Levinson (2010)) trade-embodied pollution arises as a structural phenomenon of the globalised economy, depending on the systematic difference between the composition of domestic and foreign production. These increasing differences may be responsible of a 'burden shifting' in terms of environmental impacts relocated abroad (then imported, thus appearing in a consumption view of sustainability). A burden that can depend upon differences in policy stringency (the pollution haven hypothesis), but also on structural facts of changing specialization and industry mix. Structural imbalances may appear in globalised - difficult to regulate - markets, that risk of being not sustainable if we take a worldwide perspective. Advanced countries environmental performances in the production side (e.g. EKC) may appear better than what are in reality. Systematic differences can fast change given that the production specialization of a country is usually more marked and in the development if compared to the 'consumption specialization' of a country (a relative long run phenomenon in terms of development). ETC/SCP (2011b) discusses production and consumption long term indicators with reference to SCP[4], and presents some answers to these 35 policy questions through assessing trends in 39 relevant European indicators.

3.1.2 The Evolution of EE-IOA Studies and New Research Targets

We can affirm that sector-based input-output datasets existing for EU countries offer the possibility of highlighting how emissions are indirectly associated with production. NAMEA-type tables are datasets with coefficients on emission per output that can thus be matched with I-O tables for useful integration. Integration aims at calculating economic-environmental performances by sector by including the role of trade. In other words, it aims to test the hypothesis that given different relative emission efficiency, the structure of imports and exports matters.

From a general and methodological point of view, the integration of NAMEA accounting and input-output (I-O) tables touches upon ecological/environmental economics and industrial ecology frameworks. Due

[4]Thus, this means that in a dynamic setting, consumer behaviour is changing slowly in terms of embodied environmental efficiency, compared with domestic production, thus possibly creating a net demand of pollution abroad, through import from emerging countries. Although consumption structure and behaviour can be less sensitive to environmental policies than production, there can be room for addressing consumers and their behaviour to contribute to higher efficiency in terms of vertically integrated environmental impacts. The EU strategies on Sustainable Consumption and Production pave the way to this policy direction, and analyses based on Environmental Extended Input-Output Analysis, addressing the differences between the two perspectives, can clarify the needs and implications of these policies.

to the striking increase of related works in such realms, the brief survey we provide in the next paragraph aims to give insights into recent developments and offer stimulus for future analyses rather than offering full coverage. It is worth noting that, very recently, there has been increasing interest in these environmental issues in the 'Input-Output world'. A boom of papers on environmental extended I-O was reached in 2009, that witnessed a peak (Hoekstra, 2010), with a total amount of 360 papers, from 1969 to 2010. A related field of analyses with a great relevance in the I-O arena is structural decomposition analysis (SDA), one of the most effective and widely applied tools for investigating the mechanisms influencing energy consumption and emissions and their environmental side-effects (Mazzanti and Montini, 2010). Many studies address industry. Nevertheless, services are also relevant: they are less energy intensive but present lower technological contents and can indirectly contribute to strong environmental impacts (we note the NAMEA-based disentangled analyses in Marin and Mazzanti (in press), who present industry vs. services assessments for Italy). Alcántara and Padilla (2009) analyse CO2 emissions for Spain using I-O (year 2000).

Trade is the key factor in recent extended I-O and NAMEA works that aim to deal with SCP contents[5]. We recall that the main aim is to assess direct and indirect environmental effects by attributing their relative weights to national consumption and to exports in the explanation of a country environmental performance. Currently, main efforts aim to move away from the Domestic Technology Assumption (DTA) that says that imported goods use the same technology (in terms of structure of intermediate inputs and environmental efficiency) as goods produced domestically.

A very recent example is Arto et al (2003). They show that Spain is a net emission exporter and consequently, its consumer responsibility in emissions is higher than its producer responsibility. The difference

[5]Some main streams of research can be outlined: I-O models accounting for trade and embodied emissions (through energy accounts); global multi-region input-output (MRIO) model; extension for eco-footprint analysis; comparing physical trade balance (PTB) and pollution trade balance (UTB) associated with fossil use; analysing pollution terms of trade, pollution haven tests; analysing I-O tables linked with satellite accounts.

between both types of responsibility increases by applying the physical DTA. This is substantially due to the fact that the monetary DTA estimates less embodied emissions in imports from non-Annex I countries than the physical DTA[6].

A study that brings together various frameworks highlighting flexibility of methods and usefulness of integrated use is certainly Moll et al (2007). The work shows that, according to different sectors and countries, the domestic production patterns and associated direct domestic environmental pressures are rather different. Electricity, gas and hot water production, agriculture and transport and communication services cause the majority of environmental pressures. Direct pressures from private households (mainly for heating and private transport) constitute another important source. With regard to international factors, it can be seen that a second determinant for cross-country differences in domestic direct pressures is the role of exports. When it comes to consumption and investment patterns, Moll et al (2007) show that cross-country differences are far less pronounced than production patterns. Analyses focusing on environmental impacts of consumption (by categories) are also found in Huppes et al (2005): food, heating and transport emerge as core impacting aggregation[7]. We also note the extensive IPTS 'EIPRO' (IPTS, 2006) report. In general, it is the satisfaction and organization of basic needs, i.e. eating, housing and mobility, that is responsible for the majority of production-cycle-wide environmental pressures.

In this chapter we attempt to provide complementary evidence with respect to the mentioned works. The main purpose of the current analysis is to aggregate our original Italian and Spanish data according to relevant aggregations used in other studies and to compare our bench-

[6] The physical DTA refers to the use of imports in physical quantities and using, for imports, the same physical environmental coefficient (emissions per kg of import) as domestic physical environmental coefficients (emissions per kg of domestic output). This assumes that, although of different quality (value per physical unit), the emission content of goods is closely correlated to its weight and less correlated to its value.

[7] Automobile driving and related maintenance activities are by far the largest contributing products to total environmental impacts by consumption in the EU25. However, by summing several animal-based foods (meat, meat products, poultry, dairy products), animal food products would become dominant. At the aggregate level of 12 consumption domains, food already comes up as the largest contributor to environmental problems.

mark estimates (i.e. the estimates arising from the most disaggregated model) with the estimates arising from less detailed aggregations. More specifically, our benchmark consists of a disaggregation of 50 commodities[8]. This benchmark will be compared with the subsection NACE rev. 1.1 level (accounting for 30 sectors) and with an aggregation of 16 sectors roughly corresponding to previous studies based on OECD/IEA data.

We provide new evidence through an application that focus and compare Italy and Spain, two countries with an historical experience of NAMEA and I-O table's generation, which is witnessed in the many papers published by Ecological Economics and collected in dedicated books (Costantini et al, 2011) in recent years. The choice of using Italy and Spain is nevertheless motivated by various specific facts.

From a data quality and availability point of view, we selected two of the top experiences in the EU. We observe that our projected began well before the publication and release of the first result of EUROSTAT project (summarized in the publication by Eurostat (2011), "Creating consolidated and aggregated EU27 Supply, Use and Input-output Tables, adding environmental extensions (air emissions), and conducting Leontief-type modelling to approximate carbon and other 'footprints' of EU27 consumption for 2000 to 2006", which attempts to improve the data availability situation in the EU towards a more institutionalised and homogeneous generation of data on I-O supply and use tables and air emissions, and their integration to account for embodied emissions in final demand. NAMEA data generation had been more scattered before 2011. Even after recent improvements, in the cited report itself the data quality assessment signals that a few countries present excellent status over 2000-2006. Italy and Spain are among those few, and have historically allowed many analyses, including panel econometric studies (Mazzanti and Zoboli, 2009). Germany is another country with excellent quality in all years. As example, France and the UK are countries that had not presented and which still present not excellent situations (see Eurostat (2011), Detailed tables on air emissions 2006). Germany posed problems

[8]This level of disaggregation corresponds roughly to the 2-digit NACE rev. 1.1 classification. For more details, refer to Section 3.2.

in terms of commensurability of sector aggregation, which lead us to end up with Italy-Spain comparisons in this exercise (more details on this fact are available upon request). Extensions to other countries are suggested for future research on the shoulders of the fast improving conditions of data availability (as a reference the I-O tables and NAMEA availability, including pilot projects for various countries, is summarized and available at the EUROSTAT website). The assessment of the aggregation bias in Extended Input-Output analysis is crucial to achieve robust analysis of embodied emissions in final demand (and import-export), which is a key pillar of the EU strategy on sustainable consumption and production. The methodological clarification of the bias is important to reduce the overall bias of such analyses, which is on the other hand depending on sector data commensurability and on (the relaxation of) the Domestic Technology assumption.

From an economic point of view, those are two Southern EU countries which, notwithstanding differences in their industrial composition, share on the other hand similar features in terms of level of economic development and GDP per capita (Mazzanti and Musolesi (2010) present the case of strong differences between northern and southern EU countries regarding income-green house gases structural relationships in a EKC framework). Italy is relatively more industrial and export oriented. This main significant difference can be useful to compare in the end the results and provide explanation of eventual non homogeneity.

The chapter is organized as follows. In Section 2, we review the specific empirical literature on the estimation of environmental pressures induced by domestic consumption and domestic production activities, with a specific focus on environmentally extended input-output methodologies and related potential biases. In Section 3, we describe our methodological approach, with a particular focus on the role of aggregation bias in environmentally extended input-output analyses, and our data source, stressing the value added of merging NAMEA emissions with the input-output framework. In Section 4, we report and comment on our main results. Section 5 concludes.

3.2 Methodological Issues in the Relevant Literature

Empirical analysis with an extension of the use of the statistical information derived from environmental accounts and the input-output tables requires several considerations to be made. The main aim of this chapter is linked to the investigation of the so-called aggregation bias. As suggested by Lenzen (2011), environmental I-O analyses of environmental issues are often plagued by the fact that environmental and I-O data exist in different classifications.

A recurring problem in EE-IOA is that input-output accounts and environmental statistics used as environmental extensions are often not compiled by the same statistical agency and therefore often differ with respect to the classification of economic sectors and other definitions. In these cases, analysts have to carry out data collection and harmonization procedures in order to integrate both accounts. What can happen is that: (i) environmentally sensitive sectors are sometimes more aggregated in the economic I-O database than the environmental dataset because monetary I-O tables are compiled with no environmental implications in mind; and (ii) I-O data are disaggregated into more sectors than environmental satellite data, especially for the services sectors (Lenzen, 2011).

There are two basic alternatives for dealing with such a misalignment: either environmental data have to be aggregated into the I-O classification (but some environmental sensitive data will lose their peculiarities) or I-O data have to be disaggregated based on fragmentary information (with several assumptions).

By keeping this in mind, the aggregation bias is likely to severely affect the construction of environmentally extended Multi-Region Input-output (EE-MRIO) analysis, as recently suggested by Su et al (2010) and Lenzen (2011), as well as environmentally extended Single Region Input-output accounts with specific assumption regarding the technology used (embodied in international trade, specifically those in the import data).

As we explain below, the DTA (Domestic Technology Assumption) relies on the consideration that all imported commodities are produced with the same mix of intermediate inputs (in monetary terms and as indicated by the intermediate flows in the input-output table) and with the same environmental efficiency (in terms of emissions per monetary unit of output) as domestic commodities.

Some authors (including Arto et al (2003), Peters (2008), Serrano and Dietzenbacher (2010), Turner et al (2007)) suggest moving away from the DTA because they consider it too simplistic, but they recognize that, generally, the DTA produces better estimates than ignoring imports altogether. Ideally, full information on bilateral trade plus corresponding NAMEA data by country is equivalent to analysing trade of impacts at country-by-country differentiated coefficients. However, it requires a wide and often unavailable range of data. A possibility for dealing with the latter is to include only the most important trade partners in terms of emissions embodied in imports and this, as suggested by Andrew et al (2009). For the emissions embodied in imports, Andrew et al (2009) find that the unidirectional trade model gives a good approximation to the full MRIO model when the number of regions in the model is small. Moreover, the assumption that imports are produced with DTA in an MRIO model can introduce significant errors and requires careful validation before results are used.

If we re-examine the issue of aggregation bias, the studies that have analysed the CO2 emissions embodied in international trade have also been carried out by using an input-output framework at a specific level of sector aggregation. Generally, the choice has been made to a large extent according to economic and energy data availability or, similarly, economic and environmental data availability. A finding in Su et al (2010) is that levels of around 40 sectors appear to be sufficient to capture the overall share of emissions embodied in a country's exports.

The issues related to aggregation bias and a possible DTA obviously affect the consumption[9] perspective when looking at the corresponding

[9]The consumption based emissions are computed using domestic production based emissions minus the emissions embodied in exports (demanded by final users abroad) plus the embodied emissions in imports (demanded by domestic final users) assuming that the Rest of the World has the same technology as the country analysed.

emissions. As suggested in the Introduction, the focus of the EU policy area on Sustainable Consumption and Production forces researchers to consider new tools of analysis and one of them is the EE-IOA based on NAMEA data. The notion of 'responsibility' (either for the consumer or the producer) allows some considerations.

As suggested by Gallego and Lenzen (2005), there is a sort of domination of producer-centric representation to view the environmental or social impacts of industrial production. When thinking about environmental impacts, crucial questions arise such as who is responsible for what? Moreover, the kind of pollutant considered influences policy implications when looking at the ratio between consumption-based emissions (C) and producer-based emissions (P). If we consider global pollutants, such as CO2, and C is bigger than P, the country responsibility is bigger than that reported by the official statistics. If we consider local pollutants, and C is bigger than P, the country would be displacing environmental costs to other territories.

Gallego and Lenzen (2005) propose a method of re-tracing the flow of past inter-industrial transactions to allocate responsibility for production impacts consistently among all agents such as consumer, producers, workers and investors. According to them, the input-output analysis can be used as a descriptive tool to re-trace the flow of past transactions and examine ex-post how, for example, inputs of resources or outputs of pollution were associated with these transactions.

Serrano and Dietzenbacher (2010) define two ways to evaluate the international responsibility of emissions generated by one country - in their analysis they consider Spain in 1995 and 2000 and nine gases that were shown to be equivalent: the trade emission balance (as the difference between the emissions embodied in a country's exports and imports) and the responsibility emission balance (as the difference between the responsibility of one country as a producer and its responsibility as a 'consumer').

On the basis of the highlighted and hotter methodological issues, we present below our methodological framework.

3.3 Our Methodology and Data

In this section,we outline the main features of the domestic technology assumption (DTA henceforth) and we summarize the main issues related to the assessment of the aggregation bias in input-output analysis including NAMEA data.

3.3.1 Domestic Technology Assumption

The hypothesis behind the domestic technology assumption is that the imported commodities (either as intermediate inputs or final consumption) are produced with the same mix of intermediate inputs (in monetary terms) and with the same environmental efficiency (in terms of emissions per monetary unit of output) as domestic commodities.

Serrano and Dietzenbacher (2010) formally describe how and under which conditions an environmental extended multi-regional input-output model accounting for worldwide induced emissions could be reduced to a model using only domestic data with an explicit domestic technology assumption. In addition to assumptions on technology (i.e. the structure of intermediate inputs described by the input-output matrix) and on the vector of emission coefficients, the export of the country on which the analysis is focused should represent a negligible share of world output.

Another requirement, related to the validity of the domestic technology as a proxy of world technology, is that the country produces domestically at least part of all the commodities it consumes as intermediate inputs or final products. For example, this requirement is not fulfilled when a country has no particular raw materials in its soil or subsoil (oil, coal, gas, minerals, metals, etc.) and it is completely dependent on importing these commodities. As a result, the technology for the extracting industries (section C of NACE 1.1) in the input-output tables is biased towards secondary activities within the sector (e.g. basic transformation of raw materials) and it does not describe the main activity (i.e. extraction) properly. This problem is particularly relevant in environmentally extended input-output analyses in which extracting sectors are, in general, among the most polluting industries.

Although the DTA cannot be used to interpret the results as 'actual worldwide emissions induced by domestic final demand', it gives information on the potential emissions arising because of domestic final demand if the country has produced domestically the necessary final and intermediate goods (that is, using domestic technology). Estimates using the DTA, if interpreted properly, are therefore a particularly important indicator of consumer responsibility because of its low requirement for data, the possibility of replicating its results and the straightforward and clear hypothesis behind its implementation. For this reason, we claim that estimates based on the DTA should be used as a benchmark in more complex multi-regional environmentally extended input-output analysis aimed at assessing consumer responsibility.

However, the DTA and the overall EE-IOA results might be severely biased when the commodity/sector aggregation is very low and/or when the country which is analysed relies exclusively on import for certain commodities. In the latter case, in fact, either it will not be possible to compute any domestic environmental coefficient (because both emissions and output are zero) or, if this sector is aggregated with other sectors, both the technology (the row of the matrix of technical coefficients when considering both imported and domestic intermediate inputs) and the emission coefficient of the aggregated sector could fail to represent technically-viable technologies. A possible solution to this problem, although not conclusive, would be to substitute the specific rows of the matrix of technical coefficients and the specific entries of the vector of emission coefficient for these sectors with data of similar countries which have domestic production in these sectors. However, on the one hand, this kind of manipulation is likely to unbalance the whole input-output system and on the other, the similarity is difficult to check due to the variety of dimensions included in this type of environmentally extended input-output analyses.

Before discussing the way in which aggregation is likely to introduce biases in the estimates of the level of emissions induced by final domestic demand, we introduce some notation and explain how induced emis-

Table 12: Summary of the relevant notation.

Symbol	Dimension	Description
$\mathbf{Z_d}$	$n \times n$	Matrix of domestic intermediate inputs
$\mathbf{Z_m}$	$n \times n$	Matrix of imported intermediate inputs
$\mathbf{f_d^d}$	$n \times 1$	Vector of domestic final demand for goods produced domestically
$\mathbf{f_d^m}$	$n \times 1$	Vector of domestic final demand for goods produced in foreign countries (import of final goods)
$\mathbf{f_x^d}$	$n \times 1$	Vector of foreign demand for goods produced domestically (export of final goods)
$\mathbf{f_x^m}$	$n \times 1$	Vector of foreign final demand for goods produced in foreign countries (re-export)
$\mathbf{e}$	$n \times 1$	Vector of domestic air emissions
$\mathbf{i}$	$n \times 1$	Summation vector (column vector of 1s)
$\mathbf{I}$	$n \times n$	Identity matrix
$\mathbf{S}$	$n \times n$	Aggregation matrix
$\mathbf{x_d}$	$n \times 1$	Domestic output ($\mathbf{Z_d i} + \mathbf{f_d^d} + \mathbf{f_x^d}$)
$\mathbf{x_{d+m}}$	$n \times 1$	Domestic + imported output ($\mathbf{x_d} + \mathbf{Z_m i} + \mathbf{f_d^m} + \mathbf{f_x^m}$)
$\mathbf{A_{d+m}}$	$n \times n$	Matrix of technical coefficients under the domestic technology assumption ($[\mathbf{Z_d} + \mathbf{Z_m}] < \mathbf{x_{d+m}} >^{-\mathbf{1}})$)[a]
$\mathbf{L_{d+m}}$	$n \times n$	Leontief inverse under the domestic technology assumption $(\mathbf{I} - \mathbf{A_{d+m}})^{-\mathbf{1}}$
$\mathbf{f_d}$	$n \times 1$	Domestic final demand ($\mathbf{f_d^d} + \mathbf{f_d^m}$)
$\mathbf{b}$	$n \times n$	Emission coefficients ($\mathbf{e} < \mathbf{x_d} >^{-\mathbf{1}}$)

[a] $< \mathbf{x_{d+m}} >$ refers to a diagonal matrix with the diagonal composed by the elements of the vector $\mathbf{x_{d+m}}$.

sions are computed.

The notation is summarized in table 12.

When estimating the emissions induced worldwide by domestic final demand, we need to account for the intermediate inputs induced worldwide (thus using $\mathbf{L_{d+m}}$ as Leontief inverse) and for domestic final demand only ($\mathbf{f_d}$).

Induced emissions (consumption perspective, $\mathbf{e_{cp}}$) classified by product/industry are given by:

$$\mathbf{e_{cp}} = (\mathbf{b}' + \mathbf{L_{d+m}} < \mathbf{f_d} >)' \tag{3.1}$$

while total induced emissions ($\mathbf{e_{cp}^{tot}}$) may be obtained by post-multiplying $\mathbf{e_{cp}}$ by $\mathbf{i}$[10].

[10] For an exhaustive review on the accounting definitions related to environmentally extended input-output analysis, the reader should refer to Serrano and Dietzenbacher (2010) and Moll et al (2007).

3.3.2 Aggregation Biases

The issue of the choice of the level of aggregation is crucial in any empirical analysis in economics[11]. Each aggregation results in losses of relevant information and in implicit compensations which are likely to affect the reliability of the results. However, aggregation is often unavoidable. First, the most common constraint regards the availability of sufficiently disaggregated raw data. Second, privacy legislation often prevents the diffusion of disaggregated data[12]. Third, time and computation constraints are likely to induce the researcher to employ readily available and small bases of aggregated data. Finally, when matching various sources of raw data, there is little alternative to aggregation if one or more of the sources is not sufficiently disaggregated, leading to an overall aggregation. This last issue is very common in multi-regional input-output models and the general approach involves reducing the overall level of disaggregation to the level of the most aggregated country/region[13].

In environmentally extended input-output analysis, aggregation consists of a reduction in n sectors due to data availability constraints. More generally, if either the intermediate input matrices ($\mathbf{Z_d}$ or $\mathbf{Z_m}$) or the vector of direct emissions ($\mathbf{e}$) presents low disaggregation, it is enough to force the researcher to reduce the level of aggregation of the model to the lowest 'n' dimension.

More formally, the way in which we estimate embodied emissions under different aggregations ($\mathbf{e_{cp}^{aggr}}$) is described by Eq. 3.2:

[11]In this section we refer to the aggregation of basic data as opposed to the aggregation of results. The aggregation of results of any empirical analysis in economics is a necessary step when giving an overall picture of the phenomenon under analysis.

[12]Due to privacy protection, ISTAT, the Italian National Institute of Statistics is not allowed to publish data for aggregates with less than three units and it is forced to further aggregate these branches.

[13]The aggregation to the minimum common standard is the most widely used approach (Ahmad and Wyckoff, 2003; Nakano et al, 2009). However, a noticeable exception is represented by Huppes et al (2005) who exploit the very detailed US input-output table and adapt it to the EU economic structure, thus using more disaggregated data relative to publicly available EU input-output tables. Although very interesting, this approach is affected by problems related to differences between US and EU classification structures within each macro-industry.

$$\mathbf{e_{cp}^{aggr}} = \\ ((\mathbf{e'S'} < \mathbf{Sx_d} >^{-1})(\mathbf{I} - \mathbf{SZ_{d+m}S'} < \mathbf{Sx_d} >^{-1})^{-1}\mathbf{S} < \mathbf{f_d} > \mathbf{S'})' \neq \\ \neq \mathbf{Se_{cp}} \tag{3.2}$$

where $\mathbf{S}$ is the aggregation matrix. An aggregation matrix is a rectangular matrix (in our case $m \times n$, with $m < n$) composed by 1s and 0s. The column sum of $\mathbf{S}$ will be 1 for each column while the sum of all the entries equals n. Pre-multiplying a column vector by $\mathbf{S}$ results in a new vector composed by m rows in which some of the original cells are summed up in a unique entry. When dealing with a square matrix of dimension n, an aggregate square matrix of dimension m can be obtained by pre-multiplying the original matrix by $\mathbf{S}$ ($m \times n$) and post-multiplying it by $\mathbf{S'}$ ($n \times m$).

The aggregation in input-output models is related to two main dimensions: the resolution of sector/commodity disaggregation of input-output matrices and related extensions and the level of spatial - geographical aggregation (Miller and Blair, 2009).

The issues of sector/commodity aggregation in input-output models and quantification of its bias have been investigated for a long time (Hatanaka, 1952). The main concern at that time was related to computational constraints when dealing with big matrices. Aggregation was one way of easing the computation of the Leontief inverse. However, due to tremendous improvements in computational power, the issue of aggregation is currently related to constraints on the availability of or concerns over the quality of disaggregated data. The measurement and decomposition of the bias have been investigated by Morimoto (1970)[14]. The main contribution by Morimoto (1970) is related to four theorems which identify the cases in which the aggregation bias does not arise[15].

[14]The theoretical results obtained by Morimoto (1970) do not depend on the reason that induces aggregation.

[15]An important point, which often remains implicit, is that the aggregation bias only arises when the vector of final demand is modified relative to the original vector of final demand.

To summarize, the aggregation bias in static input-output models disappears if, alternatively:

- the sectors/commodities which are aggregated are characterized by the same interindustry structure;
- the vector of final demand remains unchanged for all aggregated sectors/commodities whereas it changes for all or some of the non-aggregated sectors/commodities.

However, when dealing with extensions (e.g. environmental data extensions) either these conditions should be used together or the additional condition of 'common emissions coefficient among aggregated sectors/commodities' should be satisfied. Other works provide complementary insights. Among others, Su et al (2010) focus on a description of the aggregation bias and its generalization and they perform sensitivity analysis in order to identify a minimum level of disaggregation (around 40 sectors) to assure reliable estimates. Lenzen (2011) demonstrates that it is generally desirable to have approximations of disaggregated input-output relations when environmental information is available at a very disaggregated level instead of aggregating environmental information to the level of original actual input-output data.

In our case, the aggregation bias is likely to arise because, when assessing the consumer responsibility, we consider the vector of domestic final demand (thus excluding the vector of export) instead of total final demand. This is equivalent to estimating the effect of a particular impulse (different from the actual vector of final demand) with the risk of obtaining biased results.

The main purpose of the current analysis is to aggregate our original Italian and Spanish data according to relevant aggregations used in other studies and to compare our benchmark estimates (i.e. the estimates arising from the most disaggregated model) with the estimates arising from less detailed aggregations. More specifically, our benchmark consists of a disaggregation of 50 commodities[16]. This benchmark will be

[16]This level of disaggregation corresponds roughly to the 2-digit NACE rev. 1.1 classification. For more details, refer to Section 3.2.

compared with the sub-section NACE rev. 1.1 level (accounting for 30 sectors) and with an aggregation of 16 sectors roughly corresponding to previous studies based on OECD/IEA data sources such as Ahmad and Wyckoff (2003) and Nakano et al (2009)[17]. Table 13 summarizes the sectoral detail of each aggregation we tested.

Even if several studies acknowledge that their results depend on the choice of the level of aggregation, to our knowledge, just two of them explicitly performed a sensitivity test for aggregation bias. Wyckoff and Roop (1994) found that aggregating their analysis[18] to 6 sectors (using a disaggregation of 33 sectors as a benchmark) downward biases the carbon embodied in manufacturing imports by about 30%. Su et al (2010) perform a similar sensitivity analysis on a single country environmentally extended input-output model for China. Compared to their benchmark results obtained with a disaggregation of 122 sectors[19], the bias in the estimation of carbon emissions embodied in Chinese exports arising from aggregation is positive and around 12% when using a 10-sector aggregation whereas it almost vanishes when using a 42-sector aggregation.

3.3.3 Data Sources

The current analysis relies on input-output tables for Italy and Spain for the years 1995, 2000 and 2005 with a disaggregation of 60 sectors/commodities and on NAMEA sector-level air emissions data with a disaggregation of 50 sectors for the same years and countries. To match the environmental extensions with the input-output table, we reduced

[17]OECD/IEA estimates use a disaggregation of 17 sectors. However, both OECD input-output tables and IEA CO2 emissions from fuel combustion go beyond the 2-digit NACE Rev. 1.1 as regards sector 27. This sector is split into 'Iron and steel' (271+2731) and 'Non-ferrous metals' (272+2732). On the contrary, Italian and Spanish input-output tables and NAMEA do not allow this separation.

