OECD Science, Technology and Industry Working Papers
2017/04

The Multiprod project: A
comprehensive overview

**Giuseppe Berlingieri,
Patrick Blanchenay,
Sara Calligaris,
Chiara Criscuolo**

<https://dx.doi.org/10.1787/2069b6a3-en>

OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS

The release of this working paper has been authorised by Andrew Wyckoff, OECD Director for Science, Technology and Innovation, and by Catherine Mann, OECD Director for Economics Department.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors. Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to Directorate for Science, Technology and Innovation, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Note to Delegations:

This document is also available on OLIS under the reference code:
DSTI/CIIE(2017)4/FINAL

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

© OECD 2017

Applications for permission to reproduce or translate all or part of this material should be made to: OECD/OCDE Publications, 2 rue André-Pascal, 75775 Paris, Cedex 16, France; e-mail: rights@oecd.org.

THE MULTIPROD PROJECT: A COMPREHENSIVE OVERVIEW

Giuseppe Berlingieri, Patrick Blanchenay,
Sara Calligaris, Chiara Criscuolo (OECD)

Abstract

This paper is a technical document, designed to serve as a reference document for subsequent papers arising out of MultiProd, a project of the Committee on Industry, Innovation and Entrepreneurship and the Working Party on Industry Analysis, aimed at studying productivity patterns across countries and over time. MultiProd provides harmonised micro-aggregated data of paramount importance for investigating the extent to which different policy frameworks can shape firm productivity and examining the way resources are allocated to more productive firms.

The paper discusses the project's main contributions in relation to the current literature, in particular how using micro-aggregated data can help policy-makers understand and measure the efficiency of resource allocation in the economy, the dependence of economic activity on a small number of large firms, and wage inequality.

TABLE OF CONTENTS

OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS.....	2
TABLE OF CONTENTS	4
THE MULTIPROD PROJECT: A COMPREHENSIVE OVERVIEW	6
1. Introduction.....	6
2. The MultiProd dataset.....	10
2.1. The distributed microdata approach	10
2.2. Structure of the program and basic requirements.....	10
2.3. Measures of productivity and capital	13
2.4. Challenges	16
2.5. Country coverage	20
2.6. Output.....	23
3. Productivity heterogeneity	26
3.1. Motivation	26
3.2. Productivity Dispersion and its Evolution over Time	26
3.3. Productivity Dispersion Decomposition	28
3.4. Analysing MFP Time-Volatility	33
4. Allocative efficiency.....	34
4.1. Motivation	34
4.2. Measuring Reallocation.....	36
4.3. Measuring Misallocation.....	40
4.4. Policies, Allocative Efficiency and Productivity.....	48
4.5. The Great Recession, Reallocation and Productivity Growth.....	50
5. Granularity	54
5.1. Motivation	54
5.2. Understanding characteristics of the largest firms	54
5.3. The Contribution of the Largest Firms to Productivity	56
6. Wage Dispersion and its link to productivity.....	59
6.1. Motivation	59
6.2. A cross-country description and decompositions of wage dispersion.....	59
6.3. Policy implications regarding wage and productivity dispersion.....	64
6.4. Potential additional contributions of the project	66
7. Conclusions.....	68
APPENDIX	69
Metadata.....	69
Tables on sectoral, size and age aggregation	72
Description of the MultiProd output	73
Details of the MultiProd code	76
REFERENCES	83

Tables

Table 1.	Productivity measures used in MultiProd.....	15
Table 2.	Productivity measures used in MultiProd.....	21

Table 3.	Data coverage for manufacturing and non-financial market services sectors	21
Table 4.	Employment and capital data, recording units	22
Table 5.	Representativeness in 2011	23
Table 6.	Productivity 90-10 ratio in 2011, by country.....	26
Table 7.	Share of within-sector of variance of LP, 2011	29
Table 8.	Top three sectors in the share of within-sector LP VA variance.....	30
Table 9.	Dispersion of TFPR, manufacturing in DNK.....	43
Table 10.	Dispersion of TFPR, s in DNK.....	43
Table 11.	Dispersion of TFPQ, manufacturing in DNK.....	44
Table 12.	Dispersion of TFPQ, services in DNK	44
Table 13.	Herfindahl-Hirschman index by country in 2011, manufacturing and services sector.....	56
Table 14.	Share of within-sector of variance of wages, 2011.....	61
Table 15.	Top three sectors in the share of within-sector wage variance, 2011	62
Table A.1.	National data sources	69
Table A.2.	Sectoral classifications and coverage.....	70
Table A.3.	Optional variables	70
Table A.4.	Labour input variable	71
Table A.5.	Sectoral Aggregation	72
Table A.6.	Size Classes.....	73
Table A.7.	Age Classes.....	73
Table A.8.	Firm-specific variables used to define a cell.....	74
Table A.9.	List of statistics computed for a given quantity	74
Table A.10.	List of Basic Quantities.....	75
Table A.11.	Time transformations	76

Figures

Figure 1.	Trends for top, median and bottom decile of (log) LP	27
Figure 2.	Trends for top, median and bottom decile of (log) MFP	28
Figure 3.	OP covariance and unweighted productivity in 2011, manufacturing sector	37
Figure 4.	OP covariance trend in BEL, DNK, FRA and SWE, manufacturing sector.....	37
Figure 5.	Distribution of TFPR, manufacturing in DNK	43
Figure 6.	Distribution of TFPR, services in DNK	43
Figure 7.	Distribution of TFPQ, manuf. in DNK.....	44
Figure 8.	Distribution of TFPQ, services in DNK	44
Figure 9.	TFP gains.....	46
Figure 10.	Relative TFP gains, base year 2003.....	46
Figure 11.	Divergence of (log) productivity over time	52
Figure 12.	Value added and employment shares by quantiles of sales, manufacturing vs. services ..	55
Figure 13.	Top, median and bottom decile of real wages over time	60
Figure 14.	Change in real wages in different parts of the productivity distribution of firms over time	65

THE MULTIPROD PROJECT: A COMPREHENSIVE OVERVIEW

1. Introduction

One of the main objectives of economic research is to understand why some nations are more developed than others. A simple measure of economic development, output per capita, illustrates the large disparity found across countries. And these disparities are largely reflected in different levels of productivity across countries. Hall and Jones (1999), for example, find that output per worker in the United States (US), a measure of labour productivity, is 35 times greater than in Niger. Labour productivity differences can be partially explained by differences in physical and human capital (Caselli, 2005). However, the main reason for the disparity between these two countries comes from differences in aggregate Multifactor Productivity (MFP), which reflects the overall efficiency with which inputs are combined in the production process. More generally, Prescott (1998) suggests that differences in aggregate MFP are the main driver of international income differences found both across countries and over time. This point is further illustrated by Klenow and Rodriguez-Clare (1997), who conduct a cross-country analysis of 98 nations and suggest that 90% of the country divergence in growth of output per worker correspond to disparity in MFP growth.¹

In turn, aggregate productivity growth depends closely on MFP at the firm level. If firms increase the efficiency with which they turn inputs into outputs, they can contribute to overall efficiency gains. However, empirical evidence finds substantial heterogeneity in MFP across firms, even within narrowly defined industries. In the US manufacturing sector, the MFP ratio amongst an industry's 90th and 10th percentile plants is on average 1.92, implying that plants in the 90th percentile roughly make twice the amount with the same inputs as plants in the 10th percentile (Syverson, 2004). Such dispersion is not only found in developed countries but also in developing countries such as India and the People's Republic of China (hereafter "China"). Hsieh and Klenow (2009) find that the ratio of MFP between 90th and 10th percentiles in manufacturing industries is on average more than 5:1.

In light of the large dispersion of firms' MFP, analysing industry average productivity does not offer the complete picture: countries might display the same average but very different underlying distributions. This fact has important policy implications. For instance, low average productivity can be explained by too few firms at the top (lack of innovation) or by too many firms at the bottom (weak market selection), two different situations that would entail very different policies. To better inform policy makers, it is therefore essential to understand how firm-level productivity patterns translate into aggregate productivity.

Economists typically attribute differences in MFP across firms to either slow technology adoption or inefficient technology usage.² In addition, a growing body of literature attributes high aggregate MFP not only to the efficiency of technology use and speed of adoption but also to the efficient allocation of resources across firms. Resource allocation can impact aggregate productivity when there is a flow of inputs from low- to high-productivity firms. Conversely, when factors are allocated in an inefficient manner, aggregate productivity is adversely affected. These important issues have been investigated by two intrinsically interrelated branches of the literature. The reallocation literature typically focuses on the drivers of resource reallocation, such as creative destruction, and upscaling and downscaling

of firms, together with the factors that may influence them, such as technological change, regulation and recessions. The misallocation literature typically identifies a specific distortion or a bundle of distortions (policies and/or institutions) and examines the extent to which they adversely impact aggregate productivity. The results obtained in the misallocation literature show that distortions in the economy can have a quantitatively important effect on aggregate productivity.³

The analysis of MFP can also provide fundamental insights on more short-term fluctuations of output over time. The variations in MFP growth are in fact key to explaining changes in output over time. Disparities in productivity growth have been further accentuated by the Great Recession of 2008, a large-scale shock with many countries experiencing a substantial contraction in aggregate output (OECD 2014). Hall (2015) for example finds that in 2013 US output was 13 percentage points under its 1990-2007 trend path. A significant contribution to this shortfall, accounting for 3.5 percentage points, comes from productivity growth, which the author claims is unlikely to converge towards its earlier growth path. In the United Kingdom, labour productivity has remained remarkably weak after the recession. Firm-level evidence suggests that labour productivity growth is 14 percentage points under its pre-recession trend (Barnett et al., 2014). Therefore, understanding the impact of productivity dispersion and resource allocation at different points of the economic cycles is at the core of the current policy discussion, as they play a central role in shaping the welfare of societies and the competitiveness of countries.