[18]They employ a multi-regional environmental extended input-output model for 6 OECD countries (USA, Canada, France, Germany, Japan and the UK) to estimate the embodiment of carbon in imports of manufacturing products.

[19]Note that the benchmark results are obtained by 'disaggregating' the original vector of emissions intensities (42 sectors) in order to meet the 122-sector aggregation of the input-output tables. This operation is likely to partly affect the reliability of the estimates for the 122-sector aggregation.

Table 13: Sector aggregation

Aggregation level	Detail
50-sector aggregation	2-digit NACE Rev. 1.1 except 50-52, 65-67 and 70-74
30-sector aggregation	Sub-sections NACE Rev. 1.1 (2-digit capital letters): A (01-02), B (05) CA (10-12), CB (13-14), DA (15-16), DB (17-18), DC (19), DD (20), DE (21-22), DF (23), DG (24), DH (25), DI (26), DJ (27-28), DK (29), DL (30-33), DM (34-35), DN (36-37), E (40-41), F (45), G (50-52), H (55), I (60-64), J (65-67), K (70-74), L (75), M (80), N (85), O (90-93), P(95)
16-sector aggregation (source: Ahmad and Wyckoff (2003))	Agriculture, hunting, forestry and fishing (01-05); Mining and quarrying and petroleum refining (10-14, 23); Food products, beverages and tobacco (15-16); textiles, apparel and leather (17-19);Wood and wood products (20); Pulp, paper, printing and publishing (21-22); Chemicals (24); Other non-metallic mineral products (26); Iron and steel (271, 2731) + Non-ferrous metals (272, 2732); Fabricated metal products, machinery and equipment (28-32); Motor vehicles, trains, ships, planes (34-35); Plastics, other manufacturing and recycling (25, 33, 36-37); Electricity, gas (40); Construction (45); Transport and storage (60-62); All other services (41, 50-93 excl 60-62)

the overall level of disaggregation to 50 sectors. In this section, we discuss the features and the limitations of our base data in detail.

Input-Output Tables

The Council Regulation (EC) No 2223/96 of 25 June 1996 on the European system of national and regional accounts in the Community (the so-called ESA 1995) requires each member country to compile and submit supply and use tables annually and symmetric (domestic and import) input-output tables every 5 years to Eurostat. The regulation is very precise as regards the methodology used to collect the data and the structure of the published data but allows some flexibility as regards the choice between 'commodity-by-commodity' and 'industry-by-industry' input-output tables. On the one hand, commodity-by-commodity input-output tables better describe the actual technology in terms of intermediate commodities to produce a specific product whereas industry- -industry input-output table describe relationships among sectors regardless of the actual flows of commodities. On the other hand, most of the extensions (e.g. environmental extensions) refer to industries and not to commodities, making the 'industry-by- industry' approach more attractive (Eurostat, 2008; Miller and Blair, 2009). Out of the 31 countries which submit their input-output tables to Eurostat (EU27 plus Croatia, Macedonia, Turkey and Norway), 'industry-by-industry' tables are only supplied by 8 countries (Denmark, Italy, Hungary, Netherlands, Finland, UK, Turkey and Norway).

In our analysis, we use 'commodity-by-commodity' input-output tables in order to make the comparison between Italy and Spain possible. The procedure we use to assign 'industry' emissions to 'commodity' output is based on the hypothesis that direct emissions related to each commodity within a single industry are proportional to the share of the output of each commodity within the industry (Miller and Blair, 2009). Information on the commodity composition of industry output can be found in the make (supply) matrix.

Starting with the make matrix ($\mathbf{V}$) and the vector of total output by industry ($\mathbf{x}$), we compute a matrix which describes the commodity composition of industry output ($\mathbf{C} = \mathbf{V}' < \mathbf{x} >^{-\mathbf{1}}$). Each row of the matrix sums to 1 and indicates the relative weight of the different commodities in the total output of the industry (Miller and Blair, 2009; Roca and Serrano, 2007)[20]. To obtain the measure of direct emissions generated by the production of a specific commodity (by all of the industries producing that commodity), indicated with $\mathbf{e_{pp}}$, we multiply the transpose of $\mathbf{C}$ by the vector of direct emissions by industry ($\mathbf{e_{ii}}$):

$$\mathbf{e_{pp}} = \mathbf{C}'\mathbf{e_{ii}} \tag{3.3}$$

In the appendix of the current chapter (Industry-by-industry vs commodity-by-commodity) we compare our results obtained using the commodity-by-commodity approach for Italy with the results we obtain

[20]Note that when the make matrix is diagonal (that is, when all industries produce only their primary commodity), then the $\mathbf{C}$ matrix is an identity matrix.

using the industry-by-industry approach[21]. The estimates for total emissions induced by domestic final demand differ by less than 1% in all cases except for CO in 2000 and 2005, thus confirming the validity of the 'commodity-by-commodity' approach.

The NAMEA Data

The NAMEA approach to identify environmental pressures across production sectors was developed in the late 1980s and 1990s at the Central Bureau of Statistics of the Netherlands (CBS) under the supervision of Steven Keuning (Boo et al, 2003). NAMEA data are constituted by a matrix form statistical source where economic (output, value added, final consumption expenditures and full-time equivalent job) and environmental (emissions) indicators can be observed at sector level. In NAMEA, environmentally-relevant information is compiled consistently with the way economic activities are represented in national accounts (for an overview of NAMEA study we refer to Costantini et al (2011)). This framework divides the economy into production sectors and household consumption categories and shows how each industry branch or the household categories contribute to a set of environmental pressures. This allows quite robust analyses on dynamics, correlation, even causation regarding performance and resource productivity indicators.

Both the Italian, which dates back to 1990 (first published data in 2000), and the Spanish NAMEA include several air pollutants: carbon dioxide (CO2), nitrogen oxides (NOX), methane (CH4), sulphur oxides (SOX), nitrous oxide (N2O), ammonia (NH3), non-methane volatile organic compounds (NMVOC) and carbon monoxide (CO) among others. In the current chapter, we report results for emissions of five different substances (CO2, NOx, SOx, NMVOC, CO)[22] for which NAMEA with the same aggregation of sectors is available both for Italy and Spain[23].

[21] This comparison is not feasible for Spain because the Instituto Nacional de Estadistica (INE) does not produce industry-by-industry input-output tables.

[22] We also perform all the estimates for 12 additional substances available in the Italia NAMEA only (NH3, PM10, PM2.5, As, Cd, Cr, Cu, Hg, Ni, Pb, Se and Zn). Results are available upon request.

[23] The Spanish NAMEA used in this chapter is available on the Eurostat website with a

Figure 2: Emissions induced by domestic final demand by sector (Italy).

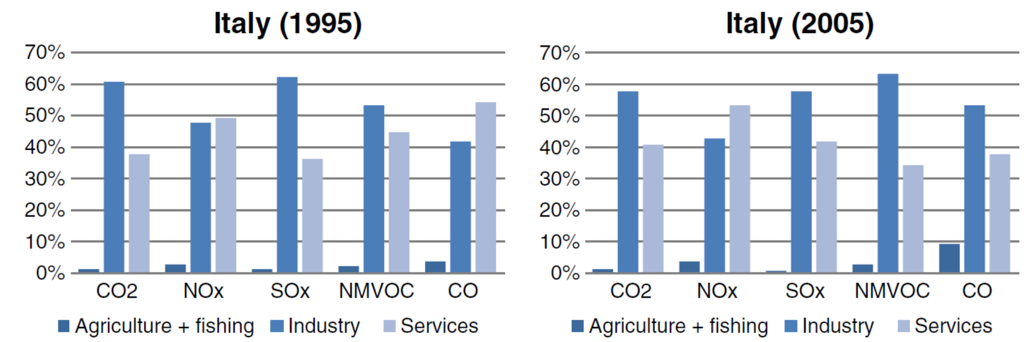
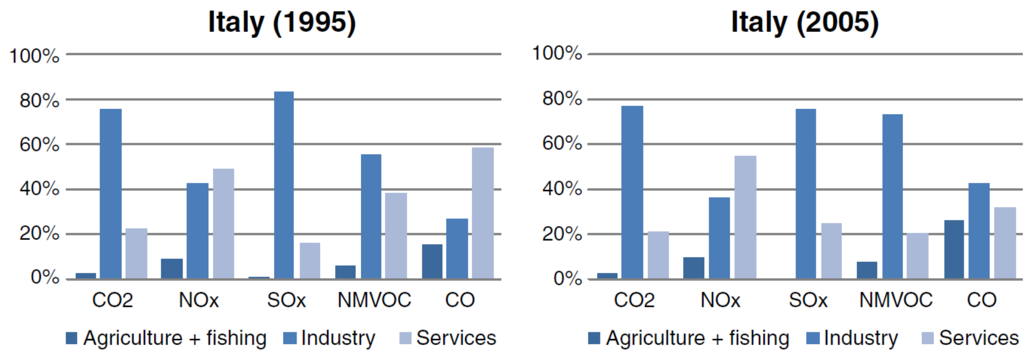


Figure 3: Direct emissions by sector (Italy).



3.4 Results and Discussion

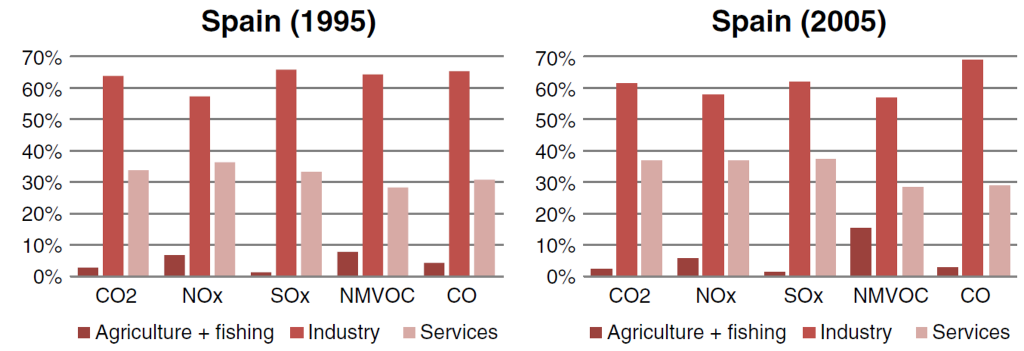
3.4.1 Overview: Consumption vs. Production Perspective in the Benchmark Case

Before facing the issue of aggregation and its related bias, in this section we briefly discuss the results for Italy and Spain of our benchmark (50 sectors) estimates for the years 1995 and 2005. The 50-sector aggregation level has been obviously considered as the benchmark; as stated by Su et al (2010), in empirical studies it is logical to take the view that the finer the level of sector disaggregation, the more refined the decomposition results obtained.

Figures 3.4.1 and 3.4.1 and figures 3.4.1 and 3.4.1 report the contri-

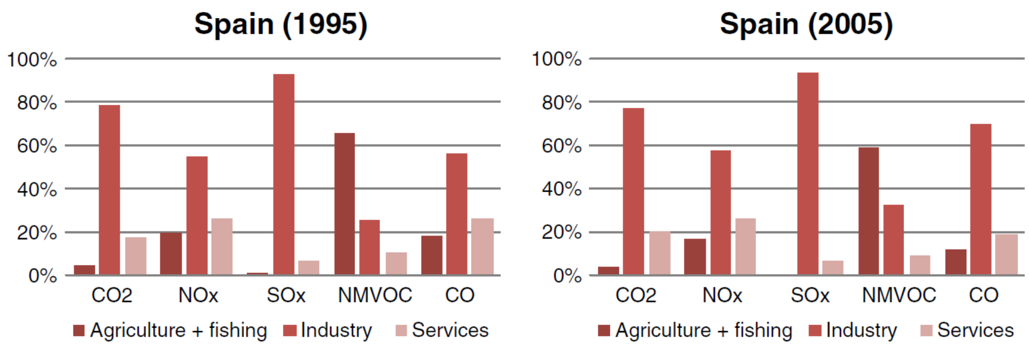
50 sector aggregation and only 5 pollutants. The Instituto Nacional de Estadistica (INE) divulgates a NAMEA with even more pollutants but with only 30 sectors and for this reason is not useful for our purposes.

Figure 4: Emissions induced by domestic final demand by sector (Spain).



[t]

Figure 5: Direct emissions by sector (Spain).



bution of three macro-sectors[24] to emissions induced by domestic final demand and domestic direct emissions for Italy and Spain respectively.

In Italy, for all emissions except NOx and CO/1995, the contribution of the demand of final products from industry is above 50%. There has been a general shift towards services in the 1995-2005 decade for CO2, NOx and SOx induced emissions. Regarding those pollutants, a weak reduction in environmental pressures caused by industrial activities from 1995 to 2005 appears; efficiency improvements in production processes and product design could be present but a composition effect cannot be excluded.

Agriculture appears almost irrelevant since most of its final products is used as intermediate inputs (the direct emissions by sector are in fact bigger that those induced by domestic final demand).

[24] Agriculture+fishing (A-B NACE Rev. 1.1), Industry (C-F NACE Rev. 1.1) and Services (G-O NACE Rev. 1.1). Results at 2-digit NACE are available upon request.

Table 14: Emissions for production and consumption perspectives (Italy, 50 sectors; in tons, CO2 in 1000 tons).

	Production perspective			Consumption perspective		
	1995	2000	2005	1995	2000	2005
CO2	360,071	368,511	389,961	348,183	355,362	376,104
NOx	1,569,712	1,233,273	1,139,097	1,507,256	1,132,557	1,035,779
SOx	1,375,635	840,127	457,795	1,374,334	774,669	398,884
NMVOC	1,064,689	713,566	584,124	1,002,686	670,275	557,370
CO	3,034,181	1,539,949	1,212,926	2,965,820	1,559,251	1,232,689

Table 15: Emissions for production and consumption perspectives (Spain, 50 sectors; in tons, CO2 in 1000 tons).

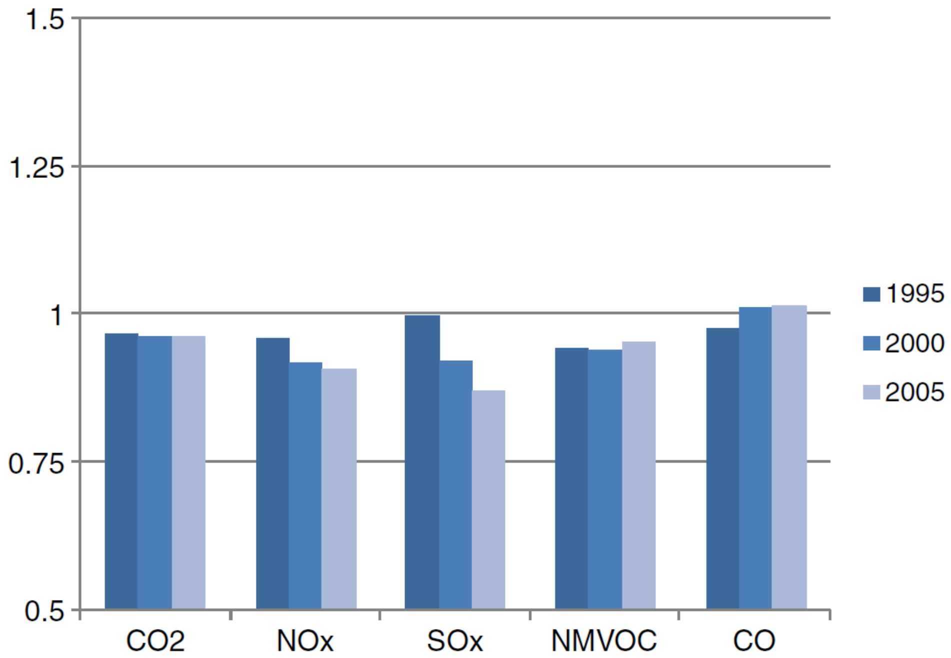
	Production perspective			Consumption perspective		
	1995	2000	2005	1995	2000	2005
CO2	208,054	248,692	294,655	220,225	306,978	382,698
NOx	1,028,209	1,155,724	1,257,268	1,074,762	1,328,240	1,560,148
SOx	1,752,362	1,453,493	1,290,977	1,891,531	2,028,020	1,750,648
NMVOC	1,865,274	1,913,460	1,987,809	2,181,989	2,380,397	2,453,815
CO	908,522	932,967	904,531	993,401	1,158,443	1,243,147

Table 3.4.1 (and figure 3.4.1) and table 3.4.1 (and figure 3.4.1) show the comparison between the consumption and production perspective for Italy and Spain respectively. A consumption/production ratio greater than 1 indicates that the emissions arising from the production needed to satisfy the domestic final demand are greater than the emissions directly generated by domestic production sectors. This is equivalent to saying that the amount of emissions embodied in imports is greater than the amount of emissions embodied in export (i.e. the country is a net exporter of emissions)[25]. The interpretation should be reversed when the consumption/production ratio is smaller than 1.

Though close to 1, the consumption/production ratios for Italy are always below unity except for CO emissions in 2000 and 2005. Furthermore, the average pattern is either stable (CO2, NMVOC and CO) or even decreasing (NOx and SOx). This result, in line with previous analyses such as Moll et al (2007) but still quite surprising for an OECD country, may have two main explanations. First, Italy maintained indus-

[25]The equivalence is explained in Serrano and Dietzenbacher (2010).

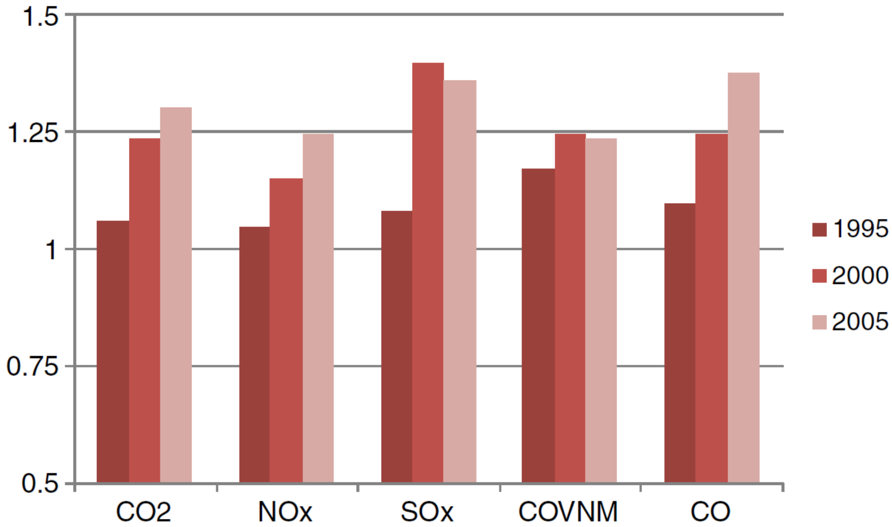
Figure 6: Consumption/production perspective (Italy, 50 sectors).



trial specialization in the manufacturing sector, especially in more traditional (and relatively energy intensive) industries, during the considered period. Second, it may be that, within each 2-digit industry, there has been a shift from polluting sub-industries (whose products, formerly produced domestically, have been substituted by import) to cleaner sub-industries. This possible shift may lead to a reduction in direct sector emissions in presence of unchanged aggregate monetary domestic output (though with a different sub-industry composition not visible in aggregate monetary data), thus artificially improving the environmental efficiency of the aggregate sector. This hidden structural change worsens the DTA prediction because it affects the sub-industry composition and the real average environmental efficiency of imports. This possible explanation further highlights the importance of using disaggregated.

The comparison between the patterns of different emissions suggests other somewhat unexpected and interesting results. Local negative externalities generated by NOx and SOx (and not by CO2) emissions, coupled with relatively strict environmental policies controlling

Figure 7: Consumption/production perspective (Spain, 50 sectors).



these emissions during the considered period[26], are expected to increase the incentive to move the production of commodities intensive in these emissions abroad (to pollution havens). This should result in an increase of emissions embodied in imports and an increase in the consumption/production ratio. However, we find the opposite. Italy, due to low stringency of environmental regulation and to lacks of enforcement, is to some extent behaving as a pollution haven within the EU (Marin and Mazzanti, in press).

Spain is characterized by the opposite situation and pattern. For all emissions/years the consumption/production ratio is greater (often far greater) than 1 and the ratio tends to increase in time, reaching the maximum for SOx in 2000 with 1.395. This means that SOx emissions induced by domestic final demand are 39.5% greater than SOx emissions directly generated by Spanish industries. These results are in line with the findings of Arto et al (2003) and Serrano and Dietzenbacher (2010).

[26]Among others, at EU level, the Council Directive 1980/779/EC substituted by the Council Directive 1999/30/EC of 22 April 1999 'relating to limit values for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air', the Council Directive 85/203/EEC of 7 March 1985 'on air quality standards for nitrogen dioxide', as last amended by Council Directive 85/580/EEC and the Council Directive 1999/13/EC 'on the limitation of emissions of volatile organic compounds due to the use of organic solvents in certain activities and installations'.

Spain was a very dynamic economy during the 90s and the early 2000s, with growth mainly driven by the construction and tertiary sectors, whereas the share of manufacturing in employment, output and value added has declined steadily[27]. This process, coupled with an increased volume of final demand of manufacturing goods (Roca and Serrano, 2007), gave rise to a rapid increase in foreign emissions to produce these goods thus worsening the balance of emissions embodied in import.

3.4.2 Aggregation Bias

In the following paragraphs, we discuss to what extent the estimates of the consumption perspective change when aggregating data.

Figure 3.4.2 and 3.4.2 show the relative magnitude of the bias in the consumption perspective emissions arising from the aggregation of sectors into 30 NACE Rev 1.1 sub-sections and in 16 sectors according to the IEA/OECD studies[28] in the Italian case.

First we note that, with few exceptions (CO2 in 1995 and CO in 1995 and 2000 for the 30-sector aggregation), a higher level of aggregation tends to overestimate the relevance of the consumption perspective, and this effect is even more evident in the 16-sector aggregation. Moreover, the bias tends to increase in time. The bias tends to be greater for the 16-sector aggregation as opposed to the 30-sector aggregation[29].

With regard to the 16-sector aggregation, the magnitude of the bias is particularly evident for SOx (with a maximum bias of almost 40% in

[27]The output share of manufacturing was 32.6%, 31.1% and 26.7% in 1995, 2000 and 2005 respectively.

[28]IEA/OECD studies such as Nakano et al (2009) and Ahmad and Wyckoff (2003) use a disaggregation of 17 sectors which, for sector 27 (manufacture of basic metals), goes beyond the 2-digit detail. IEA/OECD data distinguish between 'Iron and steel' (27.1 and 27.31) and 'Non-ferrous metals' (27.2 and 27.32). On the contrary, input-output tables and NAMEA published by ISTAT and INE treat sector 27 as a unique sector. This aggregation potentially introduces a bias in our results due to the high emissions intensity of sector 27 and to the heterogeneity in technologies and emissions intensity within sector 27.

[29]Note that there is no perfect link between the 16-sector aggregation and the 30-sector aggregation. This fact does not allow the monotonicity of the bias with respect to the number of sectors to be interpreted as a stylized fact. In fact, monotonicity is not found for Spain.

Table 16: Consumption/production perspective emissions for Italy according to different levels of aggregation.

Year	50 sectors	30 sectors	16 sectors
		CO2	
1995	0.967	0.966	1.021
2000	0.964	0.972	1.067
2005	0.964	0.977	1.077
		NOx	
1995	0.960	0.965	0.990
2000	0.918	0.974	1.016
2005	0.909	0.980	1.027
		SOx	
1995	0.999	1.001	1.093
2000	0.922	0.991	1.150
2005	0.871	0.970	1.216
		NMVOC	
1995	0.942	0.952	1.003
2000	0.939	0.956	1.035
2005	0.954	0.973	1.079
		CO	
1995	0.977	0.970	1.006
2000	1.013	1.004	1.072
2005	1.016	1.016	1.091

2005) and it is also relevant for NMVOC, CO2 and NOx.

The detailed estimates of the consumption/production perspective ratio for the different levels of aggregation (table 3.4.2) show to what extent the aggregation bias is likely to affect our main synthetic indicator, the consumption/production perspective ratio. In all cases (again except CO), moving from the benchmark result (50 sectors) to the result for 16 sectors (to be compared with the set of IEA/OECD multi-regional analyses) artificially makes Italy a net exporter of emissions even within the framework of a pure DTA. Moreover, the relative gap between consumption and production perspectives in the 16-sector case in 2005 becomes quite high for SOx (+21.6%), NMVOC (+7.9%) and CO2 (+7.7%)[30], suggesting that Italy is a net exporter of emissions.

[30]The figure for the benchmark case of the 50-sectors disaggregation was of -12.9% for SOx, -4.6% for NMVOC and -3.6% for CO2.

Table 17: Consumption/production perspective emissions for Spain according to different levels of aggregation.

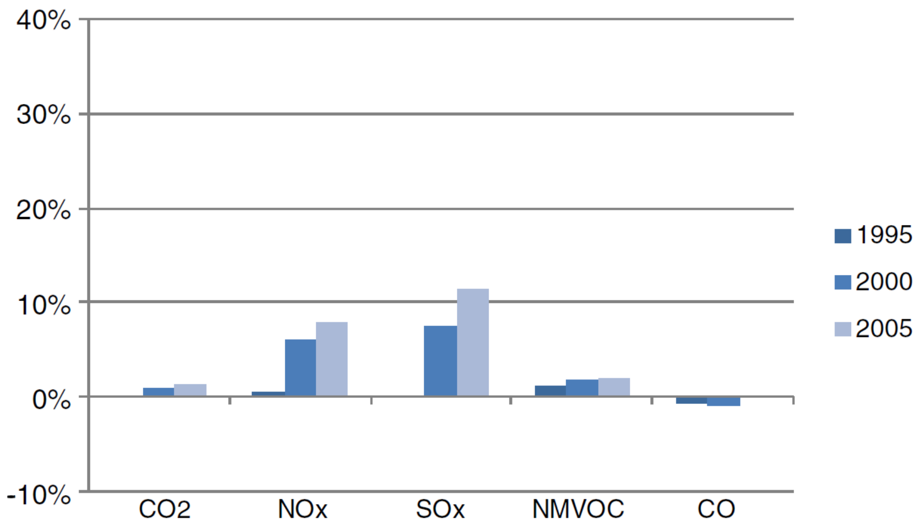
Year	50 sectors	30 sectors	16 sectors
		CO2	
1995	1.059	1.060	1.142
2000	1.234	1.176	1.288
2005	1.299	1.242	1.331
		NOx	
1995	1.045	1.062	1.096
2000	1.149	1.123	1.186
2005	1.241	1.193	1.249
		SOx	
1995	1.079	1.079	1.198
2000	1.395	1.285	1.405
2005	1.356	1.301	1.383
		NMVOC	
1995	1.170	1.049	1.079
2000	1.244	1.047	1.084
2005	1.234	1.088	1.125
		CO	
1995	1.093	1.083	1.137
2000	1.242	1.179	1.306
2005	1.374	1.283	1.453

Figures 3.4.2 and 3.4.2 report the relative aggregation bias for Spain. Results for Spain are less straightforward than the Italian ones. The bias for the 30-sector aggregation is generally negative (with the only exceptions of very small positive biases for NOx in 1995) and it is particularly high for NMVOC. No clear trend is found from 1995 to 2005. Moving to the bias for the 16-sector aggregation, it is generally positive (except for NMVOC for which it remains negative though less important relative to the 30-sector aggregation). Moreover, it tends to decrease in time for CO2, NOx and SOx and to increase for CO.