The Directorate for Science, Technology and Innovation (STI) at the OECD aims to contribute to this debate by providing new policy relevant evidence through the MultiProd project. This project provides a comprehensive picture of productivity patterns across countries over the last two decades.⁴ The length of the period considered and the breadth of country coverage allows for the analysis of trends in productivity before, during and after the Great Recession, in a large set of countries. In particular the project collects micro-aggregated data and moments of the productivity distribution that allow for a cross-country analysis of i) productivity heterogeneity; ii) allocative efficiency; iii) misallocation; iv) aggregate productivity growth and v) the link between productivity and wages.

The project is part of a larger effort within the OECD STI Directorate to exploit existing official sources of confidential data at the firm level from official surveys and administrative sources (e.g. Business Registers) in a harmonised framework to provide non-confidential micro-aggregated statistics and analysis that are comparable across countries.⁵ One of the main contributions of the MultiProd project is indeed to build cross-country harmonised micro-aggregated data of paramount importance for understanding differences in productivity performance across countries. The project relies on a *distributed microdata methodology* and the micro-aggregated results (at the cell level) are collected, checked and analysed at the OECD. An important aspect of the methodology is to ensure the comparability of results across countries. Therefore, productivity is measured in exactly the same way across countries, conditional on the available data. Ensuring harmonisation across countries entails an intrinsic trade-off between detail and coverage; for the sake of maximising the participation rate and the number of policy issues addressed by this phase of project, the choice has been made to favour breadth of coverage to depth of the information (e.g. information on skill of the workforce or hours worked have not been considered in the current version of the code).

To further ensure harmonisation and representativeness, in particular for countries where MultiProd relies on production surveys, an appropriate set of weights is built using information from business registers, which typically cover the whole population of firms. We use these to reweight production surveys in order to construct statistics representative of the whole population, which can then be comparable to statistics from countries where production data cover the whole population of firms. MultiProd is to our knowledge the first project to implement such reweighting.

The output of the algorithm is a collection of statistics at different levels of aggregation (e.g. at the 2-digit industry level) over the 1994-2012 period, depending on data availability for each country. At the time of writing, 18 countries have been successfully included in the MultiProd database, namely: Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Hungary, Italy, Indonesia, Japan, Luxembourg, the Netherlands, Norway, New Zealand, Portugal and Sweden.

The main variables are levels and growth rates of employment, investment, value added, labour productivity, MFP, and other firm characteristics at different percentiles of the firm productivity distribution, further refined by size, age, and ownership categories. Moreover, measures of within-industry productivity dispersion and the covariance with size (Olley and Pakes 1996), as well as structural measures of misallocation based on Hsieh and Klenow (2009) and Petrin and Sivadasan (2013), are produced to analyse the role of allocation and selection. Finally the project provides results from distributed regressions, with the aim of establishing a set of stylised facts for each country regarding the relationship between productivity, firm characteristics (size, age, previous performance, ownership, etc.) and structural characteristics (concentration, misallocation, sectoral policies, etc.).

The output of the project allows for various decompositions of aggregate productivity level, growth and dispersion to understand the role of particular industries or groups of firms in explaining aggregate outcomes (e.g. small vs large; multinational corporations; old vs young; low vs high productivity; etc.). For instance, changes in the overall productivity dispersion are decomposed to quantify how much of an increase in dispersion is due to an increase in within-industry dispersion and how much comes from a reallocation of resources to industries characterised by a higher dispersion. Moreover, the role of the largest firms is investigated in great detail and compared to that of the most productive firms (“frontier firms”). In an undistorted economy, the two sets of firms should coincide, so the analysis is another way of unveiling the presence of misallocation in the economy. The analysis of the largest firms in each sector can also shed light on the so-called “granular” hypothesis, which posits that aggregate fluctuations are the results of microeconomic shocks and not economy-wide shocks, as usually assumed. Such idiosyncratic shocks, even if they are uncorrelated, may not cancel on average if sectors are dominated by a small number of large firms. This fact can have important implications for policies aimed at increasing economic resilience, and highlights the importance of studying firm-level data to better understand aggregate outcomes. The MultiProd project can offer new insights on this hypothesis by analysing how much economic activity is concentrated in a small number of large firms, and how much of the observed productivity variation is indeed the result of microeconomic variations.