Unlike the Italian case, aggregation does not alter the status of Spain as net exporter of emissions for the full set of emissions and years (table 3.4.2).

The aggregation bias in EE-IOA depends both on the biasedness of

Figure 8: Aggregation bias %: 30 vs. 50 sectors (Italy).



the vector of total (worldwide) induced production[31] and on the combination of this vector with an aggregated vector of emissions coefficients (for which aggregation is made according to domestic production shares). The weighted average of emission coefficients for aggregated sectors uses as, weights, domestic production instead of worldwide-induced production, giving rise to an additional bias[32]. Table 3.4.2 and figures 3.4.2-3.4.2 are thus the result of the combination of the two biases and of the compensation of sector-level biases. The analytical and mathematical investigations of the contribution of the different sectors to the overall bias is beyond the objective of the current chapter (refer to Su et al (2010) for the analytical investigation of the bias). To give an idea of the results (available upon request), we report some facts on the bias for Italian input-output estimates[33]. The average positive bias in worldwide-induced production is about 0.36%, with 5 sectors characterized by a bias greater than 1%[34]. However, when considering the final results of the estimates for the consumption perspective, the aggregation bias is much

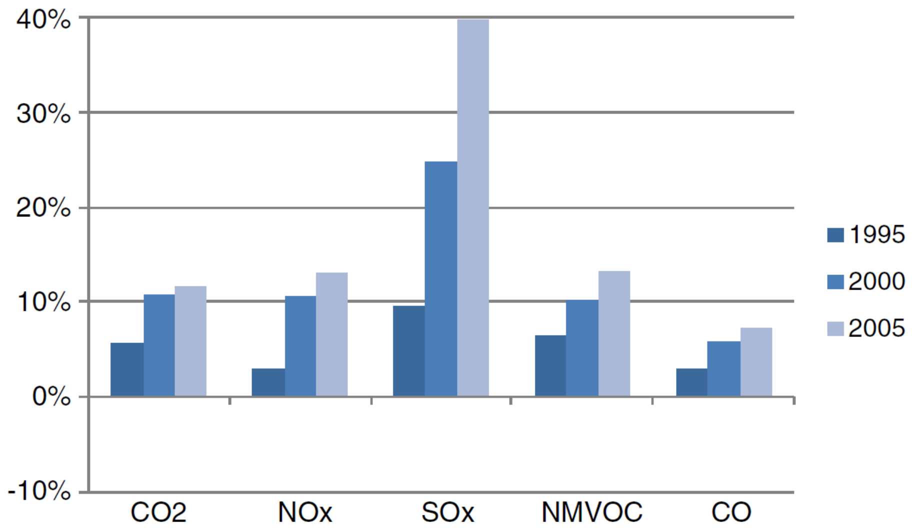
[31]In the disaggregated case the vector of worldwide-induced production is given by $(\mathbf{I} - \mathbf{Z_{d+m}} < \mathbf{x_d} >^{-1})^{-1}\mathbf{f_d}$.

[32]Note that this bias is still related to the aggregation bias in the estimates of the vector of worldwide-induced production.

[33]The results reported in the following example refer to Italian input-output tables for 2005 and CO2 emissions and to the 30-sector aggregation.

[34]CB -1.87%, DD +1.49%, A +1.43%, DJ +1.32% and DM 1.05%.

Figure 9: Aggregation bias %: 16 vs. 50 sectors (Italy).



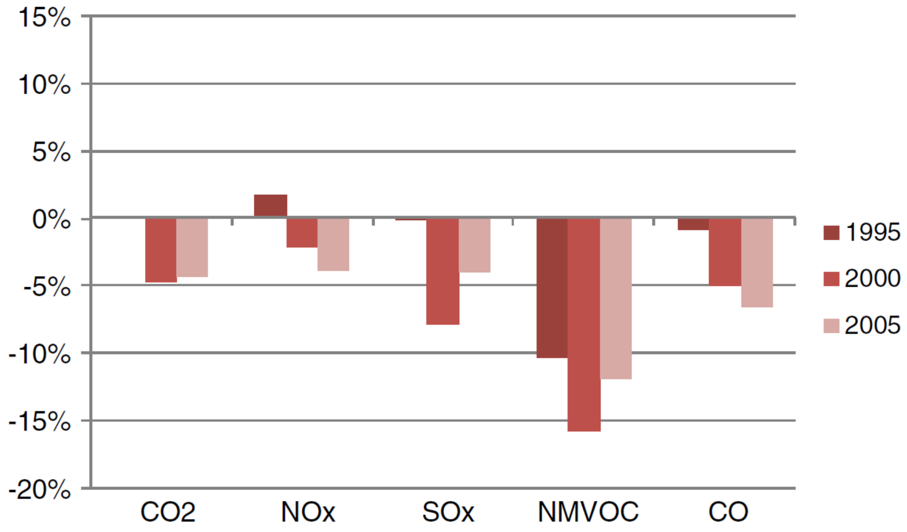
bigger. There are four sectors for which the bias is greater than 10%[35] and five sectors for which the bias ranges between 5% and 10%[36]. Sector CB (Mining and quarrying, except of energy producing materials),which is the sector characterized by the most severe bias, is composed by two very different sub-sectors (2-digit Nace): sector 13 (Mining of metal ores) with an emission coefficient of 5.4 tons of direct CO2 emissions per million of Euro and sector 14 (Other mining and quarrying) with a coefficient of 163.6 tons per million of Euro. Moreover, the share of domestic production and of worldwide-induced production of the two sub-sectors relative to the aggregate sector CB differs substantially: sector 13 (the less emission intensive) accounts for 14% of domestic production and for 24% of worldwide-induced production of sector CB. As a consequence, sector 13, the less emission-intensive, is under-weighted in the aggregate emission coefficient when considering worldwide-induced production, leading to a positive aggregation bias.

The results arising from this simple example should be kept in mind when discussing our benchmark results (50 sectors).Within the 2-digit Nace classification, there are several sectors for which we expect relevant sub-sector heterogeneity regarding emission coefficients and do-

[35]CB +48.9%, DJ -15.4%, DN +11% and DB -10.1%.

[36]J, CA, O, E and DJ.

Figure 10: Aggregation bias %: 30 vs. 50 sectors (Spain).



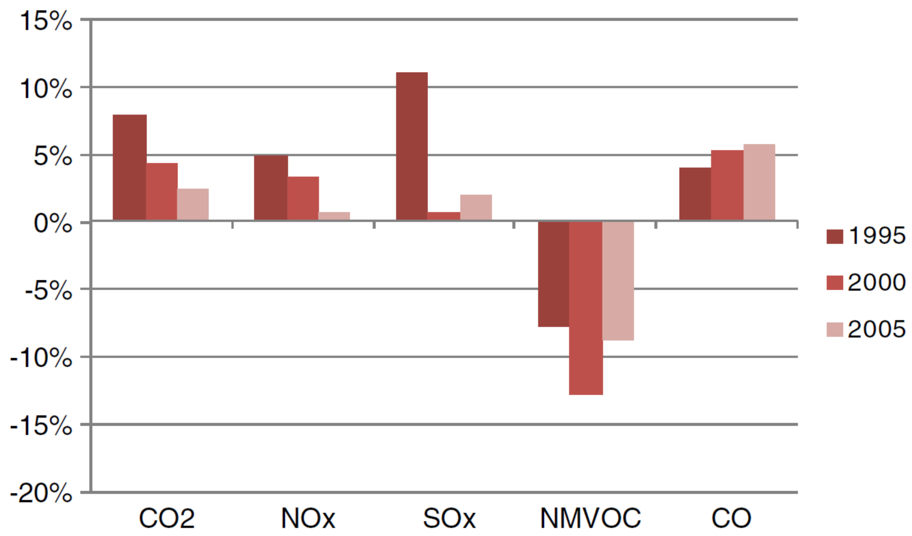
mestic worldwide-induced production patterns. This unobservable heterogeneity is source of possibly severe bias even when using the 50 sector classification. Unfortunately, no input-output table has been published yet with greater detail for most EU countries[37].

3.4.3 Comparison with Previous Studies

In the last decade, as previously indicated, some empirical studies have been conducted focusing on carbon or other pollutant embodiments in trade using international-comparable data especially from OECD sources (input-output, CO2 emissions and Bilateral Trade) (e.g. Ahmad and Wyckoff (2003); Nakano et al (2009)), Eurostat sources (e.g. Moll et al (2007)) and single country sources (e.g. Arto et al (2003); Serrano and Dietzenbacher (2010) for Spain; Su et al (2010) for China). Other recent studies and ongoing projects inherent this theme relates to the construction of a world input-output database (WIOD project) - that includes various environmental indicators - and a new environmental accounting framework using Externality Data and Input-output Tools for Policy

[37]Huppes et al (2005) use the US input-output table (with a disaggregation of about 500 sectors) and modify it to fit European aggregates. The main shortcoming of that approach is the necessity to perform several manual manipulations to the original data which limit the possibility to replicate and compare the results.

Figure 11: Aggregation bias %: 16 vs. 50 sectors (Spain).



analysis (EXIOPOL) set up respectively under the EU's 7th and EU's 6th framework Program[38]. These projects represent good examples of standardisation and harmonization processes involving input-output tables of several countries and environmental data[39].

Among the empirical studies provided by the literature, for comparison purposes, we only consider those that include Italy or Spain or both. (Ahmad and Wyckoff, 2003) in their OECD study consider 24 countries responsible in 1995 for 80% of global emissions and global GDP (in nominal prices); following this study, Nakano et al (2009) increase the former OECD analysis to 41 countries/regions so that more than 90% of world GDP is covered. The study of Moll et al (2007) includes 8 EU countries[40] selected on the basis of data availability and the high coverage purpose

[38]Some possible but preliminary applications with environmental extensions of the WIOD database have been presented in occasion of the World Bank workshop "The Fragmentation of Global Production and Trade in Value Added" (June 9-10, 2011). New studies based on the EXIOPOL and follow-up projects are currently under way as reported in the Sixth Meeting of the UN Committee of Experts on Environmental-Economic Accounting (New York, 15-17 June 2011).

[39]On the side of the standardisation of environmental accounts, the new System of Environmental and Economic Accounting (developed by the United Nation Statistics Division) represents an important development.

[40]The selected 8 economies represent more than two thirds of EU25's GDP and more than 60% of EU25's population. The geographical coverage comprises ES, UK (1995) and DE, DK, HU, IT, NL, SE (1995 and 2000).

Table 18: CO2 emissions for production and consumption perspective in Italy in different studies (Mton CO2).

Source	MRIO or DTA	Aggr. level (# sectors)	Production perspective		Consumption perspective		C/P	
			1995	2000	1995	2000	1995	2000
Nakano et al (2009)	MRIO	17	413	427	511	554	1.24	1.30
Ahmad and Wyckoff (2003)	MRIO	17	398[a]		445[a]		1.12	
Moll et al (2007)	DTA	38		358		362		1.01
Own elaboration	DTA	50	360	369	348	355	0.97	0.96

[a] 1992.

of European economic contexts.

A comparison of our CO2 results with the empirical evidence for the same pollutant found in the recent EE-IOA studies suggests that as far as the Italian case is concerned (table 3.4.3), some of the studies are affected by aggregation bias due to a small number of considered sectors. This results in a strong and significant difference among empirical findings with respect to both the consumption and production perspective emissions and the corresponding ratio. In Nakano et al (2009) and Ahmad and Wyckoff (2003), the C/P ratios reported for the Italian case, in 1995 and 2000, are larger than ours and always higher than 1[41]. The Moll et al (2007) figure is the closest to our 2000 figure for the C/P ratio (0.96); they use a 38-sector aggregation level. If we consider the sensitive results found by Su et al (2010) (levels around 40 sectors appear to be sufficient to capture the overall share of emissions embodied in a country's export), it may be considered more reliable than other authors' findings. From a policy point of view, a C/P ratio that ranges from 1.24-1.30 to 0.96-0-97

[41] However, the comparison has to take into account that there is a severe heterogeneity in the methodologies used by different authors. For example, differently with respect to our study, Ahmad and Wyckoff (2003) do not use the NAMEA data framework but IEA data; moreover they use, as in Nakano et al (2009), MRIO and not DTA. A consistent comparison of the absolute levels of CO2 emissions between IEA/OECD studies and NAMEA-based studies is not possible. In fact, IEA records CO2 emissions from fuel combustion only and, differently from NAMEA, the principle of recording the emissions generated by resident agents only is not applied in the collection of these data.

Table 19: CO2 emissions for production and consumption perspective in Spain in different studies (Mton CO2).

Source	MRIO or DTA	Aggr. level (# sectors)	Production perspective		Consumption perspective		C/P	
			1995	2000	1995	2000	1995	2000
Nakano et al (2009)	MRIO	17	236	280	275	330	1.17	1.18
Ahmad and Wyckoff (2003)	MRIO	17	235		252		1.07	
Serrano and Dietzenbacher (2010)	DTA	46	204	239	222	279	1.09	1.17
Arto et al (2003)	DTA	46		364[a]		429[b]		1.18
						453[c]		1.24
Moll et al (2007)	DTA	46	209		228		1.09	
Own elaboration	DTA	50	208	249	220	307	1.06	1.23

[a] MtCO2e.
[b] MtCO2e with Monetary DTA.
[c] MtCO2e with Physical DTA.

suggests that while large studies that involve several countries have to be encouraged because they permit macro area analysis, in the meantime if they require a low level of sectoral detail to assure countries' homogeneity and comparability, their empirical results require caution when they are interpreted.

Table 3.4.3 shows a similar comparison for Spain. With regard to this country, the empirical findings reported in the different studies are more homogeneous than the Italian case both for the absolute values of production and consumption perspective CO2 emissions and the corresponding ratio. This could be interpreted, at least partially, as a confirmation of the higher relative reliability of our 50-sector estimates. However, in the light of the Italian results, we could conclude that after a certain degree of aggregation, there is a concrete risk of having biased and volatile results which depend on the specificities of the economic structure of the country and the type of emission considered.

3.5 Conclusions

The integration of the National Accounting Matrix including Environmental Accounts (NAMEA) and input-output (I-O) tables (often referred to as Environmental Extended Input-Output Analysis - EEIOA - based on NAMEA data) represents a new way to analyse the determinants of the income-environment relationships in international settings. Moreover, EE-IOA provides analyses of the emissions embodied in domestic consumption and domestic production by considering the structure of intermediate inputs and environmental efficiency in each production sector.

A comparison of a production and consumption perspective may have relevant policy implications. A consumption and production emission ratio greater than one denotes a country that is a net exporter of emissions in the sense that it requires an amount of emissions embodied in imports, and thus produced abroad, that is greater than the amount of emissions embodied in export. Usually, the environmental policy points mainly to production activities as responsible actors of impacts to be targeted by legislation and regulation. Looking at the final consumption demand for vertically integrated domestic and international environmental impacts can push policy attention towards the possible role of consumers as actors to be targeted with particular environmental policies, together with the international responsibility for environmental externalities of pollutants' emissions produced abroad but domestically demanded.

However, similar comparisons require particular assumptions, such as the technology associated with the imported goods, and could be affected by some biases. In this chapter we have analysed and discussed the aggregation bias due to different levels of production sector aggregation for Italy and Spain in 1995, 2000 and 2005. Our empirical findings, for the Italian and the Spanish cases, show that a different sectoral aggregation significantly biases the amount of emissions both for the consumption and the production perspective. If we consider only 16 production sectors, the results obtained in the consumption perspective are quite different from those for higher levels of sector disaggregation (e.g.

50 which is our benchmark) both for the amounts of calculated emissions and for the corresponding C/P ratios. With regard to Italy, the 16-sector aggregation level in 2005 shows an emission amount for CO2, NOx and NMVOC which is more than 10% higher than those calculated with the 50-sector aggregation level. Moreover, considering SOx, the gap between 16- and 50-sector aggregations reaches almost 40%. With regard to Spain, between 16- and 50-sector aggregation levels in 2005, there are differences of below +5% for CO2, NOx and SOx, and almost 5% for CO. NMVOC shows the biggest gap for the Spanish case with an underestimation of almost -8% compared with the benchmark aggregation level due to the use of a 16-sector aggregation level.

Our results suggest that special attention must be paid when interpreting the EE-IOA of country estimated amounts of embodied emissions, both in domestic final demand and those directly associated with the production sectors when the sectoral aggregation level has a low definition as considered in some recent similar studies.

Industry-by-industry vs commodity-by-commodity

The methodology we used to employ in a consistent way commodity-by-commodity input-output tables as a proxy of industry-by-industry tables is explained in Section 3.2. While the main analysis relies on results obtained using commodity-by-commodity input-output tables, in this appendix we report the differences between the industry-by-industry approach and the commodity-by-commodity approach as regards the estimation of the emissions induced by domestic demand. This comparison is only possible for Italy because Spain does not publish industry-by-industry input-output tables. The main results are summarized in table 3.5. With the only exception of CO emissions, the absolute value of the gap for aggregate consumption perspective emissions is always below 1%. On average, the commodity-by-commodity approach tends to underestimate the emissions induced by the final demand of agriculture-fishing goods and industrial goods whereas it overestimates the emissions induced by the final demand of services. Finally, we do not observe

Table 20: Commodity-by-commodity (cc) versus industry-by-industry (ii) approach for Italy (1-ii/cc).

1995	CO2	NOx	SOx	NMVOC	CO
Agriculture + fishing	-4,74%	-4,64%	-4,34%	-4,34%	-5,41%
Industry	-2,79%	-0,83%	-2,33%	-2,08%	-1,09%
Services	2,60%	-0,70%	1,94%	1,66%	-0,14%
Total	-0,84%	-0,88%	-0,83%	-0,48%	-0,74%
2000	CO2	NOx	SOx	NMVOC	CO
Agriculture + fishing	-4,61%	-4,37%	-4,54%	-4,63%	-5,25%
Industry	-4,17%	-2,42%	-3,37%	-4,37%	-5,93%
Services	5,95%	3,24%	5,87%	7,62%	4,95%
Total	-0,55%	0,34%	-0,02%	-0,41%	-1,61%
2005	CO2	NOx	SOx	NMVOC	CO
Agriculture + fishing	-4,69%	-4,57%	-4,47%	-5,36%	-6,17%
Industry	-3,91%	-2,52%	-3,36%	-4,27%	-6,48%
Services	4,57%	2,48%	4,17%	7,62%	6,71%
Total	-0,63%	0,02%	-0,37%	-0,51%	-1,88%

relevant changes in the magnitude of the gaps over time.

Chapter 4

Do Eco-Innovations Harm Productivity Growth through Crowding Out?

4.1 Introduction

Technological progress, together with structural change and shifts in consumption patterns, has been acknowledged to be a crucial factor in achieving environmental sustainability (Jaffe et al, 2002; Popp, 2010; Popp et al, 2009). Technological progress might improve environmental performance both through increased resource efficiency and lower emission intensity in production activities and through the supply new more 'sustainable' products as substitutes to other less efficient products (e.g. energy intensive durable goods). Firms are key actors in the creation, adoption and diffusion of environmental innovations as well as the most important responsible for environmental pressures.

The economic literature on eco-innovation patterns at the micro level focuses to a great extent on the identification of the drivers of eco-innovation by firms with little attention given to the effect of eco-innovation on productivity or financial performance of firms. Moreover, most of these empirical works are based on German firms. Rennings and Ziegler (2004) use data from the German Community Innovation Survey (CIS) finding significant positive effect of environmental organizational measures (EMAS and ISO 14001), market opportunities and R&D intensity on process and product environmental innovations. Wagner (2007) uses both data on environmental patent applications and self-reported measures of eco-innovation to investigate the effect of environmental management on environmental innovations. Results for German firms show positive effect of EMS adoption on self-reported process environmental innovations and a negative effect on firms' general patenting activity. The paper by Horbach (2008) uses a discrete choice model for German manufacturing firms finding strong positive effects of technology push (knowledge capital), demand pull (social awareness of customers) and environmental policy (either mandatory or voluntary through environmental management tools) factors on environmental innovations. Horbach et al (2011) is the first relevant study investigating the determinants of different fields of environmental innovations. Their analysis, based on the German CIS for 2009, shows that the introduction of innovations aimed at reducing by-products of production activities such as the release of air, water and noise emissions are strongly related to government regulations (current and expected). On the other hand, innovations aimed at reducing material and energy use are driven by cost-savings and resource and energy taxes due to the easier appropriability of the returns from innovation through reductions in production costs. Rave et al (2011) base their analysis on German firms and on their patenting behaviour. The main results highlight the importance of a clear and strict environmental regulatory framework, of possible cost savings due to environmental innovations and of the possibility of creating new markets. Finally, results from a survey conducted by Cainelli et al (2011) show that different types of environmental innovations introduced and adopted by manufacturing firms in Emilia Romagna (Italy) is very strongly correlated to international characteristics (foreign ownership and export propensity) and networking with other firms and institutions.

While environmental innovations are expected to have, by definition, a beneficial effect on the environment[1], their effect on productivity is less straightforward. The conventional wisdom predicts that starting from a situation of optimizing firms, any policy aiming at limiting environmental by-products of firms will result in a reduction in measured productivity. These productivity losses could be reduced by introducing environmental innovations. However, productivity losses cannot be fully removed and resources devoted to generate or adopt environmental innovations should be diverted from 'optimal' research project with higher expected returns (crowding out). In this respect, Popp and Newell (2009) find significant crowding out of energy R&D expenditures on general R&D in those US industries characterized by more than 5 percent of energy R&D. However, when they consider energy patents at the firm level, evidence is more mixed, with relevant but insignificant crowding out effect.

An alternative view, promoted by Porter and van der Linde (1995), allows for the possibility of win-win outcomes. In this case, environmental regulations help to fill information gaps about available technologies and technological opportunities and they help solving the additional appropriability problem of environmental innovations (EI) due to the fact that EI reduce external, generally not priced, costs ('weak' version of the Porter hypothesis). Moreover, early introduction of environmental technologies is expected to generate early mover advantages for regulated firms, with long run positive effects on competitiveness and, eventually, on measured productivity ('strong' version of the Porter hypothesis)[2].

In this respect, Rexhauser and Rammer (2011) use the German CIS 2009 to investigate the effect of different types of environmental innovations on German firms' profits. They find that cost-reducing innovations aimed at reducing energy and material input have a positive effect on firms' profitability while regulation-induced environmental innovations,

[1]Economists and policy makers are increasingly worried about the possibility that cost and price reductions brought by environmental innovations through improvement in material and energy efficiency would result in an increased consumption of these new efficient goods, with an overall negative effect on the environment (rebound effect).

[2]For a more detailed discussion about the difference between the 'strong' and 'weak' version of the Porter hypothesis refer to Jaffe and Palmer (1997).

mainly aimed at reducing environmental pressures, have a negative but weak effect on profitability.

The aim of this chapter is to assess the drivers of environmental innovations and their effect on firm-level productivity. I employ a panel of Italian manufacturing firms for the period 2000-2007 containing information on balance sheet and income statement, EPO patent applications and polluter status in order to jointly identify the drivers of eco-innovations and their contribution to firm-level productivity. The empirical framework is that of a modified CDM model (Crepon et al, 1998) to account for eco-innovation patterns. The rest of the chapter is organized as follows. Section 2 briefly defines eco-innovation and the extent to which patent data are a useful source of information on eco-innovation. Section 3 focuses on the description of the empirical model and of the data. Section 4 discusses the results. Section 5 concludes.

4.2 Definition of environmental innovations and the role of patent data

A definition of environmental innovation is needed in order to investigate its impact on productivity and potential crowding out effects. There has been a rich debate in the economic literature about the distinctive features of environmental innovations as opposed to general innovations (Rennings, 2000). Environmental innovation (or eco-innovation) has been defined by Kemp and Pearson (2007) within the project 'Measuring Eco Innovation' as

> [...] the production, assimilation or exploitation of a product, production process, service or management or business method that is novel to the organisation (developing or adopting it) and which results, throughout its life cycle, in a reduction of environmental risk, pollution and other negative impacts of resources use (including energy use) compared to relevant alternatives.

This is a broad definition, making it difficult to measure environmental innovation in a comprehensive way. On the one hand, survey are able to describe qualitatively the whole spectrum of eco-innovation strategies of innovative firms. On the other hand, however, the broad definition of eco-innovation is likely to result in ambiguous questions in the questionnaires which are prone to misleading interpretations by surveyed people.

Patent data could act as a more objective alternative to measure eco-innovation (Oltra et al, 2009). Patents contain rich information about the technological field of the underlying innovation, especially when analysing reported IPC classes and the text contained in the patent or in the abstract. This rich information is generally exploited through the identification of relevant 'environmental' IPC classes and /or through the systematic search of 'environmental' keywords.

Nevertheless, the use of patent data as measure of environmental innovation output within the definition elaborated by Kemp and Pearson (2007) (but also more generally for all innovations[3]) is characterized by some serious limitations.

First, patents cover just part of the innovation output. Many innovations are not patented either because they cannot be patented[4] or because firms prefer to use alternative means to protect their innovations (secrecy, lead time, etc.). Moreover, the propensity to use patents as a mean of protecting innovations varies substantially across sectors and across technologies. In general, process innovations, which are very relevant when considering environmental innovations, are under-represented as opposed to product innovations.

Second, information on the ownership and actual use of patented innovations is generally lost after the patent has been granted. Patent data ignore the whole phase of 'adoption' of innovations. It is thus plausible that patented innovations are not even adopted by applicant firms which could act as specialized suppliers of (embodied or disembodied)

[3]Refer to Griliches (1990) for a survey on the advantages and limitations of patent data as a measure of innovation.

[4]An innovation can be patented if it is novel, non-obvious and commercially viable. Moreover, specific patent offices do not allow to patent specific technologies (e.g. living organisms).

knowledge to other firms which are the real adopters.

Third, patent data consider only those innovations which are 'new to the market' while they ignore those innovations which are just 'new to the firm' because of the 'novelty' requirement for patented innovations.

Finally, the distribution of the value of patents is very skewed, with a tiny proportion of extremely valuable patents and a great majority of patents with little or even no commercial value (Hall et al, 2007). Finally, patenting firms represent a very small fraction of innovative firms, leading to possibly low robustness of the results and to econometric problems when dealing with excess zeros of patent count indicators.

Despite these limitations, many recent analysis on environmental innovations were based on patent statistics. Among other, refer to Lanjouw and Mody (1996), Popp (2002), Brunnermeier and Cohen (2003), Wagner (2007) and Johnstone et al (2010).

4.3 Empirical model and data

Econometric analysis are based on an adaptation of the CDM model (Crepon et al, 1998). The CDM model is an empirical structural econometric model aimed at investigating innovation patterns of firms in a comprehensive way, considering the drivers of innovation inputs (R&D), their effect on innovation success (innovation output) and the extent to which innovation success affects firm's productivity.

4.3.1 Classical and extended CDM model

Classical CDM model

The CDM model is an empirical structural model proposed by Crepon et al (1998) to evaluate innovation patterns of firms in a comprehensive way. The model is composed by three steps. In the first step, firms decide whether to undertake formal R&D projects or not and the amount of resources to devote to R&D activities. The choice of innovation inputs is modelled with an Heckman selection model to account for incidental truncation of the R&D variable. In a second step, firms use innovation inputs and other internal or external resources to obtain commercially viable innovations. The original CDM model (Crepon et al, 1998) used two alternative measures of innovation output: share of innovative sales (which, in their case, was a categorial variable) and patent applications count (count variable). Other more recent CDM models based on CIS data used alternative measures of innovation output such as the introduction of process and / or product innovations. Finally, successful innovations affect firm's profitability and / or productivity. Innovation output is included in an extended production function as an additional input and its effect on productivity is assessed.