MultiProd also attempts to shed light on the nature of wage inequality across countries. This is of particular importance for the current policy debate given the large increase in wage inequality experienced in many countries, in particular the United States and the United Kingdom over the last 30 years (Atkinson and Piketty 2009, Machin and Van Reenen 2008). Thanks to the MultiProd output we are able to examine the evolution of the between plant/firm component of wage dispersion, which has been found to account for a large share of the wage inequality of individuals (see, for example, Dunne et al., 2004, Card et al., 2013, Card et al., 2014, Song et al., 2015). MultiProd decomposes the overall wage dispersion in the within and between contribution both for industry and productivity quantiles, and calculates the share of each industry or productivity quantile in the within (industries or quantiles) component of wage dispersion. Moreover, one can identify the effects of various policies (e.g. employee protection legislation, minimum wage, coordination in wage setting) on the dispersion of wages and productivity and the interaction of the two, as well as examine the extent to which wage variance is driven by the development of new technologies and/or increased capital intensity.⁶

MultiProd contributes to the literature in several ways. First, the project allows us to better understand the Schumpeterian process of creative destruction across countries, gauge whether resources are efficiently allocated, and ultimately unveil the ingredients of dynamism at the heart of economic

growth. Second, thanks to the data on the entire productivity distribution, it is possible to identify the group of firms at the frontier from other firms in their own country, as well as document the evolution of the gap between the two groups of firms.⁷ Third, the data also allows for a comparison of the characteristics of “national” frontier firms in different countries, as well as what drives their performance. Fourth, the time frame of the analysis permits an investigation of cross country differences in firm-level productivity performance and in allocative efficiency before, during and after the recent Great Recession. Fifth, the effectiveness of various policy frameworks aimed at shaping firm productivity and enhancing resource allocation to more productive firms (i.e. allocative efficiency) can be examined. Identifying the nature of misallocation at the firm level and across countries allows the OECD to make evidence-based policy suggestions, which could help enhance aggregate productivity growth. Finally, the analysis sheds new light on the link between productivity dispersion and wage dispersion, and examines structural factors that can exacerbate or dampen that link.

The MultiProd project has addressed a series of methodological issues that span from defining the main variables of interest, such as labour and multifactor productivity, to constructing measures of capital stock and to defining the sampling weights to obtain results representative for the entire population of firms. We detail some of these methodological considerations in Section 2, and leave the more technical or code-related issues to the Appendix and to the documentation that accompanies the source code distributed to countries.⁸ The following sections of the paper outline the pillars of the MultiProd project in the context of the current literature and of the policy questions that the project aims to answer by exploiting the novel data that are collected via the MultiProd code. For each of them, the paper describes the main methodological tools used to carry out the analysis, and provide some preliminary examples of the analyses that can be performed with the MultiProd output.

More specifically, the rest of the paper is structured as follow: Section 2 provides an overview of the methodology used, the data, and the challenges faced in the implementation of the MultiProd project. Section 3 describes the methodology adopted to analyse the evolution of productivity dispersion across sectors and over time. Section 4 discusses how MultiProd can provide new evidence on allocative efficiency, focusing both on measurement and policy issues. Section 5 provides details on the analysis of granularity and the contribution of the largest firms to productivity.⁹ Section 6 provides an overview of novel evidence and analysis linking productivity dispersion and wage inequality shedding new light on the role of policies for both productivity growth and inequality. Finally, Section 7 concludes.

2. The MultiProd dataset

2.1. The distributed microdata approach

In recent years, the policy and research communities' interest in harmonised cross-country microdata has increased significantly. This has been partly driven by improvements in computing power but, fundamentally, it reflects the recognition of the need of microdata for understanding the growing complexity in the way economies work and the heterogeneity in economic outcomes.

Significant obstacles remain, however, for transnational access to official microdata. As a result, cross-country studies based on the analysis of official microdata are rare; where they are performed, it is generally via the formation and coordination of networks of national researchers, with each team having access to their respective national microdata. Therefore, the comparability of the country level results needs to be ensured via the use of a commonly specified protocol for data collection and aggregation, and a commonly specified model for the econometric analysis. This method is called *distributed microdata approach*: it is a method of collecting statistical moments of the distribution of firm characteristics (employment, productivity, wages, age, etc.) by a centrally written routine that is flexible and automated enough to run across different data sources in different countries. The advantages of this novel data collection methodology are manifold: it puts a lower burden on individual national statistical agencies through the development of a common programme; in a related vein, the development of the programme allows for low-cost reproducibility; it also overcomes the confidentiality constraints of directly using national micro-level statistical databases; and finally, it allows for a high degree of cross-country harmonisation and comparability. It was pioneered in the beginning of the 2000s in a series of cross-country projects on firm demographics and productivity (Bartelsman, 2004; Bartelsman, et al., 2005; Bartelsman, Haltiwanger, et al., 2009).

The OECD currently follows this approach in three ongoing projects: *MultiProd*, *DynEmp*, and *MicroBeRD*.¹⁰

2.2. Structure of the program and basic requirements

2.2.1. Data needs

The data collected in MultiProd are computed by running a standardised STATA® routine on firm-level data. The program produces a set of statistics based on micro-level longitudinal information on output, inputs (labour and capital), labour costs, sector of activity, and possibly age and ownership of the firm.¹¹ The information is used to calculate firm-level labour and multifactor productivity (MFP), which are then aggregated to the 2-digit sector level, separately for each year. Moreover, some statistics are further refined by age and size classes, ownership characteristics, quantiles of the productivity distribution, and quantiles of the size distribution (defined in terms of sales or employment).

The program relies on two main file sources:

- a. Administrative data or production surveys (PS), which contains all the variables needed for the productivity analysis but may be limited to a sample of firms;¹²
- b. A business register (BR), which contains a more limited set of variables (mainly employment, sector of activity, age and ownership) but for the entire population of firms. The program also works if this information is not available. However the presence of a BR substantially improves the representativeness of the results and thus their comparability across countries, in particular when only a PS is available. Given the coverage and the continuity of the administrative information collected for each firm, BR data allow for: i) the calculation of the population

breakdowns that are necessary for obtaining the weights used in the analysis; ii) a much more precise treatment of entry and exit; iii) the calculation of more precise industry modes and conversion tables in case of changes in the industry classification at the firm level or of a change in the whole industry classification system.

The program requires minimal intervention by the participants, but the input data should respect a number of basic requirements (see the Appendix for further details):

1. The data source must be longitudinal. In the case that multiple data sources are available, the preferred one is the one that provides the best trade-off between coverage of units, time span and quality of information on inputs (employment, capital stock and intermediate inputs), output (gross output and value added) and labour costs.
2. Panel dimension: individual units need to be identified by a unique longitudinal identifier (id) that has to be constant and unique over time. If the unit exits the production survey, its identifier must not be re-assigned to another firm.
3. Sectoral coverage: the aim is to cover the full non-financial market sector, especially manufacturing and non-financial market services. If available, the code collects information at the 2-digit level for all sectors of the economy.
4. The minimum set of required variables is:
  - A unique firm identifier.
  - A calendar year to which the time-varying variables refer to.
  - A 3 or 4 digit sector identifying the main economic activity of the unit, following the ISIC (NACE) classification. While ISIC rev. 4 (NACE rev. 2) classification at least for more recent years is preferable, the program can deal with the dataset being partially or completely classified according to the ISIC rev. 3.1 (NACE rev. 1.1) classification, as outlined in more detail below.
  - Birth year of the unit. This can be outside (i.e. earlier than) the period covered by the business register or production survey. It can also be missing for some units or censored. In case the birth year is missing, the first year in which the unit appears with non-zero employment in the Business Register is considered as the entry year.¹³
  - Firm employment in headcounts or full time equivalents. Preferably, two different variables should be available: one for employees and one for persons engaged (including also working proprietors, unpaid family workers, part-time, seasonal or home workers, apprentices, etc.), ideally both should be expressed in full time equivalents. However, given constraints on data availability and for cross-country coverage and comparability, the program gives the possibility of providing the variables in headcounts. In any case, the program runs regardless of whether employment is: a) expressed in headcounts or full times equivalents, and b) is defined as employees or persons engaged only, or as two separate variables.
  - Total labour costs of the firm. The preferred measure is the total gross wage bill including social security contributions, for the same group of employees included in the employment variable(s). Average wage is simply obtained by dividing total labour costs by the number of

employees. This is the reason why it is important that the two variables (employees and labour costs) cover the same set of workers.

- Gross Output (GO) of the firm, can be sales if gross output is unavailable.
- Value Added (VA) of the firm preferably at basic prices if available. VA is our preferred measure of output used in the productivity measures. The program expects this variable to be available in the dataset, so the user should calculate it from its components (e.g. gross output minus intermediate inputs, and possible adjustments) before running the MultiProd code. As an alternative output measure, the program also uses gross output (approximated by sales if gross output is not available). The current version of the code is set to calculate Labour Productivity (LP) using VA only, while MFP is computed using both VA and GO depending on the methodology used (see discussion in Section 2.3).
- Either investment or capital stock must be provided otherwise MFP cannot be calculated. The preferred measure is annual investment (total investment across all asset classes), since the program obtains the capital stock through a common approach based on the perpetual investment method (PIM) to increase the comparability across countries. Nevertheless if an alternative capital stock measure (e.g., based on book values) is available in the micro data, it should be provided as the program uses it for: a) initialising the PIM; b) carrying out a robustness analysis. In case investment is not available, capital stock is used to carry out the main analysis.