By using predicted values of R&D and patent counts in the second and third step respectively, the CDM model is a sort of instrumental variable approach to correct for simultaneity and reverse causality issues in the various steps.

R&D equation

The first decision of firms about their innovation strategy is whether to perform any formal R&D and, eventually, its intensity. The CDM model uses a Heckman sample selection model to estimate R&D intensity. R&D expenditure is characterized by incidental truncation, with the decision to perform formal R&D depending on (unobservable) expected returns on R&D (i.e. whether expected returns exceed R&D investments). Moreover, R&D strategies of firms might be modelled as a the two-stage decision process in which firms decide, in a first stage, whether to perform any formal R&D and, in a second stage, its intensity. I estimate the Heckman selection model simultaneously with maximum likelihood.

Explanatory variables for the first step (probability of reporting positive R&D) are log of employees count, market share, log of capital intensity (fixed physical assets per employee), log of total assets (book value), age (a dummy variable for firms older than 10 years) and sector[5], year

[5]I classify sectors according to the Pavitt's taxonomy (Pavitt, 1984). Pavitt classifies manufacturing sectors according to their patterns of innovation ending up with four macro-sectors: (i) supplier-dominated sectors; (ii) scale intensive sectors; (iii) specialized suppliers sectors; (iv) science based sectors.

and macro-regional[6] dummies. I exclude (exclusion restriction) book value and firm's age in the second step (R&D intensity expressed as the log of R&D per employee) of the Heckman model, assuming that firm's age and the value of its assets affects the probability of performing R&D (extensive margin) but not its intensity (intensive margin).

It is important to bear in mind that possible severe measurement errors in the R&D variable (refer to the appendix 'Adjustments to AIDA' for further detail on the R&D variable) are likely to cause a substantially overestimated standard errors.

Patent equation

The combination of innovation inputs (R&D) with internal and external resources results in the introduction of innovations. Successful innovations have been measured in CDM models in a variety of ways. Crepon et al (1998) use patent applications count and share of innovative sales as indicators of successful innovations. Other authors (e.g. Hall et al (2009) for Italy and Griffith et al (2006) for France, Germany, Spain and the UK) used dummy variables describing the introduction of innovations, generally distinguishing between process and product innovations. In this chapter I use the number of EPO (European Patent Office) patent applications as a measure of innovation output.

Patent data are count data. When the dependent variable is a non-negative integer, OLS are likely to be biased and they could give rise to negative predicted values. In this model I use a negative binomial (NB2) regression model[7]. The baseline model to deal with count dependent variables is the Poisson model which assumes equidispersion (mean equal to the variance) of the variable of interest. This property is often violated in actual data, which are generally characterized by overdispersion[8] (Cameron and Trivedi, 1998). I use a Negative Binomial (NB2)

[6] I identify four macro-regions: North-West, North-East, Central Italy and Southern Italy and Islands.

[7] Preliminary attempts have been done with other models which deal explicitly with excess zeros (zero inflated Poisson or NB, hurdle Poisson or NB).

[8] The unconditional variance of total patent count is, in all samples, much higher than the unconditional mean. In the full sample the variance 6.5 time the mean, in the patent

model which allows the conditional variance of the dependent variable to be a quadratic function of the conditional mean[9].

The dependent variable of the patent equation is yearly EPO applications count. Explanatory variables are (predicted) log of R&D per employee, firm size (log of employees), local patent stocks (log of regional patent stocks per capita) and sector, macro-region and year dummies. Innovation input and firm size (Mansfield, 1986) are expected to affect positively the number of patent applications. As regards local patent stock, it might act either as innovation input for firms in the form of local knowledge stock or as potential alternative to internally patented innovations to be adopted through licensing or embodied in machineries and intermediate inputs. These two alternative interpretations affect firm-level patenting activity in opposite directions. Sector dummies (Pavitt's taxonomy) control for sector-specific propensity to apply for a patent. As pointed out by Pavitt (1984), the propensity to use patents as a mean to appropriate the returns to R&D is high for science-based and specialized suppliers sectors while scale intensive and supplier dominated sectors have lower propensity to patent their innovations[10]. Finally, year and regional dummies control for changes in the propensity to patent through time (due either to changes in firm's strategies or to variations in patent systems) and for geographical differences in human capital, links with local actors in the local innovation systems and innovation capabilities.

Productivity equation

The final step consists in estimating the effect of successful innovation output on firm's productivity or profitability. I use an extended Cobb-

sample the variance is 4.47 times the mean and in the polluter sample the variance is 13.36 times the mean.

[9]Being ω_i the conditional variance of the count variable, and μ_i its conditional mean, the NB2 model assumes that $\omega_i = \mu_i + \alpha\mu_i^2$. In case α is not statistically different from zero, the NB2 model converges to a Poisson model.

[10]Descriptive statistics regarding the sample of firms used in this chapter confirm Pavitt's priors about sector-specific propensity to patent: in table 23 I observe a much higher share of firm/year pairs with positive patents in science based and specialized supplier sectors (4.27% and 5.14% respectively) than for scale intensive (1.97%) and supplier dominated (0.78%) sectors.

Douglas production function in which the log of labour productivity (value added per employee) is a function of the log capital intensity (fixed physical assets per employee) and the log of (predicted) innovation output (expected number of patent applications per employees). I also include the log of employees as explanatory variable to test for constant returns to scale[11]. I include sector, year and regional dummies to account for sector, time and region specific productivity shocks. I estimate the productivity equation with OLS.

Extended CDM model

My extension to the classical CDM model consists into the separation of the innovation output (patent applications count) into two categories: environmental innovations and other innovations[12]. The approach of splitting innovation outcome in multiple categories has been extensively used in CDM models, especially by separating product innovations from process innovations (Griffith et al, 2006; Hall et al, 2009). This extension will result in two different patent equations and in a unique productivity equation in which environmental and non-environmental innovations will have different productivity effects. The separate assessment of the productivity effect of environmental and non-environmental innovations will allow me to (indirectly) test for the presence of crowding out of environmental innovations relative to other innovations[13].

An additional extension regards the special consideration, as regards

[11] A general Cobb-Douglas production function with two inputs (labour, L, and capital, K) is given by the equation $VA = AK^{\alpha}L^{\beta}$. Constant returns to scale imply that $\alpha+\beta = 1$. Dividing both sides by L, taking the log and rearranging I obtain $\log(VA/L) = \log(A) + \alpha \log(K/L) + (\beta + \alpha - 1)\log(L)$. Under the assumption of constant returns to scale, the parameter for $\log(L)$ should be zero ($\beta + \alpha = 1$). The same concept applies with more than two inputs.

[12] Note that, because of the complexity and variety of environmentally-beneficial technologies and of the approach of selecting environmental patents only by means of their IPC class, the non-environmental category of patents is likely to contain a possibly remarkable number of environmental innovations.

[13] Popp and Newell (2009) try to disentangle the presence of crowding out at the firm level in a more direct way. They check whether an increase in energy patent applications at the firm level reduces the number of other patent applications. They implicitly assume that the return of energy patents is (on average) lower than the return of other patents.

environmental innovations, for polluting firms/sectors. Polluting firms and sectors are expected to show a significant and systematic bias towards environmental innovations relative to other firms/sectors. This fact is likely to be reflected in the patent equation, with polluting firms and sectors which will probably apply for a greater number of environmental patents. Table 21 reports the results of a series of Probit regressions in which the dependent variable is the probability of filing for an environmental patent and the explanatory variables the polluter status (either polluting firm or polluting sector) and a series of controls[14]. Estimates has been performed on the sub-sample of observations with positive patent applications count. On average, the probability of being an eco-innovator is greater for polluting firms ('polluter') and polluting sectors, thus highlighting the relative bias of polluting firms and sectors towards eco-innovations relative to other firms and sectors. However, especially when controlling for firm size, the bias of polluting firms becomes statistically insignificant.

In addition, I aim at investigating whether environmental innovations affect polluting firms in a different way relative to other firms. On the one hand, polluting firms might profit from environmental innovations through their effect on the decrease of compliance costs. On the other hand, however, environmental innovations introduced to reduce compliance costs coupled with constraints in the amount financial resources that a firm can invest in R&D activities could determine a shift of the innovation strategies of firms towards innovations with lower expected returns. This test is done by interacting the predicted intensity of environmental patents with the dummy indicating that a firm is a polluter in the productivity equations.

[14] Year dummies and macro-region dummies (North-West, North-East, Central Italy and South-Islands), macro-sector dummies (Pavitt's taxonomy), firm-size (log of employees) and patent class dummies (1 patent, 2-5 patents, 6-10 patents, 11-20 patents and 21+ patents).

Table 21: Probability of filing for an environmental patent (firm/year pairs with positive patents - Probit estimates, marginal effects are shown)

All_env	(1)	(2)	(3)	(4)
Polluter	0.0825***	0.0879***	0.0275*	0.0173
	(0.0186)	(0.0189)	(0.0155)	(0.0149)
Polluting_sector	0.0115	0.00918	0.0322*	0.0437**
	(0.0139)	(0.0137)	(0.0170)	(0.0177)
Polluter	0.0819***	0.0878***	0.0239	0.0124
	(0.0187)	(0.0191)	(0.0154)	(0.0146)
Polluting_sector	0.00320	0.000331	0.0287*	0.0417**
	(0.0134)	(0.0132)	(0.0169)	(0.0177)
Pol_waste	(1)	(2)	(3)	(4)
Polluter	0.0254**	0.0259**	0.0149	0.0121
	(0.0110)	(0.0110)	(0.00994)	(0.00945)
Polluting_sector	0.0114	0.0109	0.0188*	0.0225*
	(0.00867)	(0.00858)	(0.0112)	(0.0117)
Polluter	0.0239**	0.0245**	0.0130	0.01000
	(0.0108)	(0.0109)	(0.00964)	(0.00906)
Polluting_sector	0.00905	0.00839	0.0170	0.0210*
	(0.00831)	(0.00820)	(0.0110)	(0.0115)
Renewables	(1)	(2)	(3)	(4)
Polluter	0.0289**	0.0292**	0.00832	0.00331
	(0.0116)	(0.0116)	(0.00894)	(0.00802)
Polluting_sector	-0.00430	-0.00496	0.00608	0.0106
	(0.00765)	(0.00744)	(0.00982)	(0.0102)
Polluter	0.0300**	0.0305**	0.00775	0.00242
	(0.0118)	(0.0119)	(0.00890)	(0.00788)
Polluting_sector	-0.00653	-0.00727	0.00504	0.0102
	(0.00723)	(0.00700)	(0.00969)	(0.0102)
Year d.	-	Yes	Yes	Yes
Macro_reg d.	-	Yes	Yes	Yes
Size (ln(L))	-	-	Yes	Yes
Pavitt d.	-	-	Yes	Yes
Class_patent d.	-	-	-	Yes
N	5694	5694	5694	5694

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.3.2 Data

This chapter uses a dataset containing balance sheet / income statement information on Italian manufacturing firms (AIDA, by Bureau van Dijk) which has been further merged with patent applications to the European Patent Office (EPO). For sake of brevity, refer to Appendix A for further details on the methodology used to match EPO applications and firms in AIDA and for some general descriptive statistics. I further extended the AIDA dataset by identifying the biggest polluting firms and the most emission-intensive sectors[15].

The use of administrative data as an alternative to survey data is a

[15] Refer to the appendix 'Polluting firms and polluting sectors' for further details.

Table 22: Descriptive statistics

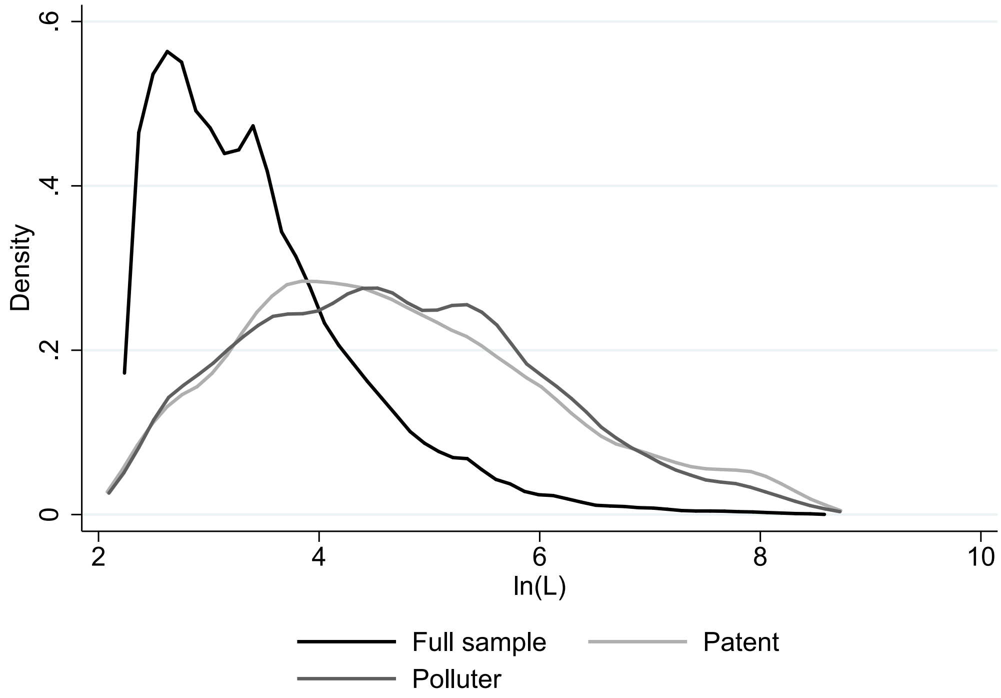
Variable	Mean	Q1	Median	Q3	Min	Max	SD/mean
			Full sample				
Book value	13478	1800	3608	8392	113.7	7795221	5.36
Employees	63.8	15	26	50	10	4985	2.919
Fixed physical assets per empl.	37.41	9.144	22.49	47.64	.6339	472.5	1.228
Value added per empl.	47.21	33.23	41.32	54.21	10.2	237.2	.4831
Age	20.05	11	18	26	0	107	.6629
R&D per empl.	1.937	.08979	.3252	1.198	2.18e-06	529.7	4.193
Perform R&D (d.)	.3184	0	0	1	0	1	1.463
Regional patent stock pc	.539	.3602	.5676	.7812	.01131	.8801	.4869
			Patent sample				
Total patents	2.092	1	1	2	1	44	1.461
Environmental patents (all)	.1507	0	0	0	0	25	4.498
Pollution and waste patents	.03548	0	0	0	0	3	6.057
Renewable energy patents	.04689	0	0	0	0	25	8.895

Table 23: Sectoral distribution of observations (sub-section Nace Rev. 1.1 and Pavitt (1984) taxonomy)

Sector	Full sample	Patent sample	Perc w/pat	Polluter sample	Perc pollut
DA	18245	88	0.48%	348	1.91%
DB	19812	135	0.68%	283	1.43%
DC	8115	67	0.83%	81	1.00%
DD	6212	23	0.37%	78	1.26%
DE	15434	103	0.67%	481	3.12%
DF-DG	11082	520	4.69%	1058	9.55%
DH	14173	465	3.28%	181	1.28%
DI	14461	111	0.77%	849	5.87%
DJ	52915	942	1.78%	2244	4.24%
DK	35990	1843	5.12%	216	0.60%
DL	21657	914	4.22%	187	0.86%
DM	6698	227	3.39%	127	1.90%
DN	18499	256	1.38%	280	1.51%
Scale intensive manuf.	88946	1752	1.97%	3452	3.88%
Science based manuf.	26006	1110	4.27%	1192	4.58%
Specialized suppliers manuf.	42024	2160	5.14%	218	0.52%
Supplier dominated goods	86317	672	0.78%	1551	1.80%
Total	243293	5694	2.34%	6413	2.64%

distinctive feature of the current chapter as opposed to similar articles either in the CDM literature or in the eco-innovation literature. On the one hand, administrative data give more objective and standardized information on firms. Balance sheet and income statement are compulsory and they are compiled according to transparent and standard compulsory criteria by all firms. These information are likely to be generally more reliable than corresponding self-reported information. However, standardization goes together with simplification of available information, limiting substantially the scope of possible empirical analysis as opposed to survey data. Besides missing qualitative information (e.g. inno-

Figure 12: Kernel distribution of log of employees count



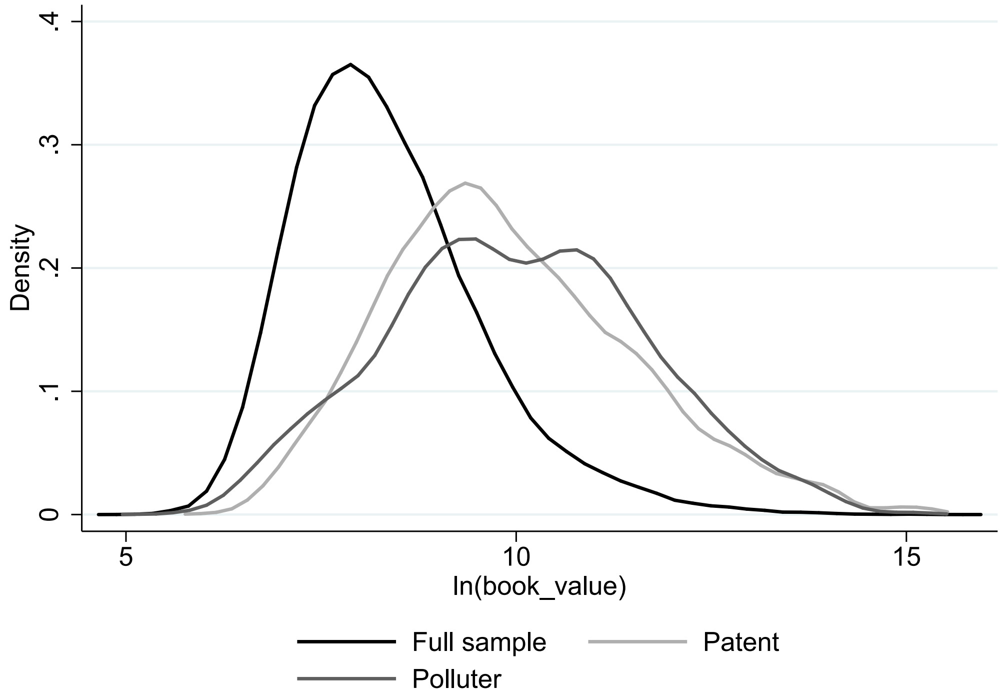
vation strategies, skill composition of labour force, perceptions of market and regulatory conditions), many important monetary variables are not shown in either balance sheet or income statement, such as investment, R&D expenditure (refer to the appendix 'Adjustments to AIDA' for further detail on the R&D asset variable reported in the balance sheet), composition of sales by product for multi-product firms, export and other relevant information. Finally, selection issues in administrative databases are slightly different relative to survey data. In general, firms are selected into an administrative database (e.g. AIDA) according to specific criteria[16] and, in theory, the 'response rate' is expected to be 100 percent[17].

In this chapter I focus on firm/year pairs with more than ten em-

[16] In this chapter I use AIDA TOP which includes, in theory, all registered companies with more than 1.5 million euros of sales and a sample (about 10 percent of population) of registered companies with less than 1.5 million euros of sales. Individual firms are excluded and inactive firms are dropped from the database after four consecutive years of inactivity.

[17] Coverage of balance sheet and income statement information in AIDA is actually not complete, with a quite substantial share of firms with a limited time coverage.

Figure 13: Kernel distribution of log of book value

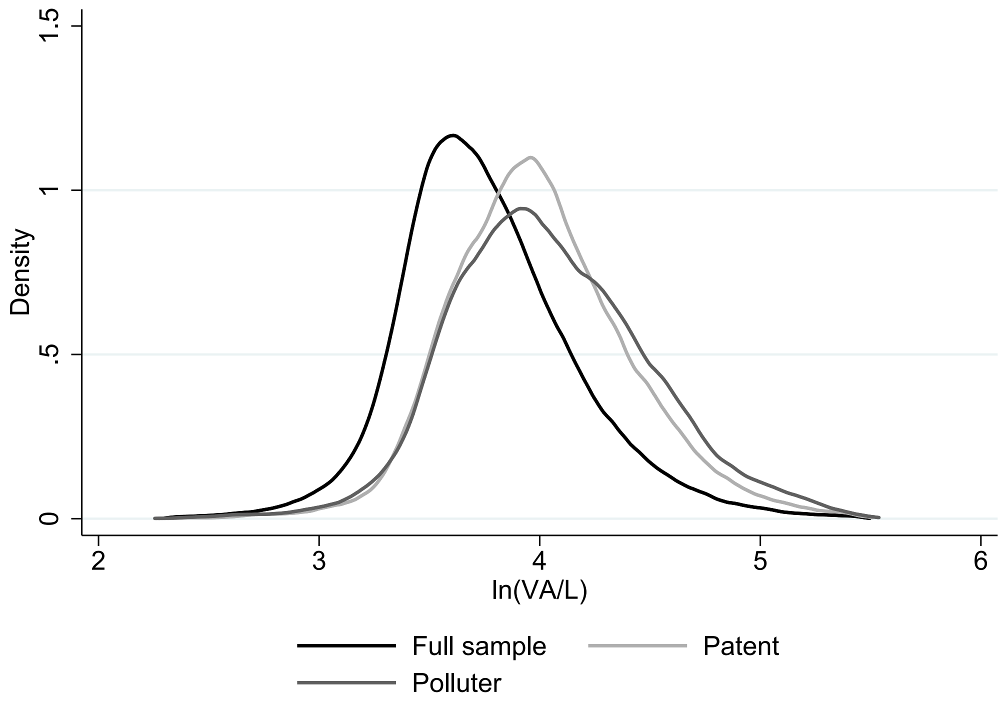


ployees and with less than 5000 employees[18]. Moreover, I removed some observations because of missing values in relevant variables and outlier observations[19]. The classical and extended CDM models are estimated for three distinct samples: (i) full sample; (ii) sub-sample including firm/year pairs with positive patents (patent sample); (iii) sub-sample including polluting firms only (polluter sample). The distribution of observations by sectors for the three different samples is reported in table

[18]Even though results with all firms are very similar to those reported here, the choice of excluding very small and very big firms is motivated by the possibility that very small family companies and huge groups are characterized by extremely different innovation patterns. The thresholds are, however, somehow arbitrary. Finally, results excluding the upper and lower tails of the size distribution of firms are more easily comparable with those of innovation surveys which generally exclude that kind of firms.

[19]Outliers were identified according to the following criteria: sales per employee smaller than 2000 euros or greater than 10000 euros, growth rate of sales greater than 150 percent or smaller than -150 percent, growth rate of employees greater than 150 percent or smaller than 50 percent, growth rate of fixed physical assets greater than 200 percent or smaller than -50 percent, growth rate of labour productivity (value added per employee) greater than 300 percent or smaller than -90 percent, first and last percentile of sales per employee, value added per employee and fixed physical assets.

Figure 14: Kernel distribution of log of value added per employee



23. Table 22 reports some descriptive statistics for the full sample and, for patent statistics, for the patent sample. It is worth noting that just 32 percent of firm/year pairs have positive R&D (71 percent for the patent sample and 53 percent for the polluter sample), which motivates the use of the Heckman selection model to correct for sample selection bias. Another interesting information concerns the distribution of patents: in most of the cases (65 percent) firms just file for one patent per year and several firms (about one quarter) file for just one patent during the considered period. Finally, about 15 percent of considered patents have been classified as 'environmental patents'.

Figures 12, 13 and 14 show, respectively, the kernel distribution (Epanechnikov kernel function) of log of employees count, log of book value and log of value added per employee for the three samples. Firms with positive patents and with big polluting plants are substantially bigger than other firms. Moreover, polluting firms tend to be slightly bigger than firms with patents[20]. Looking at the distribution of labour productivity, I observe a clear evidence that patenting firms and polluting firms are generally more productive than other firms. This is not surprising as regards patenting firms, because I expect that either patents improve productivity through technological improvement and temporary market power and (or) that more productive firms are more likely to file for patents. However, firms with big polluting plants seem to be even more productive, on average, than patenting firms, with a fatter right tail. This evidence, which might seem surprising at a first look, could depend on the peculiar sectoral distribution of polluting firms, especially concentrated in scale-intensive sectors.

EPO patent applications were sorted by priority year. Environmental patents were identified according to their IPC class[21]. I use two different sources of environmentally-relevant IPC classes: the IPC Green Inventory[22] by the World Intellectual Property Organization (WIPO) and the Indicator of Environmental Technology[23] (ENV-Tech Indicator) by the OECD. The selection of environmentally-related IPC classes by the OECD is much narrower relative to the selection by the WIPO[24]. In the sample of matched patent applications used in the current chapter, environmental patent applications identified according to the ENV-Tech Indicator (OECD) were about one third of environmental patent applications identified according to the IPC Green Inventory (WIPO). Moreover, environmentally relevant IPC classes identified by the WIPO already cover most of the IPC classes selected by the Env-Tech Indicator. In this chapter I use three different selections of environmental patent applications: (i)

[20]This is not surprising given that firms listed on the EPER and E-PRTR need to pass certain thresholds related to the size of their production plants.

[21]The article by Lanjouw and Mody (1996) was an early effort to identify environmental patents to investigate their international diffusion. Recent empirical analysis based on environmental patents (among others, Rave et al (2011) and Johnstone et al (2010)) combined both IPC class selection and keywords search in patent abstracts and/or titles to identify environmental innovations. The approach of focusing on IPC classes only is likely to underestimate the number of environmental patents, thus giving rise to more conservative estimates.

[22]http://www.wipo.int/classifications/ipc/en/est/

[23]http://www.oecd.org/environment/innovation/indicator/

[24]I excluded those IPC classes referring to nuclear energy technologies.

environmental patents identified as environmentally relevant by either WIPO or OECD (`aggr`); (ii) environmental patents in the field of renewable energy (`renew`); (iii) environmental patents in the field of waste and pollution management (`poll`).

Patent stocks per capita at the regional level were computed by using data on EPO patent applications count (based on applicants) reported in the OECD Database 'Patents by region' (OECD, 2011a). The stock is computed by means of a perpetual inventory method (the initial year was set to 1990 and the yearly geometric depreciation rate was set to 15 percent as suggested by Hall (2006)).

The rest of the variables used in the current chapter comes from the AIDA database. In absence of data on investments, I used book-value fixed capital assets as a proxy for capital stock while to approximate R&D effort I used capitalized R&D available in the section 'intangible assets' of the balance sheet. I retrieved total assets (book value of the firm) from the balance sheet and value added and sales from the income statement. Employees count, firm's age, location and main sector of activity are provided by Bureau van Dijk. Additional information on manipulation and definition of these variable is reported in the appendix 'Polluting firms and polluting sectors'.

4.4 Results

In this section I discuss the results of the econometric analysis, with a specific focus on drivers and productivity effect of environmental patents. All estimates include sector (Pavitt's taxonomy), macro-region and year dummies: results for these variables are not shown but remain available upon request. For all estimates based on the full sample, standard errors have been clustered by firm, while for the patent and polluter subsamples, standard errors have been clustered by sector (Nace Rev. 1.1 2-digit), region (NUTS-2) and year. Moreover, results including bootstrap standard errors to account for the fact that R&D (second step) and patents (third step) are estimated values are very similar.