All numeric variables must be expressed in single units and in local currency; the monetary variables are then transformed in 2005 U.S. dollars by the program.¹⁴

2.2.2. Modularity of the code

The MultiProd program has a modular structure that features independent and interchangeable modules, such that each of them contains everything necessary to execute only one specific aspect or functionality of the program. This makes the program more flexible and reusable than a single routine software: the modules can be easily adapted to different data and may also be reused (without change) in other projects. This also facilitates the "breaking down" of the project into several smaller projects, allowing to skip/modify some of them according to the needs and specific requirements and data availability and to combine with new modules to respond to different policy questions.

In particular, the MultiProd program is composed by a series of STATA® ado files. The master file is called *_MultiProd.ado*. For simplicity all inputs and options of the program are specified in an external do file named *MultiProd_Settings.do*, where they are collected through a series of STATA® global macros. The user has to fill in the options and then execute the *MultiProd_Settings.do* file; the action will launch the *_MultiProd.ado* file, which contains a series of internal settings and calls a series of other ado files that compute all the statistics generated by the program. All the details of the internal structure are contained in the documentation that accompanies the code package distributed to countries. The output files are stored in the output folder specified by the user in the program options; these files contain only micro-aggregated non-confidential data that are then transmitted to the OECD by the participating countries.

2.3. Measures of productivity and capital

2.3.1. Productivity definitions¹⁵

Productivity, in its broadest interpretation, is meant to capture the efficiency with which inputs are turned into outputs (Hulten, 2001). The MultiProd program computes a series of productivity measures that go from the least to the most data-demanding methodologies; they are summarised in Table 1.

The gross output (GO) or total revenue based LP aims at capturing the amount of output produced by a firm for a given amount of labour input. It is the crudest measure but potentially offers the highest coverage among all productivity variables. Note that even though gross output differs from total revenues in case of changes in inventories, both measures are admitted in MultiProd since their correlation tends to be very high at annual frequency. Of course one of the main problems with a gross output based LP measure is that it does not control for intermediate input usage.¹⁶ As a result, a company with significant reselling activity will probably rank very high in this measure.

Value added (VA) based LP alleviates this problem, as value added itself is the difference between output (or sales) and intermediate inputs (including resold goods, typical in retail trade). Of course, neither type of LP controls for differences in capital intensity across firms or differences in other inputs. Multifactor productivity (MFP) measures control for a wider set of inputs, using various methodologies. In MultiProd, two broad types of MFP measures are considered: index number approaches and the methods based on production function estimation.

MFP measures can take gross output or value added as the output variable, and capture both labour (measured by the number of employees or persons engaged), capital (measured by the total depreciated capital stock), and intermediates (when gross output is used, measured as total cost of intermediates) as inputs. The MultiProd program is flexible to produce both. Introducing more sophisticated measures either on the labour side (e.g. using the number of hours, workers with different skills) or on the capital side (e.g. using different types of assets, relying on capital services) would have put additional constraints in terms of country, sector and time coverage and comparability, as their availability is more limited. The choice of simpler but cruder input measures has been driven by the aim of collecting comparable statistics across countries while achieving the broadest possible coverage.

Index number measures relate output to a weighted sum of inputs, generally assuming constant returns to scale.¹⁷ They are relatively easy to calculate. Two widely used index numbers are the Solow-residuals and the Superlative-index measures.¹⁸ The first measure weights inputs using external, industry specific shares. The latter uses as weights the average between firm-level factor shares and a reference value in the industry (an average across firms or a specific firm, e.g. the median), and implicitly compares productivity levels to that reference value. As such, it is more prone to measurement error than the Solow-residual measure (see Van Biesebroeck, 2007). MultiProd provides four index-number productivity measures: a) MFP_S, a Solow-residual based MFP using external, country-industry specific labour and intermediate shares; b) MFP_SW, a Solow-residual based MFP using external, industry specific labour and intermediate shares (the cross-country-year median); c) MFP_SI: a Superlative index based MFP using labour and intermediate shares calculated as the average between the labour/intermediate share of the firm (averaged over time) and the geometric mean of firm labour/intermediate shares in the industry; d) MFP_GR: a Superlative index based MFP using firm-level labour and intermediate shares calculated as a Törnqvist discrete time approximation of Divisia shares (average over two consecutive years, MFP_GR is obtained for MFP growth only). As suggested by Syverson (2007), the labour and intermediate shares are taken as constant over time in all cases but d).^{19,20} This makes the empirical implementation more consistent with the underlying model, which assumes away any factor adjustment costs. Moreover it makes cross-country comparison easier, particularly in the case of MFP_SW.