Table 24: First step: R&D equation

	Full sample		Patent		Polluter	
Dep: ln(R&D/L)	OLS	Heckman	OLS	Heckman	OLS	Heckman
ln(L)	-0.0992***	-0.685***	-0.180***	-0.433***	-0.103***	-0.474***
	(0.0142)	(0.0239)	(0.0258)	(0.0340)	(0.0395)	(0.0460)
Market_sh	0.855***	2.516***	0.995*	2.801***	0.215	0.506*
	(0.283)	(0.343)	(0.513)	(0.634)	(0.261)	(0.299)
ln(K/L)	0.128***	0.00818	0.139***	0.102***	0.328***	0.184***
	(0.0126)	(0.0138)	(0.0337)	(0.0380)	(0.0420)	(0.0456)
Constant	-1.319***	3.230***	-0.386*	1.827***	-2.215***	1.781***
	(0.0797)	(0.163)	(0.222)	(0.269)	(0.270)	(0.354)
Perform R&D	Full sample		Patent		Polluter	
ln(L)		0.143***		-0.0758		0.0905**
		(0.0114)		(0.0515)		(0.0357)
Market_sh		-1.847***		-2.081***		-0.350**
		(0.212)		(0.328)		(0.165)
ln(K/L)		-0.0234***		-0.0461**		0.0154
		(0.00585)		(0.0210)		(0.0241)
ln(book_value)		0.426***		0.378***		0.234***
		(0.00987)		(0.0458)		(0.0315)
Age>10		0.0212*		0.0683		-0.0834**
		(0.0126)		(0.0486)		(0.0388)
Constant		-4.658***		-2.742***		-2.884***
		(0.0568)		(0.243)		(0.157)
Chi sq		1235.0		239.2		284.1
sigma		2.407		2.193		2.492
rho		-0.731		-0.808		-0.803
lambda		-1.758		-1.771		-2.002
Chi sq (rho)		1112.1***		224.1***		201.6***
Log likelihood	-162806.1	-283964.1	-8257.8	-11200.0	-7206.1	-11060.6
N	77470	243293	4052	5694	3415	6413

Standard errors in parentheses; * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 25: Second step: Patent equation (All patents)

	Full	Patent	Polluter
ln(R&D/L)*	0.262	0.356***	0.904**
	(0.199)	(0.0728)	(0.412)
ln(L)	1.229***	0.456***	1.290***
	(0.131)	(0.0284)	(0.179)
ln(reg_pat_stock_pc)	0.112	0.136**	0.0644
	(0.142)	(0.0578)	(0.401)
Constant	-9.427***	-2.417***	-9.728***
	(0.953)	(0.356)	(2.293)
Chi sq	3202.5	1311.9	419.2
alpha	10.25	0.228	10.95
Log likelihood	-29051.5	-9631.0	-2213.9
N	243293	5694	6413

Standard errors in parentheses
* $p<0.1$, ** $p<0.05$, *** $p<0.01$

4.4.1 Classical CDM model

Before moving to the results on eco-innovation patterns, it is worth discussing the outcome of the classic CDM model. The first step (R&D equation - table 24) shows a significant selection bias (the correlation of disturbances between first and second step of the Heckman, ρ, is negative

Table 26: Third step: productivity equation (All patents)

Dep: ln(VA/L)	Full	Patent	Polluter
ln(K/L)	0.118***	0.101***	0.187***
	(0.00132)	(0.00685)	(0.00873)
ln(patent/L)*	0.381***	0.431***	0.114***
	(0.0108)	(0.0669)	(0.0335)
ln(L)	0.00595***	0.320***	0.0517***
	(0.00191)	(0.0453)	(0.00541)
Constant	6.368***	3.925***	3.961***
	(0.0883)	(0.0863)	(0.267)
R sq	0.211	0.182	0.322
F	1664.3	57.44	123.4
N	243293	5694	6413

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

and significantly different from zero) in OLS estimates for all samples. The probability of reporting positive R&D intangible assets is positively and significantly related to firm size both in terms of employees count (insignificant for the patent sample only) and book value (log of total assets). The positive effect of firm size on the probability of performing R&D is a standard result in CDM models: bigger firms are more likely to be willing to incur the initial sunk costs of R&D activities, they have easier access to credit and they are more capable of bearing the risk related to R&D investments whose returns are highly uncertain. Moreover, the book value of the firm measured with total assets is a key criterion for the type of reporting system chosen by the firm (refer to the appendix 'Adjustments to AIDA' for further details). As expected, R&D intensity is negatively related to firm size (log of employees count), with big firms characterized by a relatively lower R&D intensity than small firms[25]. Market share is negatively related to the probability of performing R&D while it affects its intensity positively. On the one hand, firms holding a dominant market position have little incentive to innovate and they prefer to defend their dominant position rather than exploring new markets or changing their production technology. This idea is in line to the 'creative destruction' theory (Aghion and Howitt, 1992) according to which technological leaders have no incentive to further innovate because they would destroy their own current rents. On the other hand,

[25] For a detailed discussion on the relationship between firm size and R&D and patents refer to Cohen and Klepper (1996).

however, once they decide to innovate, there is strong incentive to exploit the current large customer base (product innovations) by introducing new products. When considering process innovations, the incentive depend on the fact that the expected unitary cost savings is spread on a large scale of production. Capital intensity is generally negatively related to the probability of performing formal R&D (except for polluting firms) but affects R&D intensity positively (not significantly for the full sample). The negative effect on the binary choice about R&D might be explained by the greater incentive to increase physical assets in absence of knowledge assets, which require substantial initial sunk costs. However, in case a firm invests in knowledge capital (R&D), complementarity between knowledge and physical assets seems to arise. Finally, older firms have higher probability to perform R&D in the full and patent (although insignificantly) samples while older polluting firms have lower propensity to perform R&D relative to younger polluting firms.

The second step (table 25) has been performed by including the predicted log of R&D intensity into a patent equation estimated with a Negative Binomial regression (NB2 version, with the variance of the disturbance expressed as a quadratic function of the conditional mean). I report estimated coefficients which can be interpreted as semi-elasticities for logarithmic independent variables (expected relative changes in patent applications count for a relative change in the independent variable) and, for dummy variables (once exponentiated) as relative change in patent applications count when the variable switches from zero to one (Cameron and Trivedi, 1998)[26].

R&D intensity affects positively innovation output expressed as patent applications count, the effect being insignificant in the full sample only. The absence of significance for the full sample may depend on the extremely high proportion of observations with no patent applications in the full sample (97.7 percent) relative to the polluter sample (93 percent of observations with no patent) and the patent sample (all observa-

[26]More detailed results such as marginal effects are not reported but available upon request.

tions with positive patents)[27]. The relative sensitivity to R&D intensity is greater for the polluter sample than for the patent sample. This result may relate to a greater effect of R&D intensity on the extensive margin (probability of patenting) relative to the extensive margin (number of patents).

Firm size plays a positive role also for innovation output conditional on innovation input (R&D) due to the generally higher propensity to file for patents for big than for small innovative firms. A greater sensitiveness to firm size for the full and polluter sample than for the patent sample may also depend on different effects of firm size on intensive and extensive margins. Firm size affects patent propensity very strongly (intensive margin) while the elasticity of patent counts (extensive margin) with respect to firm size, conditional on patenting (patent sample), is lower than one, with small patenting firms holding, on average, more patents per employee relative to big firms. Finally, regional patent stock per capita turns out to be positively related to firms' innovation output[28].

The third step (table 26) contains the predicted expected patent applications count as explanatory variable (more precisely, the log of expected patent applications per employee). Predicted innovation success affects positively and significantly labour productivity, the effect being greater in the full and patent samples (elasticity of about 0.4) relative to the polluter sample (elasticity of about 0.1). This means that the value of each patent in terms of productivity improvement in polluting firms is about one quarter of the value of each patent for the full sample. This great divergence may partly depend on the bias of polluting firms towards technological domains characterized by lower productivity potential such as environmental innovations, thus confirming the concerns about the possibility to observe crowding out.

The elasticity of value added per employee with respect to capital intensity (fixed physical assets per employee) is positive and significant.

[27] R&D is insignificant for the full sample also when considering the choice to file for patents as a dichotomous choice (yes/no).

[28] The investigation of knowledge spillovers is not the core of the current analysis. A proper investigation would require more refined measures of local knowledge stocks such as spatially weighted regional stocks.

Under the assumption of perfect competition and constant returns to scale, the elasticity of labour productivity with respect to capital intensity should represent the capital share of labour income[29]. Capital share tends to be substantially greater for polluting firms relative to other firms probably due to the concentration of polluting firms in scale-intensive sectors. Firms in the patent sample are characterized by strong increasing returns to scale (the log of employee counts is significantly greater than zero and its magnitude is quite relevant). This result, not found for the other samples, might be caused by the temporary market power assigned to patent applicants as a consequence of IPRs protection.

Most of the results of the classical CDM model confirm prior expectations and give a reasonable description of innovation patterns of Italian manufacturing firms.

4.4.2 Extended CDM model

Results for the second step (Patent equation) regarding environmental patents are reported in table 27. The effect of the various explanatory variables on non-environmental patents ('no_env' column) are very similar to those estimated for total patent applications in the classical CDM model for all samples.

Looking at the equation for environmental patents I observe a great heterogeneity across samples. Similarly to non-environmental patents, R&D intensity has a significant positive effect for the patent and polluter samples only, being positive but insignificant for the full sample. This asymmetry is in line with the one found when considering all patents. Moreover, also in the case of environmental patents, the sensitivity to R&D intensity is greater for the polluter sample than for the patent sample. However, both in the patent and polluter sample, the sensitivity to R&D is greater for environmental patents than for non-environmental patents (more than double). Patented environmental innovations seem

[29]Firm-level estimates of production functions tend underestimate the share of capital stock relative to national accounting measures in which capita receive about one third of national income. This bias is generally related to attrition problems and to measurement errors in the capital stock (Eberhardt and Helmers, 2010).

Table 27: Second step: Patent equation (all environmental patents)

	Full sample		Patent		Polluter	
	No_env	Env	No_env	Env	No_env	Env
ln(R&D/L)*	0.248	0.272	0.328***	0.812***	0.947**	2.421***
	(0.200)	(0.319)	(0.0726)	(0.201)	(0.425)	(0.892)
ln(L)	1.222***	1.215***	0.446***	0.612***	1.315***	1.933***
	(0.131)	(0.207)	(0.0286)	(0.0774)	(0.180)	(0.454)
ln(reg_pat_stock_pc)	0.139	-0.0768	0.159***	-0.213	-0.00723	1.216**
	(0.148)	(0.215)	(0.0602)	(0.155)	(0.407)	(0.497)
Poll (air)		-0.306		-0.427		-0.206
		(0.520)		(0.288)		(0.298)
Poll (water)		0.0772		-0.365		0.102
		(0.553)		(0.224)		(0.311)
Poll (haz_waste)		0.613		0.247		-1.050
		(0.443)		(0.281)		(0.695)
Poll (no_haz_waste)		-0.0580		0.223		0.550*
		(0.457)		(0.317)		(0.318)
Poll (other)		0.0788		-1.265*		-1.330
		(0.654)		(0.761)		(0.994)
Polluting_sect		-0.620***		0.132		-2.250***
		(0.207)		(0.163)		(0.461)
Constant	-9.629***	-10.70***	-2.578***	-4.012***	-9.560***	-20.34***
	(0.986)	(1.482)	(0.373)	(0.935)	(2.329)	(3.687)
Chi sq	3152.1	701.4	1164.2	207.3	430.3	2013.8
alpha	10.43	28.38	0.245	3.898	11.01	14.40
Log likelihood	-27620.1	-3807.6	-9457.4	-2252.6	-2086.0	-428.6
N	243293	243293	5694	5694	6413	6413

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 28: Third step: productivity equation (All environmental patents)

	Full sample		Patent		Polluter
Dep: ln(VA/L)	(1)	(2)	(1)	(2)	(1)
ln(K/L)	0.117***	0.115***	0.0975***	0.0975***	0.198***
	(0.00133)	(0.00133)	(0.00724)	(0.00708)	(0.00836)
ln(no_env/L)*	0.420***	0.433***	0.328***	0.303***	0.0676***
	(0.0131)	(0.0132)	(0.0807)	(0.0802)	(0.0253)
ln(env/L)*	-0.0308***	-0.0455***	0.0824**	0.0733**	-0.0152***
	(0.00552)	(0.00568)	(0.0347)	(0.0351)	(0.00534)
polluter ×		-0.0183*		-0.0383**	
ln(env/L)*		(0.0101)		(0.0183)	
polluter		-0.0521		-0.228	
		(0.0997)		(0.139)	
ln(L)	0.00510***	0.000740	0.307***	0.280***	0.0421***
	(0.00194)	(0.00196)	(0.0465)	(0.0472)	(0.00493)
Constant	6.409***	6.377***	4.156***	4.111***	3.439***
	(0.0924)	(0.0925)	(0.128)	(0.130)	(0.221)
Net effect		-0.0639***		0.0340	
for polluter		(0.0109)		(0.0396)	
R sq	0.211	0.214	0.183	0.184	0.322
F	1564.6	1413.6	54.97	55.88	113.0
N	243293	243293	5694	5694	6413

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to be more R&D-intensive than other innovations probably due to their greater complexity and pervasiveness (Cainelli et al, 2011). Innovation success in the field of environmental technologies is slightly more sensitive to firm size relative to other technologies, especially so for polluting

Table 29: Third step: productivity equation (separate effect for 'env' and 'no_env' - all environmental patents)

	Full sample		Patent		Polluter	
Dep: ln(VA/L)	(1)	(2)	(1)	(2)	(1)	(2)
ln(K/L)	0.1184***	0.1215***	0.1028***	0.1028***	0.194***	0.211***
	(0.0013)	(0.0013)	(0.0067)	(0.0072)	(0.0082)	(0.0071)
ln(no_env/L)*	0.3868***		0.4114***		0.0688***	
	(0.0112)		(0.0672)		(0.0253)	
ln(env/L)*		0.0266***		0.1449***		-0.0154***
		(0.0049)		(0.0289)		(0.0053)
ln(L)	0.0052***	0.0356***	0.3075***	0.1273***	0.046***	0.0357***
	(0.0019)	(0.0017)	(0.0455)	(0.0196)	(0.0048)	(0.0042)
Constant	6.4556***	3.5538***	3.9348***	4.0052***	3.610***	2.885***
	(0.0926)	(0.0514)	(0.0906)	(0.1254)	(0.2071)	(0.0685)
R sq	0.2109	0.2021	0.1814	0.1792	0.3215	0.3217
F	1662.51	1589.53	56.64	55.88	123.09	120.94
N	243293	243293	5694	5694	6413	6413

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 30: Second step: Patent equation (pollution and waste patents)

	Full sample		Patent		Polluter	
	No_env	Env	No_env	Env	No_env	Env
ln(R&D/L)*	0.267	-0.279	0.352***	0.329	1.006**	0.411
	(0.201)	(0.364)	(0.0731)	(0.294)	(0.420)	(1.132)
ln(L)	1.236***	0.731***	0.457***	0.293***	1.345***	0.834*
	(0.132)	(0.217)	(0.0285)	(0.109)	(0.181)	(0.502)
ln(reg_pat_stock_pc)	0.108	0.417	0.134**	0.243	-0.0213	3.113**
	(0.143)	(0.372)	(0.0590)	(0.283)	(0.411)	(1.314)
Poll (air)		1.521*		0.900**		1.103***
		(0.844)		(0.453)		(0.370)
Poll (water)		0.550		0.344		0.150
		(0.611)		(0.411)		(0.449)
Poll (haz_waste)		-0.649		-0.654		-1.710***
		(0.718)		(0.469)		(0.657)
Poll (no_haz_waste)		-0.122		0.188		0.777
		(0.592)		(0.540)		(0.482)
Poll (other)		2.029***		1.663*		1.830*
		(0.756)		(0.949)		(1.087)
Polluting_sect		-0.290		0.367		-2.496***
		(0.309)		(0.270)		(0.776)
Constant	-9.457***	-12.47***	-2.430***	-6.309***	-9.593***	-25.98***
	(0.962)	(2.206)	(0.361)	(1.665)	(2.336)	(7.394)
Chi sq	3193.5	383.8	1297.4	55.89	448.3	3907.3
alpha	10.36	46.33	0.238	6.530	11.21	0.825
Log likelihood	-28627.1	-1313.1	-9605.4	-839.7	-2173.2	-122.8
N	243293	243293	5694	5694	6413	6413

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

firms. Crossing the hurdle is more difficult for environmental patents than for other patents and bigger firms tend to be much more intensive in environmental patents than small firms. As far as local knowledge stock is concerned, no significant effect is found for the full and patent sample while polluting firms rely to a great extent on local knowledge for their eco-innovation success. While, on average, polluting firms and sec-

Table 31: Third step: productivity equation (pollution and waste patents)

	Full sample		Patent		Polluter
Dep: ln(VA/L)	(1)	(2)	(1)	(2)	(1)
ln(K/L)	0.115***	0.114***	0.101***	0.0994***	0.194***
	(0.00136)	(0.00135)	(0.00689)	(0.00680)	(0.00823)
ln(no_env/L)*	0.598***	0.612***	0.358***	0.291***	0.0687***
	(0.0266)	(0.0283)	(0.0672)	(0.0690)	(0.0243)
ln(env/L)*	-0.0797***	-0.0873***	0.0671***	0.0941***	0.00434
	(0.00899)	(0.00977)	(0.0246)	(0.0266)	(0.00385)
polluter × ln(env/L)*		0.0562***		-0.0428***	
		(0.0114)		(0.0147)	
polluter		0.757***		-0.321**	
		(0.134)		(0.138)	
ln(L)	-0.0155***	-0.0202***	0.325***	0.296***	0.0473***
	(0.00302)	(0.00303)	(0.0467)	(0.0473)	(0.00498)
Constant	7.241***	7.286***	4.160***	4.234***	3.654***
	(0.132)	(0.136)	(0.136)	(0.141)	(0.215)
Net effect for polluter		-0.0311***		0.0513**	
		(0.0143)		(0.0248)	
R sq	0.213	0.214	0.183	0.186	0.322
F	1564.1	1413.2	56.70	58.26	118.1
N	243293	243293	5694	5694	6413

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 32: Second step: Patent equation (renewable energy patents)

	Full sample		Patent		Polluter	
	No_env	Env	No_env	Env	No_env	Env
ln(R&D/L)*	0.265	0.108	0.343***	1.126***	0.870**	3.106**
	(0.200)	(0.366)	(0.0724)	(0.287)	(0.417)	(1.321)
ln(L)	1.233***	1.023***	0.454***	0.620***	1.282***	2.091***
	(0.132)	(0.222)	(0.0284)	(0.109)	(0.179)	(0.662)
ln(reg_pat_stock_pc)	0.135	-0.563*	0.158***	-0.745***	0.0393	1.539
	(0.144)	(0.300)	(0.0584)	(0.245)	(0.407)	(0.945)
Poll (air)		0.0137		-0.0280		0.312
		(0.481)		(0.344)		(0.422)
Poll (water)		-0.0411		-0.213		-0.222
		(0.611)		(0.230)		(0.366)
Poll (haz_waste)		0.236		-0.345		15.48
		(0.497)		(0.356)		-
Poll (no_haz_waste)		0.125		0.544		0.406
		(0.525)		(0.387)		(0.446)
Poll (other)		-18.76***		-19.56***		1.855
		(0.640)		(0.897)		(1.192)
Polluting_sect		-0.716**		0.000406		-0.260
		(0.357)		(0.360)		(1.159)
Constant	-9.607***	-8.051***	-2.566***	-2.149	-9.591***	-42.82***
	(0.966)	(1.963)	(0.359)	(1.587)	(2.314)	(5.788)
Chi sq	3183.7	1297.7	1279.8	664.0	405.5	-
alpha	10.23	54.95	0.232	8.721	11.19	1.115
Log likelihood	-28606.7	-1457.4	-9573.2	-933.4	-2184.7	-129.5
N	243293	243293	5694	5694	6413	6413

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

tors have greater propensity to patent in fields related to environmental technologies (see table 21), there is no clear specific pattern when considering the status of the firms in terms of type of 'pollution' (air, water, hazardous and non-hazardous waste). The only significant coefficients

Table 33: Third step: productivity equation (renewable energy patents)

	Full sample		Patent		Polluter
Dep: ln(VA/L)	(1)	(2)	(1)	(2)	(1)
ln(K/L)	0.118***	0.116***	0.101***	0.0985***	0.186***
	(0.00134)	(0.00133)	(0.00685)	(0.00674)	(0.00853)
ln(no_env/L)*	0.354***	0.347***	0.432***	0.394***	0.114***
	(0.0112)	(0.0109)	(0.0692)	(0.0726)	(0.0338)
ln(env/L)*	-0.0119**	-0.0181***	0.00603	0.0130	0.00403***
	(0.00555)	(0.00472)	(0.0125)	(0.0140)	(0.00118)
polluter × ln(env/L)*		0.0131*		-0.00765	
		(0.00737)		(0.0138)	
polluter		0.266***		-0.00812	
		(0.0849)		(0.127)	
ln(L)	0.00654***	0.00293	0.325***	0.302***	0.0507***
	(0.00192)	(0.00193)	(0.0452)	(0.0467)	(0.00528)
Constant	6.046***	5.935***	3.974***	3.975***	4.030***
	(0.126)	(0.114)	(0.0956)	(0.0956)	(0.271)
Net effect for polluter		-0.0050		0.0053	
		(0.0076)		(0.0151)	
R sq	0.211	0.213	0.182	0.184	0.323
F	1565.4	1413.8	54.38	51.92	116.4
N	243293	243293	5694	5694	6413

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

are found for ‘other polluting firms’ (with an average lower level of environmental patenting in the patent sample) and for ‘non-hazardous waste firms’ (with an average higher level of environmental patenting in the polluter sample). Finally, it is interesting to note that while, conditional on patenting (patent sample), firms in polluting sectors have a higher (but again insignificant) propensity to file for environmental patents relative to firms in other sectors, the propensity to file for environmental patents in the full and polluter sample for firms in polluting sectors is much lower due to the general lower propensity to patent for firms operating in these sectors.

Looking a the productivity equations (table 28), I observe a strong positive effect of non-environmental patents on labour productivity, with a coefficient which is slightly greater than that found for all patents in the classical CDM (table 26) for the full sample while it is slightly smaller for the patent and polluter samples. On the other hand, looking at the productivity effect of environmental patents, I observe a significant and negative effect on labour productivity for the full and the polluter samples while the effect turns out to be positive and significant (but substantially smaller than the effect of non-environmental patents) for the patent sample[30]. These results, especially the negative signs, are a possible evidence that environmental innovations tends to crowd out resources from other innovations which are more profitable, at least in the short run. The small positive effect of environmental patents for the patent sample could be interpreted as actual crowding out only if the average cost, for example expressed in terms of R&D inputs, of each additional environmental patents is substantially smaller than the average cost of non-environmental patents. This is not likely to be the case. By computing average marginal effects for the two patent equations I estimate that, on average, an increase of 10 percent in R&D intensity is expected to increase the number of non-environmental patents by 0.06 and the number of environmental patents by 0.0123. Environmental patents turn out to be at the same time 'more costly' (in terms of R&D input) and less remunerative (in terms of better labour productivity) than non-environmental patents, which suggests the possible presence of a crowding out effect also for the patent sample. Finally, polluting firms within the full and patent sample are more strongly characterized by crowding out than other firms[31]. Polluting firms are expected to face more stringent environmental policies than other firms. This asymmetry in the policy environment forces them to bias their innovation patterns towards less productivity-enhancing innovations (eco-innovations) to reduce compliance costs[32]. However, this differential effect I find for polluting firms could be partly

[30]By construction, predicted patent intensity in terms of environmental and non-environmental patents are highly correlated. When including predicted environmental patents only and non-environmental patents only (table 29), I observe that environmental patents affect productivity positively and significantly in the full and patent samples and negatively and significantly in the polluter sample. These results are slightly different from those found in table 26 for environmental patents where the estimated effect of environmental patents is systematically smaller. This could be related to the fact that environmental patents do not affect productivity negatively (full sample) or insignificantly (patent sample) per se but only when controlling for general innovation success of the firm with the only exception of the polluter sample, where the effect is always negative.

[31]The net effect for polluting firms in the patent sample turns out to be still positive but insignificant.

[32]The reasoning about compliance costs is partly weakened by the use of patents as a measure of environmental innovations. Patent information say nothing about the actual adoption of innovations. It is thus plausible that a substantial amount of environmental patents is filed by specialized suppliers of 'green' technologies and that the underlying innovation will be employed (i.e. adopted) by other firms to reduce compliance costs.

related to smaller average returns to innovation output (considering all patents) for polluting firms (refer to table 26) relative to other firms.

Estimates of the second and third steps have been done on subsets of environmental patents, namely pollution and waste management patents (tables 30 and 31) and patents related to renewable energy technologies (tables 32 and 33). Pollution and waste management patents seem to be less 'R&D intensive' than other types of environmental patents, with predicted R&D intensity being always insignificant, as well as less sensitive to firm size in all samples. Differently from the full set of environmental patents, I found a very strong bias towards pollution and waste patents for firms with a high level of air pollution while local knowledge stock is a relevant input for environmental innovations in polluting firms only. Looking at the productivity equation, patents in the fields of pollution and waste have an effect which is similar to that found for the full set of environmental patents with two notable differences. First, the crowding out in the full sample is now mitigated rather than exacerbated for polluting firms. Second, the effect in the polluter sample is now insignificant. The reduction of compliance costs due to environmental innovations may partly mitigate the crowding out effect for polluting firms with particularly emission and waste intensive production plants.

Finally, looking at patents in renewable energy technologies, their sensitivity to R&D intensity and firm size is very similar to that one estimated for the full set of environmental patents. The productivity effect of renewable energy patents is negative or insignificant (positive, significant but negligible for the polluter sample). This result is somehow striking, especially when considering the recent rapid growth of the markets for this kind of technologies and by policy measures aimed at favouring the diffusion and adoption of these technologies. Moreover, renewable energy technologies are generally developed by specialized suppliers rather than by polluting firms, thus weakening the potential to give rise to crowding out effects. Part of the explanation could be attributable to the fact that the markets for renewable energy are still very fragmented, uncertain and underexploited.

To conclude this section, some comment is needed on the generality and scope of the results. All results refer to static and short-term relationship, with no consideration of long run effects of R&D on innovation output and, more importantly, of innovation output on productivity. In the specific case of environmental technologies, it could be the case that static crowding out is counterbalanced by long run positive effects on productivity, especially because the market for environmental technologies is a new market with great potentials for growth.

4.5 Concluding remarks

In this chapter I investigate innovation patterns of Italian manufacturing firms, with a specific focus on determinants and productivity effects of environmental innovations. The CDM model describes innovation patterns coherently with expectations. Focusing on environmental innovations, there is evidence of a systematic difference in the effect of usual drivers of innovation output relative to other innovations and a significant bias for environmental innovations by polluting firms and sectors. Moreover, environmental innovations systematically differ from other innovations in their effect on firm's productivity. Environmental innovations either guarantee a return which is substantially lower than that of non-environmental innovations or they slightly reduce labour productivity. This result, coupled with constrained financial resources to be devoted to R&D activities, is a possible evidence of crowding out of environmental innovations relative to non-environmental innovations. It is important to stress that the evidence of crowding out refers to short term indicators of productivity. It is reasonable to assume, however, that the positive effects of policy-induced environmental innovations on competitiveness (and possibly measured productivity) predicted by the 'strong' version of the Porter Hypothesis (Porter and van der Linde, 1995) will show up in the medium-long run because they mainly depend on early-mover advantages of eco-innovators and on the creation of new markets for 'green' technologies.

Further research should be carried out to build a coherent theoretical framework in order to identify the channels through which crowding out occurs. Moreover, these results, based on administrative data and a very reduced set of explanatory variables, should be confirmed by similar models based on more comprehensive data sources such as innovation surveys (e.g. the Community Innovation Survey or other national or regional innovation surveys) for which is generally possible to consider both the creation and the adoption of innovations by firms.

Polluting firms and polluting sectors

The AIDA database has been further extended with information on polluting plants and on pollution-intensive sectors. Italian large polluting plants are reported by the EPER (European Pollutant Emission Register) and the E-PRTR (European Pollutant Release and Transfer Register) registers. The EPER has been introduced by the IPCC (Integrated Pollution Prevention and Control) Directive (96/61/CE). EPER includes all facilities and plants above a certain threshold of air or water pollution. The year of reference is 2006. The E-PRTR substituted the EPER register starting from the year 2007 onwards. The E-PRTR complements information on large polluting plants (used in this chapter to identify polluting firms) with diffused emission sources with great details. Differently from the EPER, the E-PRTR includes waste-intensive plants.

To identify pollution intensive sectors, I used information on 18 different types of air emissions[33] reported at the 2-digit Nace level by the NAMEA (National Accounting Matrix including Environmental Accounts) dataset for Italy (coverage: 1990-2008). I identified as pollution intensive sectors those sectors for which the yearly emissions intensity (emissions per monetary output) for any type of emissions ranked at least fifth for at least five times. I identified as polluting intensive sectors the following 2-digit Nace codes: 15 (Food products and beverages), 16 (Tobacco products), 17 (Textiles), 20 (Wood products), 21 (Pulp and paper), 23 (Coke, refined petroleum and nuclear fuel), 26 (Other non-

[33]Carbon dioxide, N2O, methane, NOx, SOx, ammonia, NMVOC, carbon oxide, particulate matter (<10 micron and <2.5 micron) and a series of heavy metals.

metallic products), 27 (Basic metals), 30 (Machinery and computers) and 37 (Recycling).

Adjustments to AIDA

I deflated firm-level value added to 2000 prices according to a 2-digit value added deflator (Istat). Fixed physical assets, total assets and R&D were deflated to 2000 prices according to a 2-digit fixed asset deflator (Istat). Market share was computed as the share of firm's reported sales relative to total sales for firms in AIDA in the same 3-digit Nace sector. This is a rough measure because it does not consider either multi-product firms, whose market share is probably overestimated, and the competition by firms not included in AIDA (on average smaller than firms included in the database), leading to a general overestimation. Due to the lack of yearly investment data in AIDA, I use the balance sheet value of fixed physical assets as a measure of capital stock. Finally, my R&D variable consists in the amount of capitalized R&D expenditure which is reported within intangible assets. According to the Italian law, the capitalization of R&D expenditure is voluntary and possible only when the utility of the investment is expected to last for more than one year. Moreover, firms satisfying a combination of requirement correlated to firm size, are allowed to file a reduced-form balance sheet with no separation of R&D assets within the broader category of intangible assets. This problem of censoring, combined with the usual issue of sample selection in reporting positive R&D, is likely to harm seriously the reliability of the R&D intensity variable. Possible measurement errors might result in over-estimated standard errors in the R&D equation.

Appendix A

Matching of PATSTAT Applications to AIDA Firms

A.1 Introduction

The use of patent data as a measure of innovation output, as opposed to other measures of innovation input such as R&D expenditure, has been proposed by economists since the late 60s (Comanor and Scherer, 1969). The advantages of patent data as output measures as opposed to other proposed measures are manifold as well as their limitations[1]. Among other, applied economic literature on innovation patterns used significant inventions (Pavitt (1984) investigates the sectoral patterns of technical change by using a collection of about 2,000 significant innovations in Britain since 1945), share of innovative products (either new to the market or new to the firm) sales at the firm level (Crepon et al, 1998; Griffith et al, 2006), binary measures at the micro level such as the introduction of process or product innovations (Griffith et al, 2006)[2]. As opposed to these

[1]For a detailed review of the literature the relevance of patent statistics refer to Pavitt (1985), Griliches (1990) and OECD (2009).

[2]The use of innovative sales and binary measures has been favoured by the inclusion of these measures into the questionnaire of the various waves of the Community Innova-

alternative measures of innovation output, patent data have the advantage of being collected for the whole population of patents and they are an objective measures, with no space for biases arising from self reported measures.

An issue related to the use of patent data regards their integration with other data sources at the micro level. Patent data are collected for legal and administrative purposes which exempt them from specific requirement on the collection. This lack of standardization in the collection of patent data poses a number of issues when trying to use them for statistical analysis. The main problems are: (i) the absence of a unique identifier for applicants and inventors; (ii) the presence of typing mistakes in textual fields such as inventor's or applicant's name or location. These issues, coupled with the huge amount of patents, inventors and firms database, increase the cost[3] of using patent data at the micro level[4].

A first systematic attempt to integrate patent data at the firm level with other micro data sources has been performed by the NBER (National Bureau of Economic Research) within the Productivity program from 1978 through 1988 (Bound et al, 1984; Hall et al, 1988). The aim of the project was to build an integrated database on U.S. publicly trade manufacturing firms with information on balance sheet, income statement, R&D and patent applications in order to investigate a variety of issues related to innovation patterns, productivity and firms' value at the micro level. Starting with a panel of about 2,600 large manufacturing firms available in Compustat, they matched about 300k patent applications to the USPTO (United States Patent and Trademark Office) for the period 1965 through 1981[5]. They combined names harmonization with manual matching aiming at minimizing false positive and false negative

tion Survey (CIS). For a discussion on the use of CIS data in microeconometric analysis of innovation patterns refer to Mairesse and Mohnen (2010).

[3]Standardization, disambiguation and matching of patent data are source of a variety of costs for the researcher: (i) monetary costs (e.g. salary for research assistants); (ii) time lags between the start of the research project and the moment in which the researcher is able to perform any statistical analysis; (iii) inaccuracies in the standardization, disambiguation and matching leading to measurement errors and biases.

[4]Aggregate figures at the macro, regional and sectoral level are currently released by the OECD in a systematic way.

[5]During the considered period, the USPTO was just reporting granted applications.

at the same time. After a first round based on exact matching and rough approximate matching, Hall et al (1988) manually checked all matches and recursively manually matched all possible unmatched firms and applicants. They repeated this procedure for each update of data of Compustat and USPTO from 1978 through 1988. This procedure, although quite effective, has been considered very costly (both in monetary and time terms) and difficult to extend to databases containing information on small and medium enterprises. The matching between USPTO applicants and companies in Compustat has been recently updated (Cockburn et al, 2009; Hall et al, 2001).

Preliminary or final datasets created by this project were employed in very influential articles published by researchers affiliated to the NBER through the 80s. Pakes and Griliches (1980) investigate the relationship between R&D expenditures and patent counts for 121 large corporation in 1968-1975. They find a very strong cross-sectional correlation and a significant, though weaker, correlation within firm. Moreover, they attempt to investigate the extent to which past R&D affects current patent counts, finding strong contemporaneous correlation and smaller (though still significant) positive correlation between current patent counts and past R&D expenditures. Hausman et al (1984) develop an econometric method aimed at investigating count data in a panel setting. They apply their econometric method in the investigation of the relationship between R&D expenditures and patent counts at the firm level[6], trying to determine the lag structure of R&D in influencing innovation outcome measured with patent counts. Hall et al (1986) extends the analysis of Hausman et al (1984) to a larger (642 firms) but shorter (1972-1979) panel. Finally, Griliches et al (1988) go beyond the simple R&D to patent relationship and investigate (i) the relationship between innovation and stock market value; (ii) the value of patents, taking advantage of data on patent renewal fees; (iii) the presence of knowledge spillovers.

A more recent attempt to integrate patent data with other firm-level database has been carried out by Grid Thoma and colleagues (Thoma

[6]The sample used by Hausman et al (1984) consists in 128 firms for the period 1968-1974.

and Torrisi, 2007; Thoma et al, 2010). The aim of their project was to create automatic routines and algorithms to harmonize, disambiguate and match lists of applicants' and firms' names. The project aimed at matching patent applications at the USPTO and the European Patent Office (EPO) available in the Worldwide Patent Database (PATSTAT) by companies included into the AMADEUS database (Bureau van Dijk)[7]. The approach used for this project, as opposed to the previous NBER project on patents, allows the possibility to have slightly significant shares of false matches and false negatives but it extends the matching to SMEs and it tries to use more efficient and automatized methods. Great effort has been put in the creation of routines for names harmonization in order to correct the most common typing mistakes and standardize as many name conventions as possible[8]. This effort in names harmonization allows to have great results from exact matching and it enables to perform effective approximate matching. In the stage of approximate matching, information on location of applicants and firms is used to reduce the risk of false matches and scores are computed both in terms of string similarity functions and in terms of token distance (see Thoma et al (2010) for further details). Due to the huge number of applicant / firm matches, no global manual check is performed[9]. The final result consisted in 131,065 companies included in AMADEUS identified as EPO applicants corresponding to about one million of EPO applications in 1979-2008.

Helmers et al (2011) focus on a single EU country, the U.K., and harmonize and match firms in FAME (*Financial Analysis Made Easy*, the British counterpart of AMADEUS, including the entire population of British registered firms for the 2000-2007 period) with patent applications at the EPO and the IPO (Intellectual Property Office, the British patent office) and with trademarks. Helmers et al (2011) just consider exact matching, with much of the effort put on names harmonization.

[7] The version of AMADEUS employed by Thoma et al (2010) includes about 10 millions of companies for the years 1998-2006.

[8] In order to increase the likelihood of identifying all matches, they also included name variations of single entities applying for PCT/WIPO as an additional dictionary.

[9] Thoma et al (2010) perform a manual check of approximate matches for a small sample of applicant / firm pairs (76 pairs), finding that the share of false matches is quite low (3 false matches, 3.9 percent)

The matching allowed to cover, as regards year 2003, about 83 percent of EPO applications (57 percent as regards IPO applications and 86 percent as regards trademarks) by British business entities. Helmers et al (2011) also report some descriptive analysis on the distribution of patent applications and trademarks by firm size, location, sector and technology. In section 3.2 I replicate part of these descriptive analysis for Italian firms in AIDA.

To conclude this brief review of the relevant literature, it is worth reporting some information on the project APE-INV (Academic Patenting in Europe) aimed at matching inventors reported in PATSTAT with academic researchers and professors (Lissoni et al, 2010). The project, started in June 2009 and still ongoing, is leaded by the KITeS (Centre for Knowledge, Internationalization and Technology Studies of the Università Commerciale Luigi Bocconi, Milan, Italy) and it is funded by the European Science Foundation (ESF). As opposed to firms / applicants matching, inventors / researchers matching suffers relatively more of disambiguation problems (high frequency of several name / surname pairs in both lists) and relatively less of names harmonization.

This chapter describes the methodology and the results of the matching of Italian patent applicants (and corresponding patent applications) and Italian firms included in the AIDA (Bureau van Dijk) database. The combination of patent data with other non-survey data is likely to attenuate the high risk of selection bias (Mairesse and Mohnen, 2010) of innovation surveys although it generally limits the variety of research questions that might be addressed relative to innovation surveys. It is worth noting, however, that the AIDA database does not contain the population of Italian firms and there is a bias 'by construction' due to the exclusion of inactive firms after four year of inactivity.

The remainder or the chapter is structured as follows. Section 2 discusses the main issues related to the matching of PATSTAT with other databases of firms (2.1) and it focuses on the description of the data sources (2.2) and of the methodology I used for the matching (2.3). Section 3 discusses the results of the matching (3.1) and it shows some stylized facts arising from some simple descriptive analysis (3.2). Section 4 concludes.

A.2 Data and methodology

A.2.1 Where is the missing link?

The issue of matching data from different sources is common in applied economic research and researchers are asked to put a lot of effort in time consuming tasks not directly related to their research. Moreover, it happens often that the matching is characterized by uncertainty due to the absence of a unique identifier in the different sources. This lack of consistency among sources leads severe measurement errors, missing values (generally non-random) and small samples which reduce the reliability of any estimate.

The collection of data on patent applications is beyond the scope of databases on firms which generally focus on balance sheet information and other demographic data (year of incorporation, legal status, location, sector of activity, etc.). Patent data are made available in specialized database such as the 'EPO Worldwide Patent Statistical Database' (also known as PATSTAT) delivered by the European Patent Office twice a year. However, these databases do not include any unique identifier for applicants[10] because their primary unit of analysis is the patent application and not the applicant. This problem affects analyses performed using the applicants as unit of reference even within patent databases due to:

- the possibility of name variations for each firm (due to actual changes in the denomination of a firm, change in name conventions or typing mistakes);
- duplication of the same name for different firms.

Patent offices lack of consistency when collecting information on patents. It often happens that names of applicants and inventors are col-

[10]Possible candidates as unique identifiers for firms are the registration number at the Chamber of Commerce or the fiscal code which uniquely identifies each firm.

lected using different name conventions, without any consideration on whether the applicant or inventor has been already reported as applicant or inventor in previous applications. Moreover, this missing link among applicants and inventors through time makes typing mistakes more common. Finally, the absence of a unique identifier for applicants and inventors limits the possibility to distinguish between duplication of names due to multiple applications by the same inventor / applicant from distinct inventors / applicants with the same name.

The consequence of these problems for analyses performed within patent databases is the introduction of biases in statistics (applications count, citations count) at the applicant/inventor level. Corporate applicants, which contribute to about 81 percent of Italian EPO applications between 1977 and 2009, are more prone to the problem of name variations relative to the problem of duplication than inventors, giving rise to an expected negative bias (underestimation of applications / citations count). On the contrary, the bias is expected to be positive for statistics at the inventor level due to the high frequency of coincidence of common names.

These problems are exacerbated when the names of applicants are harmonized and matched with external lists of firms, especially when these lists do not represent the entire population of firms as it is the case in this chapter. Harmonization possibly gives rise to the transformation of distinct names into identical harmonized names. When one of these false duplicates is not included in the list of names because the list itself represents just a fraction of the whole population there exists the possibility that the harmonized applicant name is matched to the wrong duplicate in AIDA. In order to minimize this risk, additional information such as location of both applicants and firms in AIDA could be used to identify false exact matches.

A.2.2 Data

In this study I used three different sources of data: the AIDA database, the Worldwide Patent Database (PATSTAT) and the results of the match-

Table 34: # of firms with non-missing balance sheet information

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Agric. Mining	1721	2300	3063	3225	3888	4103	4238	4423	26961
MH-tech manuf	10626	12075	14688	15179	17544	18426	19091	19915	127544
Low-tech manuf	23905	27268	33148	34179	39897	42223	44009	46271	290900
EGW, construction	7785	9780	14807	16054	23438	26366	29127	33276	160633
Wholesale, retail, hotel	27489	32670	43103	45110	56263	60862	64865	69969	400331
Transport and telecom	4673	5551	7437	7932	10224	11184	11811	12750	71562
Finance, real estate	2503	3571	4221	5504	11805	13431	14967	17584	73586
Computer	1449	1902	3100	3325	4411	4718	4905	5173	28983
R&D services	134	174	263	281	363	386	417	461	2479
Business activities	3345	4875	7954	8757	12036	13200	14460	15818	80445
Other services	2512	3533	5691	6171	8209	8778	9237	9852	53983
Total	86142	103699	137475	145717	188078	203677	217127	235492	1317407

Table 35: # of firms with non-missing balance sheet information (share of total)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Agric. Mining	2.0	2.2	2.2	2.2	2.1	2.0	2.0	1.9	2.0
MH-tech manuf	12.3	11.6	10.7	10.4	9.3	9.0	8.8	8.5	9.7
Low-tech manuf	27.8	26.3	24.1	23.5	21.2	20.7	20.3	19.6	22.1
EGW, construction	9.0	9.4	10.8	11.0	12.5	12.9	13.4	14.1	12.2
Wholesale, retail, hotel	31.9	31.5	31.4	31.0	29.9	29.9	29.9	29.7	30.4
Transport and telecom	5.4	5.4	5.4	5.4	5.4	5.5	5.4	5.4	5.4
Finance, real estate	2.9	3.4	3.1	3.8	6.3	6.6	6.9	7.5	5.6
Computer	1.7	1.8	2.3	2.3	2.3	2.3	2.3	2.2	2.2
R&D services	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Business activities	3.9	4.7	5.8	6.0	6.4	6.5	6.7	6.7	6.1
Other services	2.9	3.4	4.1	4.2	4.4	4.3	4.3	4.2	4.1
Total	100	100	100	100	100	100	100	100	100

ing by Thoma et al (2010).

AIDA

AIDA is a commercial database, maintained by Bureau van Dijk, about Italian firms. AIDA contains balance sheet, income statement and other information, such as location, sector, year of incorporation, ownership, participations in other firms, covering a 10 year time window. The version of AIDA I used is AIDA TOP. As opposed to the full version (AIDA SMALL + MEDIUM + TOP), it includes all firms with a reported turnover greater than 1.5 million euros while it includes just a small proportion (about 10 percent) of firms with a reported turnover below 1.5 million euros.

AIDA TOP is affected by three types of selection. First, there is a selection into the full AIDA database which contains about 1 million of firms as opposed to the more than 4 million of firms reported by Istat as active firms in Italy. This type of selection is not explicitly disclosed by Bureau van Dijk. Second, within AIDA, AIDA TOP severely underrepresents small firms. Finally, firms which are inactive for more than four years are generally fully removed from the database thus reducing substantially the coverage in the first years of the database and introducing an additional selection bias by considering only surviving firms (whose features are likely to differ from those of firms which exited the market).

A final consideration regards the group / subsidiary status of firms: in the current chapter I did not proceeded to consolidate patent applications and financial accounts of subsidiary firms into the corresponding business group.

In April 2011, date in which I extracted the data, AIDA TOP contained 272,475 companies. Tables 34 and table 35 show, respectively, the absolute and relative sectoral distribution[11] (by year) of firms (firm / year pairs) with non-missing balance sheet information. The overall coverage (share of firms with non-missing balance sheet information) is increasing in time moving from 86,142 (35.5 percent) firms in 2000 to 235,492 (88.8 percent) firms in 2007. This increase is the combination of the entry of new firms, the extension of the coverage of existing firms and the fact that formerly active firms which are now inactive are generally not reported. The dynamics at the sector level is quite smooth except for the *Finance, real estate* sector in which I observe a significant increase in coverage starting from 2004 (the share of the sector moved from 3.8 percent in 2003 to 6.3 percent in 2004, corresponding to an increase of about 50 percent in absolute terms). This jump is probably due to a change in the selection rules for AIDA TOP occurred in 2004. Figure 16 shows the geographical distribution (by province) of firm / year pairs with non-missing balance sheet information. Besides the province which include

[11]Macro-sectors are defined as follows (Nace Rev. 1.1 codes): Agriculture and Mining 01-14; Medium-High Technology Manufacturing 23-35; Low Technology Manufacturing 15-22 and 36-37; EGW (Electricity, Gas and Water supply), Construction 40-45; Wholesale, Retail, Hotels 50-55; Transport and Telecommunication 60-64; Finance, Real Estate 65-71; Computer 72; R&D services 73; Business activities 74; Other Services 75-95.

the most important urban areas (Milano, Roma, Torino and Napoli, accounting for about 27 percent of total firm / year pairs), high concentration of firms is found in other provinces in Lombardia (Monza e Brianza, Varese, Bergamo, Como, Lecco), Prato (especially firms in the textile sector), provinces in the northern-east part of Italy (Trieste, Padova, Treviso, Vicenza, Verona, Venezia) and in provinces of Emilia-Romagna around Bologna (Bologna itself, Rimini, Modena). Low density of firms is instead found in most central and southern provinces.

PATSTAT

The EPO Worldwide Patent Statistical Database (PATSTAT) is a commercial database prepared by the European Patent Office on behalf of the OECD Taskforce on Patent Statistics. It contains information on patent applications for more than 80 countries. Information reported by PATSTAT includes: (i) applicants' and inventors' names and addresses; (ii) title and abstract of patent applications; (iii) priority, patent families and PCT links; (iv) bibliographic information (citation links); (v) classification of patents by technology classes.

I retrieved information from the version released in April 2011 (PATSTAT is released twice each year, in April and October) on all patent applications at the European Patent Office filed from 1977 through 2009. For each application (`appln_id`) I retrieved information on application date and priority date[12] (`appln_date` and `prio_date`), applicants' name (`person_name` and `doc_std_name`) and address (`person_address`) and IPC classes (`ipc_class_symbol`)[13]. Total applications by Italian applicants (results not shown but available upon request) follow a quite smooth increasing dynamics until 2006 while there

[12]The priority date is the date in which an application for a specific invention is filed for. After this first application, the applicant is allowed to apply for the same invention to other patent offices within 12 months claiming protection for that invention since the priority date.

[13]I did not extract and use information on continuations and technical relation so that raw counts of patent applications in the following section will contain double counting of patented innovations of applications due to multiplicity of applicants for the same application or due to distinct applications (continuations, technical relations) for the same innovation.

is a small drop in 2007 and a huge drop (about half of patent applications relative to 2006) in 2008 due to the well known truncation problem (Hall et al, 2001). Truncation in patent data for the years close to the date of collection is due to delays in the publication of EPO applications. EPO applications are published within eighteen months since application or priority, leading to an underestimation for application counts in the last three years of coverage of patent databases.

Matching by Thoma et al (2010)

The results of the matching done by Thoma et al (2010) have been recently disclosed[14]. They matched 4,796 Italian firms[15] in AMADEUS as applicants to the EPO corresponding to about 24k EPO applications (reference period: 1978-2006) and they published the complete lists of harmonized names and locations of companies in AMADEUS and applicants in PATSTAT. This base of data was used in the current work for two purposes: (i) the inclusion of the matches identified by the study; (ii) the use of AMADEUS and PATSTAT harmonized names as additional name variations for firms in AIDA with the corresponding unique identifier in Bureau van Dijk products.

A.2.3 Methodology

This section focuses on the details of the methodology I used to match EPO applicants reported in PATSTAT to firms in AIDA. Much of the effort was aimed at improving exact matching, with recursive rounds of harmonization and improvement of the cleaning routines. In addition to that, I extended the coverage by including approximate matches which were manually checked.

The matching has been performed in basically nine steps:

1. preliminary check on small samples of the main problems of names harmonization for both the list of applicants in PATSTAT and firms in AIDA;

[14] http://www.researchoninnovation.org/epodata/

[15] Thoma et al (2010) assign patent applications filed by subsidiary firms to the group.

2. recursive harmonization of names and improvement of the routines at each step of the harmonization;

3. identification of duplicates in the list of firms in AIDA;

4. exact matching of non-duplicate harmonized names;

5. harmonization of addresses;

6. exact matching of duplicate harmonized names using harmonized addresses;

7. identification of candidate pairs for the approximate matching and creation of similarity measures;

8. manual check of approximate matches;

9. inclusion of EPO applications matched by Thoma et al (2010) and treatment of applications matched with multiple applicants.

The starting point of the harmonization procedure is the set of harmonization routines published in the new homepage of the NBER Patent Data Project[16]. These routines consist in a first set of general cleaning and standardization commands: elimination of punctuations, standardization of special characters, elimination of double spaces, transformation of lower cases into upper cases and unification of acronyms. A second set of commands focuses on the standardization of common name conventions. The standardization concerns both the juridical status of the firm (e.g. SOCIETÀ PER AZIONI, SOC PER AZIONI, SOC PER AZ are all standardized as SPA) and other common words in firms names that could be written in several ways or levels of abbreviation (e.g. INDUSTRIES vs IND, MANUFACTURING vs MANUF vs MFG, INTERNATIONAL vs INT). Finally, a new string variable, called *stem name*, is created by eliminating from firm names the juridical status and some common words and abbreviations (e.g. MFG, INT, IND). The stem name will be used by a Perl programme to identify candidate approximate matches.

[16]https://sites.google.com/site/patentdataproject/

These routines have been further adapted and improved to fit Italian names. I ran the same name harmonization routines both on the list of applicants in PATSTAT[17] and on the list of firms in AIDA[18]. After that I ran some routines to standardize addresses in AIDA and PATSTAT. Differently from AIDA, the address (street, street number, ZIP code, city, country, etc.) is written in a unique field[19]. I performed some basic cleaning on the addresses, especially focusing on common abbreviations (e.g. 'S.' instead of 'Santo') and English translation of the name of big cities (e.g. Rome-Roma, Milan-Milano, Venice-Venezia, etc.). Finally, I identified all unique firms whose name was repeated in the list of AIDA firms (I define these occurrences as duplicates)[20].

Once names and addresses harmonization has been performed, I identified all exact matches with coinciding address. The criterion I used to identify matching addresses was to consider the matching of either the municipality (matching of the municipality reported in AIDA with any substring including the municipality name in the address in PATSTAT) or the ZIP code (or its reduced version with 3 or 4 digits). Within these matches, I manually checked duplicates names which shared the same set of patent applications. This category of matches depend on the coincidence of both name and location for firms in AIDA and applicants in PATSTAT. When possible, I kept the matches for which the location matched better (e.g. full ZIP vs 3-digit zip, road and city vs city alone) while I removed all the remaining ambiguous matches. Exact matches with coinciding location accounted for 42,376 matched patent applications. Finally, I matched all those non-duplicates exact matches for which the location was different in PATSTAT and AIDA (3,510 patent applications).

[17] I used both the list of applicants in table `tls206_person` (field `person_name`) and the list in table `tls208_doc_std_nms` (field `doc_std_name`) of the PATSTAT database.

[18] I included, when available, the past denomination of the firm. Moreover, I added all firm names already matched by Thoma et al (2010).

[19] The ZIP code has been extracted by identifying, within the unique field of the address, all numbers (without spaces) composed by five digits.

[20] In some cases the problem of duplicates might be very severe, as in the case of firms whose name is FUTURA SRL (60 occurrences), SIRIO SRL (46 occurrences) and PEGASO SRL (45 occurrences).

I then moved to approximate matching. To identify possible matches, I used the Perl application published in the new website of the NBER Patent Data Project[21]. Once I identified candidate approximate matches (about 36,000 pairs as regards EPO applicants and about 60,000 pairs as regards EPO applicants), I created various indicators of string similarity.

A first measure is the simple Levensthein distance (Levensthtein, 1966)[22] between harmonized names in AIDA and PATSTAT for candidate matches. The Levenshtein distance computes the number of single operations (deletion of a character, inclusion of a character, substitution of a character and displacement of a character) needed to transform one string into another string[23]. Especially when comparing long strings (or long strings with short strings), a relative measure is more appropriate. For this reason I considered also the ratio between the Levinshtein distance and the maximum or minimum length (in terms of number of characters) of the two strings. Finally, I computed another measure to account for the possibility that some unnecessary substring was added in one of the two strings. This measure is given by the difference between the Levenshtein distance and the absolute value of the difference between the length of the two strings ($LEV(A, B) - |length(A) - length(B)|$)[24]. Finally, I identified all cases in which one of the two strings represented a substring of the other string[25].