Turning to the production function based measures of MFP, the simplest benchmark is the residual from an OLS regression of the production function, usually run separately for each industry. However, the estimates obtained through this exercise are likely to be inconsistent and biased. In the case of the labour coefficient, for instance, firms with higher productivity hire more workers. Productivity is not directly observed and it enters the error term. Firms' behaviour will therefore introduce a positive correlation between the error term and the labour input, yielding standard OLS estimates that are inconsistent and biased.²¹ Fixed effect estimators can only partially solve the problem, since they take into account only a time-invariant firm-specific productivity effect and cannot control for cases where productivity shocks vary across firms and time.

A well-known solution for this problem was proposed by the semi-parametric approach of Olley and Pakes (1996), who took investment as a proxy variable for unobserved MFP.²² However, investment proves to be a 'lumpy' variable and hence potentially a poor proxy for productivity.²³ The two-step method of Levinsohn and Petrin (2003) takes a similar approach, but instead of investment, it uses intermediate inputs as proxies. However, Akerberg et al. (2006) pointed out that the identification of the coefficient on the variable input is likely to be problematic in Levinsohn and Petrin (2003).²⁴

More recently, Wooldridge (2009) developed a technique that builds upon the Levinsohn and Petrin (2003) method but also overcomes the identification problem. It is a quicker one-step procedure that obtains consistent standard errors without bootstrapping. This is the methodology adopted in MultiProd and is implemented using a value added based MFP only, as in Akerberg et al. (2006). The reason for this approach is that there are more severe identification issues when using gross output as dependent variable of the production function estimation to derive MFP (see Bond and Söderbom 2005).

Overall, the theoretical advantage of the estimation based methods (no assumption of constant returns to scale needed) comes at a cost: their data requirement is usually higher, both in terms of the number and the quality of the available variables. Finally, it is interesting to note that, irrespective of the sophistication of the measures, it has been shown that firms' productivity levels and dynamics vary a lot, even in narrowly defined industries (see Bartelsman and Doms, 2000, and Syverson, 2011).

Although all these MFP measures are implemented in the code, version v.1.00 of MultiProd is actually set to compute only two measures of MFP: an index-based measure (MFP_SW: Solow residuals using cross-country industry-specific labour and intermediate shares) and the semi-parametric measure of Wooldridge (MFP_W).

Table 1. Productivity measures used in MultiProd

Productivity measures		Definitions ⁱ
Labour productivity (<i>LP</i>)	Based on gross output: <i>LP_GO</i>	$\frac{\text{Real gross output (or sales, } GO)}{\text{Employment (L)}}$
	Based on value added: <i>LP_VA</i>	$\frac{\text{Real value added (VA)}}{\text{Employment (L)}}$
Multi factor productivity (<i>MFP</i>): index-numbers ⁱⁱ	Solow-residual <i>MFP_S</i> with country and industry specific labour/intermediate shares (β_l^S, β_{ii}^S : median of shares across each year by country and industry)	$va - \beta_l^S l - (1 - \beta_l^S)k$ or $go - \beta_l^S l - \beta_{ii}^S ii - (1 - \beta_l^S - \beta_{ii}^S)k$
	Solow-residual <i>MFP_SW</i> with industry specific labour/intermediate shares ($\beta_l^{SW}, \beta_{ii}^{SW}$: median across countries and years by each industry)	$va - \beta_l^{SW} l - (1 - \beta_l^{SW})k$ or $go - \beta_l^{SW} l - \beta_{ii}^{SW} ii - (1 - \beta_l^{SW} - \beta_{ii}^{SW})k$
	Superlative index <i>MFP_SI</i> , following Griffith et al. (2009) ⁱⁱⁱ	$\bar{va} - \beta_l^{SI} \bar{l} - (1 - \beta_l^{SI}) \bar{k}$ or $\bar{go} - \beta_l^{SI} \bar{l} - \beta_{ii}^{SI} \bar{ii} - (1 - \beta_l^{SI} - \beta_{ii}^{SI}) \bar{k}$
	Superlative index MFP growth: <i>MFP_GR</i> ^{iv}	$\Delta va - \beta_l^{GR} \Delta l - (1 - \beta_l^{GR}) \Delta k$ or $\Delta go - \beta_l^{GR} \Delta l - \beta_{ii}^{GR} \Delta ii - (1 - \beta_l^{GR} - \beta_{ii}^{GR}) \Delta k$
Multi factor productivity (<i>MFP</i>): production function based ⁱⁱ	Residuals from Ordinary Least Squares estimates <i>MFP_OLS</i>	$va - \beta_l^{OLS} l - \beta_k^{OLS} k$ or $go - \beta_l^{OLS} l - \beta_{ii}^{OLS} ii - \beta_k^{OLS} k$
	Residuals from Ordinary Least Squares with firm fixed effect estimates <i>MFP_FE</i>	$va - \beta_l^{FE} l - \beta_k^{FE} k$ or $go - \beta_l^{FE} l - \beta_{ii}^{FE} ii - \beta_k^{FE} k$
	Residuals from Wooldridge (2009) type estimates <i>MFP_W</i>	$va - \beta_l^W l - \beta_k^W k$