After that, I ranked the candidate pairs according to various combinations of measures of string similarity and I proceeded to the manual identification of matches with a quite conservative approach[26]. I also ranked

[21] As an alternative, I used the RECLINK user-written Stata command Blasnik (2007). However, RECLINK performed worse than the Perl application both in terms of speed and in terms of effectiveness.

[22] The computation of the Levinshtein distance in Stata has been performed with the user-written module LEVENSHTEIN (Reif, 2010).

[23] To transform TABLE into CABLE there is need of just one operation (substitution of T with C) which corresponds to a Levinshtein distance of 1. To transform TABLE into CATTLE three operations are needed: substitution of T with C, substitution of B with T, inclusion of a T before the L.

[24] In this case, MARIO ROSSI SPA and MARIO ROSSI SPA DI M ROSSI will have a score of 0 while MARIO ROSSI SPA and MARIA ROSSI SPA will have a score of 1.

[25] MARIO ROSSI SPA is a substring of MARIO ROSSI SPA DI M ROSSI.

[26] The combination of measures of string similarity was changed in every case in which, scrolling the list down to lower level of similarity, no match was found for about 50 pairs.

candidate matches according to the similarity of their address (when available). Approximate matching, overall, allowed to match 1,704 patent applications (location coincided for 1,283 patent applications).

Finally, I added all the matches found by Thoma et al (2010) regarding Italian firms. From these matches, I removed those applications which were already matched previously to partly correct for the fact that Thoma et al (2010) assigned patent applications of subsidiary firms to their controlling company. The inclusion of matches identified by Thoma et al (2010) consisted 14,226 EPO applications.

The execution of the nine steps gave as result the matching of 8,892 EPO applicants and 49,369 EPO applications that could be split into four broad categories of matches[27]:

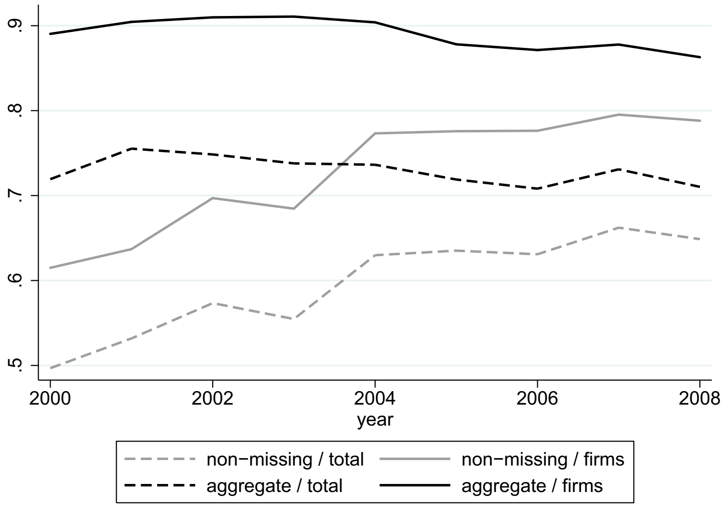
- Exact matching of firms for which the address of the applicants and the address in AIDA somehow coincides (42,376 EPO applications);
- Exact matching of non-duplicates firms with non-coinciding location (3,510 EPO applications);
- Approximate matching (1,704 EPO applications);
- BvD codes and publication numbers for EPO applications already matched by Thoma et al (2010) (14,226 EPO applications).

The first two categories are expected to be the most reliable types of matches, with an expected low share of false matches while the approximate matching is the category for which the share of false positive is expected to be higher. The possible false matches in the first categories could arise because of the failure to identify duplicates in the list of AIDA firms or because the applicant in PATSTAT corresponds to a firm which is not in the AIDA database. The version of AIDA employed in the current chapter is, in fact, a quite small non-random sample of the larger

The procedure stopped when all possible combination of measure were analysed.

[27] Note that the sum of applicants and applications, when they are split into the four categories, is greater than the aggregate figure because several matches pertain to more than one category.

Figure 15: Coverage (%) of the matching AIDA/PATSTAT



population of Italian firms, making possible to fail to identify all of the duplicates.

Each application/applicant pair was tagged according to the category of matching. In case of application/applicant pairs reported in multiple categories of matching, the pair was assigned to the category with the greater expected reliability[28]. Information on the source of matching could be useful in empirical analysis when choosing between the option of using the largest sample (with moderate probability of false matches) and the option of the sample with the lowest expected share of false matches (with a reduced sample size).

A.3 Results

A.3.1 Result of the matching PATSTAT/AIDA

The matching between firms in AIDA and EPO applicants for the period 2000-2007[29] resulted in 5,485 EPO applicants and 23,501 EPO ap-

[28]The rank is: (i) exact matching firms with the same location; (ii) exact matching of firms with different address; (iii) approximate matching; (iv) Thoma's matching.

[29]In the discussion of the results I do not consider applications in 2008 and 2009 because their total count is much lower than in the previous years. This depends on the well known truncation of patent data due to delays between the actual application and its publication

Table 36: Transition matrix for applications at EPO and UK Patent Office for firms in FAME taken from Helmers et al (2011) (all sectors - in %)

	No patent	1 patent	2-5 patents	6-10 patents	11-20 patents	20+ patents	Total
No patent	80.66	16.35	2.87	0.09	0.03	0	100
1 patent	71.24	19.92	8.23	0.44	0.13	0.03	100
2-5 patents	40.75	26.82	26.89	4.91	0.67	0.07	100
6-10 patents	7.6	15.2	37.22	26.61	11.66	1.7	100
11-20 patents	3.63	3.3	20.46	29.37	30.69	12.54	100
20+ patents	1.04	1.55	1.04	5.7	22.28	68.39	100
Total	75.99	17.48	5.39	0.68	0.29	0.17	100

Table 37: Transition matrix (EPO applications - all sectors - in %)

	No patent	1 patent	2-5 patents	6-10 patents	11-20 patents	20+ patents	Total
No patent	99.55	0.35	0.09	0	0	0	100
1 patent	74.57	16.35	8.55	0.47	0.06	0	100
2-5 patents	45.47	21.55	26.84	5.09	0.86	0.2	100
6-10 patents	7.39	9.13	39.57	27.39	15.22	1.3	100
11-20 patents	0.86	5.17	15.52	25.86	41.38	11.21	100
20+ patents	0	0	5.56	8.33	19.44	66.67	100
Total	99.3	0.47	0.19	0.02	0.01	0.01	100

plications. In addition, I matched 5,008 EPO applicants and 24,120 EPO applications for the period 1977-1999. Given that the version of AIDA I used covers the period 2000-2009 only, the matching for 1977-1999 is likely to miss many applicants who exited the market before 2000. However, data on patent applications for the period 1977-1999 could be useful when creating stock measures at the firm level.

As a rough measure of coverage of the population of Italian, I first computed the number of applications to the EPO for which the applicant name contained the strings 'SRL', 'SPA', 'SNC', 'SAS', 'SAPA' and 'COOP'[30] which identify the juridical status of any Italian commercial society. The general coverage of the matching (i.e. including all matches) is shown in figure 15.

On average, I was able to match more than 80 percent of EPO applications (on average 82.8 percent for the period 1977-2009, with a peak of 91 percent in 2002) filed by Italian firms.

Figure 15 reports trends of different measures of coverage, combining on the one hand either all matched firms in AIDA (`aggregate`) or only

(Hall et al, 2001). Matched applications for 2008 and 2009 are 2,616.

[30]These strings were searched from the list of harmonized names.

Table 38: Transition matrix (EPO applications - manufacturing - in %)

	No patent	1 patent	2-5 patents	6-10 patents	11-20 patents	20+ patents	Total
No patent	98.78	0.95	0.26	0.01	0	0	100
1 patent	74.06	16.22	9.17	0.55	0	0	100
2-5 patents	44.96	22.12	26.4	5.37	0.87	0.29	100
6-10 patents	8.43	4.82	42.17	26.51	16.27	1.81	100
11-20 patents	0	6.25	17.5	21.25	43.75	11.25	100
20+ patents	0	0	4.35	15.22	19.57	60.87	100
Total	98.11	1.25	0.53	0.07	0.03	0.02	100

Table 39: Transition matrix (EPO applications - medium-high tech manufacturing - in %)

	No patent	1 patent	2-5 patents	6-10 patents	11-20 patents	20+ patents	Total
No patent	97.76	1.71	0.5	0.02	0	0	100
1 patent	71.74	17.56	10.08	0.62	0	0	100
2-5 patents	41.92	22.12	27.77	6.64	1.33	0.22	100
6-10 patents	6.8	4.76	41.5	28.57	17.01	1.36	100
11-20 patents	0	7.25	13.04	24.64	44.93	10.14	100
20+ patents	0	0	3.03	15.15	21.21	60.61	100
Total	96.33	2.29	1.08	0.18	0.09	0.04	100

those firms with non-missing balance sheet information (`non-missing`) and on the other hand either total patent applications (`total`) without distinction regarding the type of applicants or applications by firms only (`firms`). According to the different measures, the coverage for 2000-2007 never falls below 50 percent and it is always around 90 percent when considering the weakest measure of coverage (applications by any AIDA matched firms over applications by firms in PATSTAT).

A.3.2 Some preliminary descriptive evidence on patenting patterns

In this section I discuss some descriptive evidence arising from the matched dataset, with a special focus on the distribution of applicants and applications across time, sector, firm size class, location and technology. All of the following results refer to the sub-sample of firm/year pairs for which was possible to retrieve at least the book value of the firm. Moreover, most of the statistics were computed on the sub-sample of manufacturing firms. The special focus on manufacturing is justified by the fact that innovations generally covered by IPRs consist, most of

Table 40: % of firms with at least one EPO application

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Agric. Mining	0.06	0.00	0.03	0.06	0.10	0.05	0.07	0.07	0.06
MH-tech manuf	3.60	3.69	3.53	3.49	3.32	3.44	3.35	3.32	3.44
Low-tech manuf	0.94	0.94	0.94	0.91	0.87	0.97	0.91	0.88	0.92
EGW, construction	0.17	0.16	0.14	0.13	0.14	0.14	0.11	0.15	0.14
Wholesale, retail, hotel	0.20	0.15	0.18	0.16	0.14	0.12	0.13	0.14	0.15
Transport and telecom	0.02	0.13	0.07	0.04	0.08	0.06	0.11	0.07	0.07
Finance, real estate	0.44	0.39	0.33	0.24	0.13	0.13	0.09	0.10	0.16
Computer	0.21	0.26	0.13	0.24	0.25	0.23	0.16	0.25	0.22
R&D services	8.21	6.90	5.70	6.05	5.51	5.96	7.19	4.99	6.09
Business activities	0.63	0.68	0.50	0.41	0.34	0.38	0.31	0.35	0.40
Other services	0.04	0.14	0.05	0.08	0.13	0.09	0.11	0.11	0.10
Total	0.84	0.81	0.73	0.70	0.61	0.62	0.59	0.57	0.66

Table 41: # of EPO applications

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Agric. Mining	1	0	2	2	5	2	3	3	18
MH-tech manuf	1138	1164	1434	1354	1525	1664	1649	1656	11584
Low-tech manuf	396	463	499	536	577	661	666	654	4452
EGW, construction	19	21	34	38	50	52	51	88	353
Wholesale, retail, hotel	76	108	114	117	106	105	106	135	867
Transport and telecom	1	45	71	85	103	95	74	52	526
Finance, real estate	35	45	57	51	58	35	41	40	362
Computer	3	6	6	10	15	17	9	14	80
R&D services	62	94	71	110	127	73	90	50	677
Business activities	39	54	58	76	84	113	94	106	624
Other services	2	11	6	10	15	13	18	21	96
Total	1772	2011	2352	2389	2665	2830	2801	2819	19639

the times, in product or process innovations which will be commercially exploited by the manufacturing sector. Moreover, most of these innovations are created within the manufacturing sector.

I report transition matrices for all sectors, for manufacturing firms and for manufacturing firms pertaining to medium-high technology sectors[31] (refer to tables 37, 38 and 39). Each row describes the distribution of the number of patent applications per firm at *t+1* for firms which were in the size class identified by the row at *t*. Transition matrices allow to in-

[31]Medium-high technology firms are, following the definition by OECD, those firms belonging to the following Nace Rev. 1.1 sectors: 23 (coke, refined petroleum products and nuclear fuel), 24 (chemicals and chemical products), 25 (rubber and plastic products), 26 (other non-metallic mineral products), 27 (basic metals), 28 (fabricated metal products, except machinery and equipment), 29 (machinery and equipment n.e.c.), 30 (office machinery and computers), 31 (electrical machinery and apparatus n.e.c.), 32 (radio, television and communication equipment and apparatus), 33 (medical, precision and optical instruments, watches and clocks), 34 (motor vehicles, trailers and semi-trailers) and 35 (other transport equipment).

Table 42: # of EPO applications (% of total)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Agric. Mining	0.06	0.00	0.09	0.08	0.19	0.07	0.11	0.11	0.09
MH-tech manuf	64.22	57.88	60.97	56.68	57.22	58.80	58.87	58.74	58.98
Low-tech manuf	22.35	23.02	21.22	22.44	21.65	23.36	23.78	23.20	22.67
EGW, construction	1.07	1.04	1.45	1.59	1.88	1.84	1.82	3.12	1.80
Wholesale, retail, hotel	4.29	5.37	4.85	4.90	3.98	3.71	3.78	4.79	4.41
Transport and telecom	0.06	2.24	3.02	3.56	3.86	3.36	2.64	1.84	2.68
Finance, real estate	1.98	2.24	2.42	2.13	2.18	1.24	1.46	1.42	1.84
Computer	0.17	0.30	0.26	0.42	0.56	0.60	0.32	0.50	0.41
R&D services	3.50	4.67	3.02	4.60	4.77	2.58	3.21	1.77	3.45
Business activities	2.20	2.69	2.47	3.18	3.15	3.99	3.36	3.76	3.18
Other services	0.11	0.55	0.26	0.42	0.56	0.46	0.64	0.74	0.49
Total	100	100	100	100	100	100	100	100	100

Table 43: % of firms with at least one EPO application (manufacturing - NACE Rev. 1.1)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
15	0.20	0.28	0.49	0.39	0.27	0.29	0.27	0.34	0.32
16	0.00	0.00	0.00	0.00	0.00	6.67	0.00	0.00	0.86
17	0.29	0.65	0.50	0.68	0.55	0.92	0.58	0.59	0.61
18	0.38	0.16	0.13	0.06	0.21	0.29	0.28	0.35	0.24
19	0.61	0.52	0.44	0.42	0.61	0.66	0.50	0.46	0.53
20	0.00	0.31	0.32	0.23	0.06	0.24	0.17	0.42	0.23
21	0.60	1.19	1.12	0.92	1.14	0.86	1.45	1.20	1.08
22	0.14	0.24	0.05	0.18	0.26	0.29	0.10	0.10	0.17
23	0.69	1.25	1.13	0.56	1.03	0.00	0.00	0.50	0.62
24	3.83	3.85	3.53	4.15	3.78	3.78	3.33	2.99	3.63
25	2.60	2.64	2.39	2.39	2.66	2.99	2.63	2.29	2.58
26	0.77	0.74	0.51	0.63	0.49	0.67	0.59	0.59	0.61
27	1.54	0.71	1.33	1.03	1.50	1.06	0.82	1.25	1.15
28	1.46	1.34	1.33	1.48	1.30	1.39	1.46	1.30	1.38
29	4.08	3.97	4.04	3.82	3.59	3.85	3.58	3.71	3.80
30	2.03	1.72	1.23	1.73	1.17	2.62	2.35	2.53	1.98
31	2.96	2.72	2.75	2.86	2.81	2.77	2.81	2.72	2.79
32	2.19	3.79	2.95	2.97	2.75	2.40	2.31	3.12	2.79
33	3.28	4.21	3.64	3.21	3.53	3.71	4.39	3.83	3.75
34	4.00	4.70	3.52	3.81	3.92	3.44	4.05	3.88	3.89
35	1.62	1.61	1.99	1.49	1.63	1.40	1.81	1.71	1.66
36	1.26	1.14	1.49	0.86	0.73	0.93	0.93	0.96	1.01
37	0.00	0.00	0.00	0.00	0.19	0.18	0.00	0.00	0.06
Total	1.44	1.42	1.32	1.29	1.15	1.20	1.13	1.07	1.22

vestigate in a compact way the degree of persistence of a variable. They have been used by two recent articles dealing with patent data (Helmers et al, 2011; Hingley and Bas, 2009) finding that persistence in patent application is increasing in the size of patenting activity in the past and that many firms file for one patent only during their active life.

First note that a very small proportion of firm/year pairs report positive patent applications (ranging from 0.7 percent for the aggregate figure of EPO application to 3.67 percent for EPO applications in medium-high technology manufacturing sectors) and that the proportion of patenting

Table 44: # of EPO applications (% of total - manufacturing - NACE Rev. 1.1)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
15	0.64	1.09	1.63	1.76	1.35	1.18	1.39	1.00	1.27
16	-	-	-	-	-	0.13	-	-	0.02
17	0.39	1.15	0.71	1.09	0.88	1.47	1.01	1.46	1.06
18	0.26	0.18	0.15	0.05	0.46	0.42	0.59	0.79	0.39
19	0.52	0.42	0.36	0.47	0.60	0.67	0.68	0.58	0.55
20	-	0.18	0.20	0.21	0.05	0.29	0.21	0.38	0.20
21	0.77	1.15	0.86	0.73	0.88	1.18	1.39	0.92	1.00
22	0.13	0.30	0.05	0.21	0.33	0.46	0.13	0.17	0.23
23	0.06	0.12	0.10	0.05	0.09	-	-	0.04	0.05
24	12.69	13.71	16.07	16.55	14.45	13.04	12.00	10.09	13.45
25	6.83	7.22	5.39	6.43	6.64	5.89	6.76	5.59	6.30
26	1.67	1.52	1.02	1.19	1.02	1.56	1.52	1.63	1.39
27	1.35	0.67	0.97	0.93	1.16	1.26	0.80	0.96	1.01
28	9.08	10.98	10.37	12.34	11.48	11.11	11.37	11.22	11.06
29	28.85	27.00	28.52	26.30	28.16	27.98	28.53	28.69	28.04
30	0.52	0.24	0.20	0.36	0.23	0.55	0.55	0.83	0.45
31	6.57	7.89	7.32	7.57	6.74	7.99	8.50	10.34	7.97
32	14.17	11.35	11.03	10.06	10.22	9.93	7.86	6.38	9.84
33	4.12	4.67	4.98	3.73	4.32	4.42	5.79	4.63	4.62
34	5.80	4.98	3.86	4.93	4.88	3.87	4.23	5.21	4.67
35	0.58	0.79	0.92	0.73	1.86	2.23	2.24	2.88	1.64
36	3.80	3.09	3.56	2.33	1.81	2.15	2.28	2.54	2.62
37	-	-	-	-	0.05	0.04	-	-	0.01
Total	100	100	100	100	100	100	100	100	100

Table 45: % of firms with at least one EPO application (manufacturing)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
North-West	2.18	2.08	2.01	1.88	2.02	2.28	2.16	2.10	2.10
North-East	2.03	2.22	2.01	2.09	2.25	2.59	2.44	2.48	2.28
Central Italy	1.06	1.20	1.15	1.38	1.17	1.43	1.16	1.25	1.23
South and islands	0.33	0.50	0.36	0.40	0.36	0.45	0.29	0.43	0.39
Total	1.76	1.82	1.70	1.74	1.76	2.04	1.86	1.88	1.83

firms increases when moving from the figure for all firms to manufacturing and then to medium-high technology manufacturing sectors. Second, all firms which file more than 20 EPO applications in a specific year will also patent at least 2 EPO application the following year (about 85 percent of these firms will apply for 11 or more EPO applications the following year). Finally, firms with no patent have a very small probability of filing for 6 or more patents next year (probability always below 0.02 percent). This last empirical regularity shows that becoming a great innovator is a cumulative long run process.

Comparing the results I obtained for aggregate Italian EPO applications with those obtained by Helmers et al (2011) (reported here in table 36) on EPO and IPO patent applications, I observe that the lower-right

Table 46: # of EPO applications (% of total - manufacturing)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
North-West	59.36	53.53	56.94	56.42	50.16	47.35	52.54	49.35	52.92
North-East	28.59	32.37	29.24	29.82	37.15	38.61	34.82	34.87	33.43
Central Italy	10.92	12.88	12.71	12.52	11.06	12.82	11.23	13.57	12.23
South and islands	1.13	1.22	1.11	1.25	1.63	1.23	1.40	2.21	1.42
Total	100	100	100	100	100	100	100	100	100

Table 47: % of firms with at least one EPO application (manufacturing)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Micro firms	0.51	0.39	0.45	0.33	0.40	0.46	0.41	0.42	0.42
Small firms	0.90	0.79	0.87	0.90	0.92	1.15	1.08	1.09	0.98
Medium firms	3.76	3.90	4.01	3.92	4.34	4.91	4.48	4.73	4.28
Large firms	16.51	16.93	16.47	16.49	16.50	17.28	17.01	17.10	16.81
Total	1.76	1.82	1.70	1.74	1.76	2.04	1.86	1.88	1.83

part of the transition matrix (firm / year pairs with 2 or more patent applications) is almost identical while the probabilities of transition for the categories with one or no patents differ substantially, with Italian firms with only one application per year being less persistent in their patenting activity[32]. This fact suggests that, once the hurdle of filing for patents is passed, Italian innovative firms behave similarly to other European innovative firms, as found by Lotti and Schivardi (2005).

Table 40 shows the propensity to patent of firms in different macro-sectors expressed as share of firm / year pairs with positive patent applications. R&D services and manufacturing (especially medium-high technology manufacturing sectors) tend to patent more than other sectors. On the one hand, patents are an output measure of firms in the R&D services sector. Innovations introduced by this sector are generally transferred and licensed to firms in the industrial (especially manufacturing) sector. Patents allow firms in the R&D service sector to appropriate of (at least part of) the commercial value of the innovations. The rest of the service sectors show a very low propensity to patent as well as the

[32]This systematic difference depends on two differences between my analysis and the one of Helmers et al (2011). First, Helmers et al (2011) includes the universe of UK firms while my sample under-represents small firms. Second, Helmers et al (2011) combine both patents filed at the national patent office and patents filed at the EPO while I focus on patents filed at the EPO only.

Table 48: # of EPO applications (% of total - manufacturing)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Micro firms	3.77	2.30	3.75	2.36	2.93	2.79	2.95	3.37	3.04
Small firms	14.16	12.73	15.23	16.12	19.71	20.21	18.35	20.05	17.34
Medium firms	21.71	25.90	25.37	24.38	28.50	29.71	26.78	27.14	26.33
Large firms	60.36	59.06	55.65	57.14	48.86	47.29	51.91	49.45	53.29
Total	100	100	100	100	100	100	100	100	100

Table 49: EPO applications by technology domain (OST7 classification and other classifications - % of total patents - manufacturing)

	2000	2001	2002	2003	2004	2005	2006	2007	Total
Electrical engineering; Electronics	21	16	19	18	11	11	15	15	15
Instruments	13	12	12	11	10	10	12	11	11
Chemicals; Materials	12	12	15	13	13	12	11	9	12
Pharmaceuticals; Biotechnology	7	8	8	7	6	7	5	5	7
Industrial processes	26	31	27	29	32	29	26	25	28
Mechanical eng.; Machines; Transport	30	28	26	29	30	29	30	29	29
Consumer goods; Civil engineering	15	15	19	16	21	21	19	22	19
Environmental patents (OECD)	2	3	2	3	4	3	2	3	3
Environmental patents (WIPO)	7	7	5	7	7	7	7	8	7
Environmental patents (OECD+WIPO)	7	8	6	8	7	8	7	9	8
ICT patents (OECD)	21	17	17	18	9	8	13	12	14

EGW, construction, agriculture and mining sectors.

Table 41 shows the sectoral distribution of patent applications while table 42 shows the relative contribution of each sector to the total number of patent applications. Manufacturing sectors alone account for about 81 percent of EPO applications while the R&D services sector, despite its very high propensity to patent, contribute little to the aggregate number of patents due to the reduced number of firms pertaining to this sector. As expected, within manufacturing medium-high technology sectors are characterized by greater propensity to patent and greater contribution to overall patents relative to low technology sectors.

A specific focus on the propensity to patent and on the distribution of patent applications within manufacturing is needed (see tables 43 and 44). The propensity to patent is steadily above 2 percent for 7 sectors, all of them pertaining to the medium-high technology category (Nace Rev. 1.1 codes 24, 25, 29, 31, 32, 33 and 34). On the contrary, most of the remaining manufacturing sectors show a very low propensity to patent, generally lower than 1 percent. When considering the contribution of each sector to total manufacturing patent applications, the distribution is even more skewed, with five sectors (Nace Rev. 1.1 codes 29, 24, 32, 28 and 31) accounting for about 70 percent of EPO applications by manufacturing firms.

Tables 45 and 46 and figures 17 and 18 show the geographical distribution (by macro-regions and by province) of patent applications and applicants[33]. Patent propensity of manufacturing firms is much greater in northern regions (about 2-2.6 percent) than in central (about 1-1.4 percent) and southern (0.3-0.5 percent) regions. The same pattern has been found for the contribution to aggregate manufacturing applications, with about 86 percent of EPO applications filed by firms located in northern regions and just 1-2 percent of EPO applications filed by firms located in southern regions. This clear geographical concentration of the patenting activity is probably a combination of different sectoral mix, different local systems of innovations and different endowment of physical and human capital across macro-regions.

The distribution of patenting activity by province (in this case results refer to all economic sectors) highlights the high concentration of patent applications in few areas: provinces in Lombardia and Veneto between Milano and Venezia, provinces in Emilia-Romagna between Bologna and Piacenza, Torino and Roma while no southern province has relevant patenting activities. Patent propensity follows a similar pattern, with the only notable difference of low patent propensity in Rome and high patent propensity in the Marche region and in the provinces of Chieti (Abruzzo) and Isernia (Molise).

Tables 47 and 48 report, respectively, patent propensity and relative distribution of patent applications filed by manufacturing firms by firm size[34]. The contribution of large firms to total patent applications is very

[33] Also in this case I focus on firm/year pairs for which balance sheet information was available.

[34] According to the European Commission (Recommendation 2003/361/EC), macro classes of firms by size are defined as follows: (i) micro firms are defined as firms with less than 10 employees and a turnover below 2 millions euros or a book value below 2 millions euros; (ii) small firms are defined as firms with 11 to 50 employees and a turnover between 10 and 2 millions euros or a book value between 10 and 2 millions euro; (iii) medium-sized firms are defined as firms with 51 to 250 employees and a turnover between 10 and 50 millions euros or a book value between 10 and 43 millions euros; (iv) large firms are defined

important (about 50-60 percent) and, more generally, patent propensity and contribution to aggregate applications is decreasing in firm size, as expected.

Finally, I briefly discuss the distribution of EPO applications by technology class (table 49). I classified EPO applications according to various technology classifications derived from IPC classes. First, I used the ISI-OST-INPI classification (8th edition 2006, Schmoch (2008)) which groups the several hundred thousands IPC classes into 30 or 7 macro-areas. The upper part of table 49 reports the classification in 7 macro-areas[35]. The distribution across technological fields is extremely persistent, with no significant shift among technology field in the considered period. Slightly more than a quarter of manufacturing EPO applications is in the field of mechanical engineering, machines, transport and/or in the field of industrial processes, which are the two most popular fields. Pharmaceutical and biotechnology patents are just a small proportion of manufacturing patents (about 8 percent) but they are highly concentrated in the narrow pharmaceutical sector. Secondly, I identified environmental patents according to two different selections of environmental IPC classes: the 'IPC Green Inventory'[36] created by the WIPO (World Intellectual Property Organization) and the UNFCCC (United Nations Framework Convention on Climate Change) including Environmentally Sound Technologies (ESTs) and the 'Series of patent search strategies for the identification of selected environment-related technologies'[37] developed by the OECD. The approach followed by the OECD is more restrictive than the approach followed by the WIPO (which, on the contrary, includes most of the environmental patents already identified by the OECD). The share of environmental patents is quite low (about 8 percent for aggregate environmental patents and 3 percent for environmental patents identified by the OECD) and it does not show any clear increasing or decreasing trend. Finally, I consider ICT (Information and

in a residual way as those firms which are not included into any of the previous classes.

[35] The percentages do not sum up to 100 because several applications contain multiple IPC classes pertaining to different ISI-OST-INPI categories.

[36] http://www.wipo.int/classifications/ipc/en/est/

[37] http://www.oecd.org/dataoecd/4/14/47917636.pdf

Communication Technology) patents as defined by the OECD[38] according to their IPC class. ICT patents represent a significant though decreasing share of total EPO applications by manufacturing, shrinking from about 21 percent in 2000 to about 12 percent in 2007.

A.4 Conclusions

This chapter describes the creation of an integrated base of data on firm financial accounts and on firm patent applications for Italy. The administrative nature of financial accounts and the relevance and flexibility of patents as a measure of innovation output will allow to answer to a variety of research questions on the innovation patterns at the firm level, bearing in mind potential selection biases and possible measurement errors related to false positive and negative matches and to errors in financial accounts.

This integrated base of data can be easily extended with the rich information contained in patent data such as citation links, patent families, PCT applications and information on inventors.

[38] http://www.oecd.org/dataoecd/34/34/40807441.pdf

Figure 16: # of firm / year pairs by province (period 2000-2007 - only firms with balance sheet information)

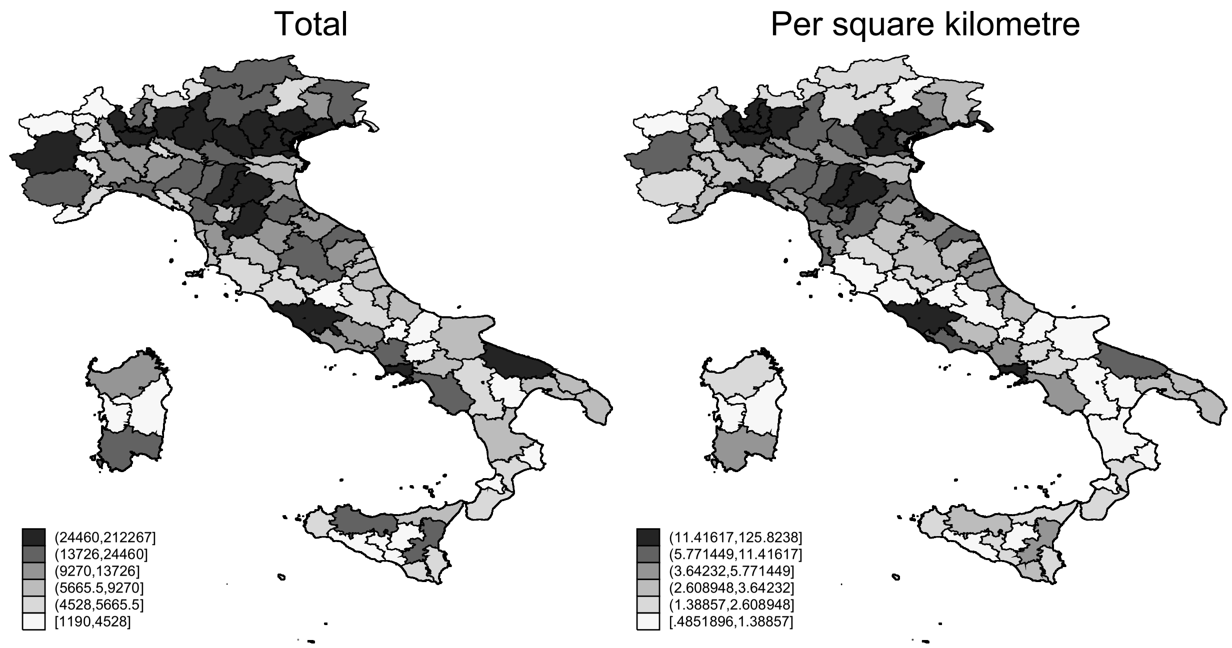


Figure 17: # of patent applications by province (period 2000-2007 - only firms with balance sheet information)

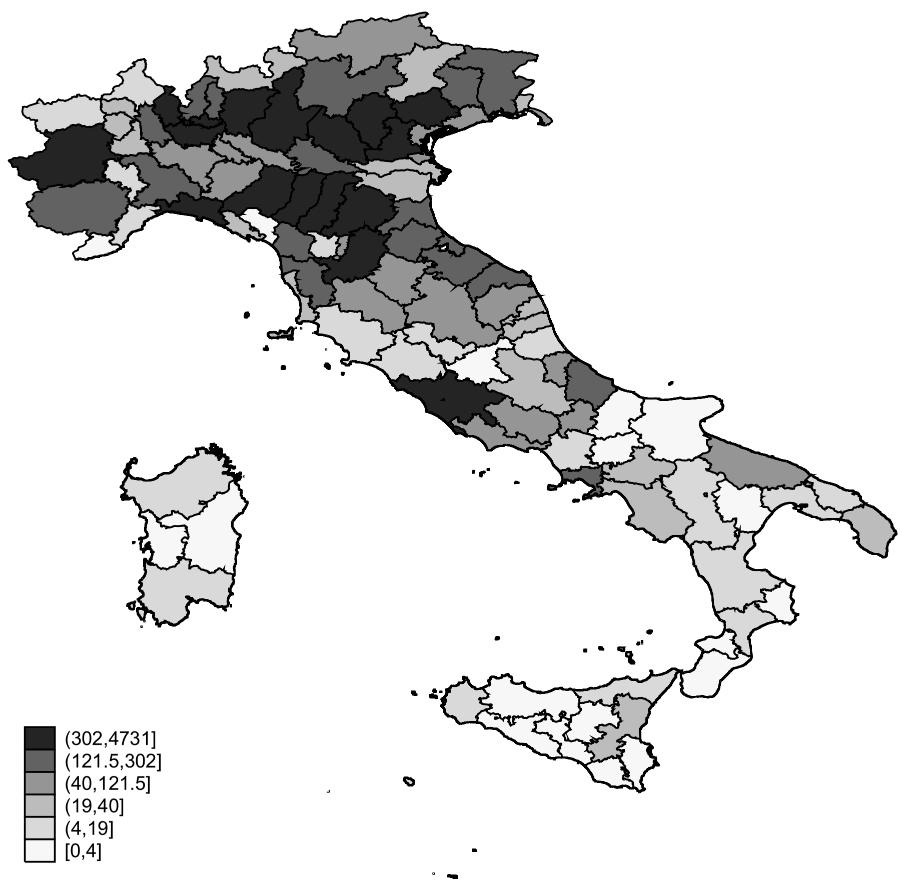
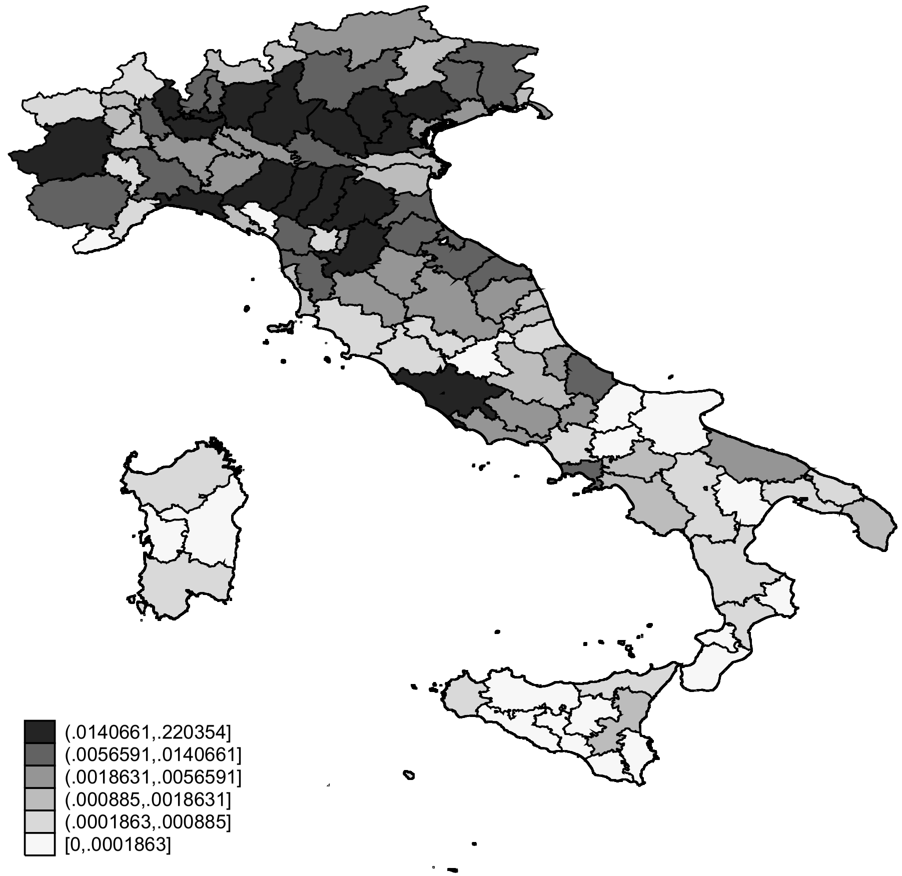


Figure 18: Share of firms with at least one patent application by province (period 2000-2007 - only firms with balance sheet information)



References

Aghion P, Howitt P (1992) A model of growth through creative destruction. Econometrica 60(2):323–51 91

Ahmad N, Wyckoff A (2003) Carbon dioxide emissions embodied in international trade of goods. OECD Science, Technology and Industry Working Papers 15, OECD 48, 51, 52, 60, 65, 66, 67, 68

Alcántara V, Padilla E (2009) Input-output subsystems and pollution: An application to the service sector and CO2 emissions in Spain. Ecological Economics 68(3):905–914 38

Andrew R, Peters G, Lennox J (2009) Approximation and regional aggregation in Multi-Regional Input-Output Analysis for national carbon footprint accounting. Economic Systems Research 21(3):311–335 43

Arto I, Roca J, Serrano M (2003) Accounting for embodied emissions in international trade: physical vs monetary trade flows, mimeo 38, 43, 59, 65, 68

Blasnik M (2007) RECLINK: Stata module to probabilistically match records. Statistical Software Components, Boston College Department of Economics 117

Bleischwitz R, Welfens P, Zhong Xiang Z (2009) Sustainable Growth and Resource Productivity: Economic and Global Policy Issues. Greenleaf Publishing, Greenleaf Publishing 34

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87(1):115–143 28

Boo AD, Bosch P, Gorter CN, Keuning SJ (2003) An environmental module and the complete system of national accounts. Occasional Working Paper NA-046, CBS 54

Bound J, Cummins C, Griliches Z, Hall BH, Jaffe AB (1984) Who does r&d and who patents? In: R&D, Patents, and Productivity, NBER Chapters, National Bureau of Economic Research, Inc, pp 21–54 105

Brunnermeier SB, Cohen MA (2003) Determinants of environmental innovation in US manufacturing industries. Journal of Environmental Economics and Management 45(2):278–293 77

Cainelli G, Mazzanti M, Zoboli R (2011) Environmental innovations, complementarity and local/global cooperation: Evidence from north-east Italian industry. International Journal of Technology, Policy and Management 11(3/4):328–368 73, 95

Cameron AC, Trivedi PK (1998) Regression Analysis of Count Data. No. 9780521632010 in Cambridge Books, Cambridge University Press 79, 92

Carrión-Flores CE, Innes R (2010) Environmental innovation and environmental performance. Journal of Environmental Economics and Management 59(1):27–42 9

Cockburn IMA, Agrawal A, Bessen J, Graham JHS, Hall BH, MacGarvie M (2009) The NBER patent citations datafile updated. Tech. rep., URL `https://sites.google.com/site/patentdataproject/Home` 106

Cohen WM, Klepper S (1996) A reprise of size and R&D. Economic Journal 106(437):925–51 91

Cole MA (2004) Trade, the pollution haven hypothesis and the environmental Kuznets curve: examining the linkages. Ecological Economics 48(1):71–81 34

Cole MA, Elliott RJ, Shimamoto K (2005) Industrial characteristics, environmental regulations and air pollution: an analysis of the UK manufacturing sector. Journal of Environmental Economics and Management 50(1):121–143 12

Comanor WS, Scherer FM (1969) Patent statistics as a measure of technical change. Journal of Political Economy 77(3):392–98 104

Copeland BR, Taylor MS (2004) Trade, growth, and the environment. Journal of Economic Literature 42(1):7–71 34

Corradini M, Costantini V, Mancinelli S, Mazzanti M (2011) Environmental and innovation performance in a dynamic impure public good framework. Departmental Working Papers of Economics - University 'Roma Tre' No. 0141, Department of Economics - University Roma Tre 8, 31

Costantini V, Mazzanti M (2010) On the green side of trade competitiveness? environmental policies and innovation in the EU. Working Papers 2010.94, Fondazione Eni Enrico Mattei 9

Costantini V, Mazzanti M, Montini A (2011) Hybrid Economic Environmental Accounts. Routledge, London 40, 54

Crepon B, Duguet E, Mairesse J (1998) Research, innovation, and productivity: An econometric analysis at the firm level. NBER Working Papers 6696, National Bureau of Economic Research, Inc xxii, 2, 75, 77, 78, 79, 104

Eberhardt M, Helmers C (2010) Untested assumptions and data slicing: A critical review of firm-level production function estimators. Economics Series Working Papers 513, University of Oxford, Department of Economics 94

ETC/SCP (2011a) Key messages on material resource use and efficiency in Europe, eTC/SCP Topic Centre on Sustainable Consumption and Production, Copenhagen 34

ETC/SCP (2011b) Progress in sustainable consumption and production in Europe, ETC/SCP European Topic Centre on Sustainable Consumption and Production, Copenhagen 37

European Commission (2009) Eu industrial structure 2009, performance and competitiveness, Brussels 35

Eurostat (2001) Methods for estimating air emissions from the production of goods imported into the uk, Prepared for DG Regional Policy and Eurostat, Working Paper 2/2001/B/5. European Commission, Eurostat, Brussels 34

Eurostat (2008) Eurostat Manual of Supply, Use and Input-Output Tables. Eurostat Methodologies and Working Papers, Luxembourg 53

Eurostat (2011) Creating consolidated and aggregated EU27 supply, use and input-output tables, adding environmental extensions (air emissions), and conducting Leontief-type modelling to approximate carbon and other 'footprints' of EU27 consumption for 2000 to 2006, Eurostat, Luxembourg 35, 40

Frankel JA, Rose AK (2005) Is trade good or bad for the environment? sorting out the causality. The Review of Economics and Statistics 87(1):85–91 34

Gallego B, Lenzen M (2005) A consistent input-output formulation of shared producer and consumer responsibility. Economic Systems Research 17(4):365–391 44

Griffith R, Huergo E, Mairesse J, Peters B (2006) Innovation and productivity across four European countries. Oxford Review of Economic Policy 22(4):483–498 79, 81, 104

Griliches Z (1990) Patent statistics as economic indicators: A survey. Journal of Economic Literature 28(4):1661–1707 76, 104

Griliches Z, Pakes A, Hall BH (1988) The value of patents as indicators of inventive activity. NBER working papers, National Bureau of Economic Research, Inc 106

Hall B, Lotti F, Mairesse J (2009) Innovation and productivity in SMEs: empirical evidence for Italy. Small Business Economics 33(1):13–33 79, 81

Hall BH (2006) R&D, productivity and market value. IFS Working Papers W06/23, Institute for Fiscal Studies 89

Hall BH, Griliches Z, Hausman JA (1986) Patents and R&D: Is there a lag? International Economic Review 27(2):265–83 106

Hall BH, Cumminq C, Laderman ES, Mundy J (1988) The R&D master file documentation. NBER Technical Working Papers 0072, National Bureau of Economic Research, Inc 105, 106

Hall BH, Jaffe AB, Trajtenberg M (2001) The NBER Patent Citation Data File: Lessons, insights and methodological tools. NBER working papers, National Bureau of Economic Research, Inc 106, 114, 120

Hall BH, Thoma G, Torrisi S (2007) The market value of patents and R&D: Evidence from european firms. NBER Working Papers 13426, National Bureau of Economic Research, Inc 77

Hatanaka M (1952) Note on consolidation within a Leontief system. Econometrica (20):301–303 49

Hausman J, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52(4):909–38 106

Helmers C, Rogers M, Schautschick P (2011) Intellectual property at the firm-level in the UK: The Oxford firm-level intellectual property database. Economics Series Working Papers 546, University of Oxford, Department of Economics xiii, 107, 108, 120, 123, 124, 125

Hingley P, Bas S (2009) Numbers and sizes of applicants at the European Patent Office. World Patent Information 31(4):285–298 123

Hoekstra R (2010) A complete overview of environmental input-output articles, Paper presented at the I-O Society Conference, Sidney, 20-25 June 2010. 38

Horbach J (2008) Determinants of environmental innovation–new evidence from German panel data sources. Research Policy 37(1):163–173 73

Horbach J, Rammer C, Rennings K (2011) Determinants of eco-innovations by type of environmental impact. The role of regulatory push/pull, technology push and market pull. ZEW Discussion Papers 11-027, ZEW - Zentrum fr Europaische Wirtschaftsforschung / Center for European Economic Research 73

Huppes G, Koning AD, Van Nielsen SSRHLOP, Guinee J (2005) Environmental impacts of consumption in the EU: high resolution input output tables with detailed environmental extensions. Journal of Industrial Ecology (10):129–146 39, 48, 65

IPTS (2006) Environmental impacts of products (EIPRO), IPTS/ESTO Project IPTS, European Commission Technical Report Series, IPTS, Seville 39

Jaffe A, Newell R, Stavins R (2002) Environmental policy and technological change. Environmental & Resource Economics 22(1):41–70 72

Jaffe AB, Palmer K (1997) Environmental regulation and innovation: A panel data study. The Review of Economics and Statistics 79(4):610–619 74

Jaffe AB, Newell RG, Stavins RN (2005) A tale of two market failures: Technology and environmental policy. Ecological Economics 54(2-3):164–174 7

Johnstone N, Hascic I, Popp D (2010) Renewable energy policies and technological innovation: Evidence based on patent counts. Environmental & Resource Economics 45(1):133–155 77, 88

Kemp R, Pearson P (2007) Final report mei project about measuring eco-innovation. URL `http://www.merit.unu.edu/MEI` 75, 76

Lanjouw JO, Mody A (1996) Innovation and the international diffusion of environmentally responsive technology. Research Policy 25(4):549–571 7, 77, 88

Lenzen M (2011) Aggregation versus disaggregation in input-output analysis of the environment. Economic Systems Research 23(1):73–89 42, 50

Levensthtein VI (1966) Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady 10(8):707–710 117

Levinson A (2010) Offshoring pollution: Is the United States increasingly importing polluting goods? Review of Environmental Economics and Policy 4(1):63–83 36

Lissoni F, Maurino A, Pezzoni M, Tarasconi G (2010) APE-INV's "name game" algorithm challenge: A guideline for benchmark data analysis and reporting. Tech. rep., Academic Patenting in Europe - APE-INV, URL `http://www.esf-ape-inv.eu` 108

Lotti F, Schivardi F (2005) Cross country differences in patent propensity: A firm-level investigation. Giornale degli Economisti 64(4):469–502 125

Lovely M, Popp D (2011) Trade, technology, and the environment: Does access to technology promote environmental regulation? Journal of Environmental Economics and Management 61(1):16–35 8

Mairesse J, Mohnen P (2010) Using innovations surveys for econometric analysis. NBER Working Papers 15857, National Bureau of Economic Research, Inc 105, 108

Mansfield E (1986) Patents and innovation: An empirical study. Management Science 32(2):173–181 80

Marin G, Mazzanti M (in press) The evolution of environmental and labor productivity dynamics - sector analyses and decoupling/recoupling trends on a 1990-2007 Italian NAMEA. Journal of Evolutionary Economics 10, 12, 34, 38, 59

Mazzanti M, Montini A (2010) Embedding the drivers of emission efficiency at regional level - analyses of NAMEA data. Ecological Economics 69(12):2457–2467 38

Mazzanti M, Musolesi A (2010) Carbon abatement leaders and laggards, Nota di Lavoro, 149, FEEM, Milan 10, 41

Mazzanti M, Zoboli R (2009) Environmental efficiency and labour productivity: Trade-off or joint dynamics? a theoretical investigation and empirical evidence from italy using namea. Ecological Economics 68(4):1182–1194 12, 34, 40

Miller RE, Blair PD (2009) Input-Output Analysis - Foundations and Extensions, 2nd edn. Cambridge University Press, Cambridge, Mass. 49, 53

Moll S, Vrgoc MM, Watson D, Femia A, Pedersen OG, Villanueva A (2007) Environmental input-output analyses based on NAMEA data - a comparative European study on environmental pressures arising from consumption and production patterns, ETC/RWM Working Paper 2/2007. European Topic Centre - Resource and Waste Management, Copenhagen 39, 47, 57, 65, 66, 67, 68

Moomaw WR, Unruh GC (1997) Are environmental Kuznets curves misleading us? the case of CO2 emissions. Environment and Development Economics 2(04):451–463 10

Morimoto Y (1970) On aggregation problems in input-output analysis. Review of Economic Studies 37(1):119–26 49

Nakano S, Okamura A, Sukurai N, Suzuki M, Tojo Y, Yamano N (2009) The measurement of CO2 embodiments in international trade: Evidence from the harmonised input-output and bilateral trade database. Statistical Analysis of Science, Technology and Industry Working Paper 3, OECD 48, 51, 60, 65, 66, 67, 68

Newell RG, Jaffe AB, Stavins RN (1999) The induced innovation hypothesis and energy-saving technological change. The Quarterly Journal of Economics 114(3):941–975 9

Nicoletti G, Scarpetta S (2003) Regulation, productivity and growth: OECD evidence. Economic Policy 18(36):9–72 11, 13

OECD (2009) OECD Patent Statistics Manual. OECD 104

OECD (2011a) Patents by regions. http://dx.doi.org/10.1787/data-00509-en 89

OECD (2011b) Resource productivity in the G8 and OECD, OECD, Paris 34

Oltra V, Kemp R, de Vries FP (2009) Patents as a measure for eco-innovation. Cahiers du GREThA 2009-05, Groupe de Recherche en Economie Thorique et Applique 76

Pakes A, Griliches Z (1980) Patents and R&D at the firm level: A first look. NBER Working Papers 0561, National Bureau of Economic Research, Inc 106

Pavitt K (1984) Sectoral patterns of technical change: Towards a taxonomy and a theory. Research Policy 13(6):343–373 78, 80, 104

Pavitt K (1985) Patent statistics as indicators of innovative activities: Possibilities and problems. Scientometrics 7(1-2):77–99 104

Peters GP (2008) From production-based to consumption-based national emission inventories. Ecological Economics 65(1):13–23 43

Popp D (2002) Induced innovation and energy prices. American Economic Review 92(1):160–180 9, 77

Popp D (2006) International innovation and diffusion of air pollution control technologies: the effects of NOx and SO2 regulation in the US, Japan, and Germany. Journal of Environmental Economics and Management 51(1):46–71 7, 8

Popp D (2010) Innovation and climate policy. NBER Working Papers 15673, National Bureau of Economic Research, Inc 72

Popp D, Newell RG (2009) Where does energy R&D come from? Examining crowding out from environmentally-friendly R&D. NBER Working Papers 15423, National Bureau of Economic Research, Inc 74, 81

Popp D, Newell RG, Jaffe AB (2009) Energy, the environment, and technological change. NBER Working Papers 14832, National Bureau of Economic Research, Inc 6, 72

Porter ME, van der Linde C (1995) Toward a new conception of the environment-competitiveness relationship. Journal of Economic Perspectives 9(4):97–118 9, 74, 101

Rave T, Goetzke F, Larch M (2011) The determinants of environmental innovations and patenting: Germany reconsidered. Ifo Working Paper Series Ifo Working Paper No. 97, Ifo Institute for Economic Research at the University of Munich 73, 88

Reif J (2010) STRGROUP: Stata module to match strings based on their Levenshtein edit distance. Statistical Software Components, Boston College Department of Economics 117

Rennings K (2000) Redefining innovation – eco-innovation research and the contribution from ecological economics. Ecological Economics 32(2):319–332 75

Rennings K, Ziegler A (2004) Determinants of environmental innovations in Germany: Do organizational measures matter? a discrete choice analysis at the firm level. ZEW Discussion Papers 04-30, ZEW - Zentrum fr Europaische Wirtschaftsforschung / Center for European Economic Research 72

Rexhauser S, Rammer C (2011) Unmasking the porter hypothesis: Environmental innovations and firm-profitability. ZEW Discussion Papers 11-036, ZEW - Zentrum fr Europaische Wirtschaftsforschung / Center for European Economic Research 74

Roca J, Serrano M (2007) Income growth and atmospheric pollution in Spain: An input-output approach. Ecological Economics 63(1):230–242 53, 60

Scarpetta S, Tressel T (2002) Productivity and convergence in a panel of OECD industries: Do regulations and institutions matter? OECD Economics Department Working Papers 342, OECD Publishing 11, 13, 15

Schmoch U (2008) Concept of a technology classification for country comparisons. Final report to the World Intellectual Property Organisation, WIPO 128

Serrano M, Dietzenbacher E (2010) Responsibility and trade emission balances: An evaluation of approaches. Ecological Economics 69(11):2224–2232 35, 43, 44, 45, 47, 57, 59, 65, 68

Su B, Huang H, Ang B, Zhou P (2010) Input-output analysis of co2 emissions embodied in trade: The effects of sector aggregation. Energy Economics 32(1):166–175 42, 43, 50, 51, 55, 63, 65, 67

Thoma G, Torrisi S (2007) Creating powerful indicators for innovation studies with approximate matching algorithms. a test based on PATSTAT and

Amadeus databases. KITeS Working Papers 211, KITeS, Centre for Knowledge, Internationalization and Technology Studies, Università Bocconi, Milano, Italy 106

Thoma G, Torrisi S, Gambardella A, Guellec D, Hall BH, Harhoff D (2010) Harmonizing and combining large datasets - An application to firm-level patent and accounting data. NBER Working Papers 15851, National Bureau of Economic Research, Inc 107, 111, 114, 115, 116, 118

Turner K, Lenzen M, Wiedmann T, Barrett J (2007) Examining the global environmental impact of regional consumption activities - part 1: A technical note on combining input-output and ecological footprint analysis. Ecological Economics 62(1):37–44 43

Wagner M (2007) On the relationship between environmental management, environmental innovation and patenting: Evidence from German manufacturing firms. Research Policy 36(10):1587–1602 73, 77

Wyckoff AW, Roop JM (1994) The embodiment of carbon in imports of manufactured products : Implications for international agreements on greenhouse gas emissions. Energy Policy 22(3):187–194 51