Notes: All output and input variables are firm and year specific, but for simpler exposition, firm and time indices are omitted.

i) Small letters (*va*, *l*, *k*, *ii*) denote natural logarithms of, respectively, Value Added, Labour, Capital and Intermediates.

ii) For simpler exposition, MFP measures are defined in natural logarithms.

iii) Variables with a tilde on top ($\bar{va}$, $\bar{l}$, $\bar{k}$) indicate log-differences from the industry and year mean value. β_l^{SI} , β_{ii}^{SI} denote the average of the labour and intermediate share in the reference firm (taken to be the geometric average across firms in the same country*industry) and the current firm.

iv) Variables with a Δ indicate log-growth. β_l^{GR} , β_{ii}^{GR} denote the two-period average (e.g.: $\beta_l^{GR} = \frac{\beta_{lt} + \beta_{lt-1}}{2}$) of the firm-level labour and intermediate share.

2.3.2. Capital Stock Calculation

For the MFP calculations a measure of capital stock is needed. In the baseline case, the program defines the capital stock variable through the perpetual inventory method (PIM) in order to increase the comparability of results across countries. Then, in case an alternative measure of capital stock is directly available in the micro data, the program uses it to initialise the PIM and carry out a robustness analysis.

Perpetual Inventory Method

In the baseline case the code obtains capital using the Perpetual Inventory Method (PIM) methodology. In order to use this methodology firm-level investment data, preferably over a sufficiently long time period, are needed. Moreover, as common in the literature (e.g. Olley and Pakes, 1996, Levinsohn and Petrin, 2003), the initial value is set to the capital stock reported by the firm in the initial year, whenever this is available.²⁵ If, on the other hand, the capital stock is not available (or missing in the first year), the code calculates the initial capital stock as the average of two proxies described below:

1. Average firm investment in all years divided by the depreciation rate (from STAN), borrowing a similar idea from Mueller (2008);
2. Country-industry-year specific capital-labour ratio (K/L) from STAN multiplied by firm-level employment (L).

The two proxies for the initial capital stock have a correlation of 0.7 in the micro data used for testing, and, as expected, the capital stock obtained from the industry average capital-labour ratio is usually larger since it tends to overestimate capital for smaller firms. The code has an option to set the first n years of the calculated capital stock to missing, in order to drop the initial noise due to the PIM method.

The code simply treats consecutive observations as consecutive years; this implies that whenever the firm is missing in some years, the PIM method essentially assumes that capital is constant in the years in between. A similar situation arises when it is only investment to be missing in some years, in fact the code sets investment to zero whenever missing and the capital stock evolution will depend only on depreciation in those years (unless it is missing for all years for that firm, in which case investment is set to missing).

Alternative measure

For those countries whose micro data already include a measure of capital stock (from book values or other measures of capital stock obtained by country researchers), the code creates a second measure of capital (K_{Rob}). The code fills in missing values by interpolating the provided information, which makes K_{Rob} coherent with the measure of capital (K) built using the PIM. In the micro data used for testing, the two capital stock measures have a correlation of 0.96. For the countries that have the capital stock information but not the investment data, K_{Rob} becomes the main (and only) measure of capital.

2.4. Challenges

2.4.1. Weighting

Since the data sources for productivity analysis may cover only a subset of the population (e.g. production surveys), MultiProd adopts a weighting strategy for the firm level samples coming from production surveys. This approach makes the results representative for the entire population of firms and comparable across countries. The strategy entails the following two main steps:

1) Preparation of the population structure from the Business Registry (BR)

The code prepares the population structure (number of firms by year, industry, size class) from the BR or from census years. The industry used to obtain the population structure is the most detailed available. The number of size classes that are used to define the population breakdowns is 7 in the baseline, with thresholds at 5, 10, 20, 50, 100 and 250 employees.²⁶

The program also adopts a special treatment for variables like annual growth rates or 3/5 year annualised growth rates. Several population structure variables are created depending on which lag-length of growth rates are used in the analysis; this is done in order to reflect the actual availability of "growth rates" in the population of firms (for instance, in case of a "true" exit, a firm would have missing growth rates even if the population of firms were available).

2) Calculation of actual weights

The weights are based on the population information contained in the business register (whenever available), but the actual calculation takes place after the outlier filtering (see below) and just before the final output statistics are calculated, so that variable specific weights can be computed. The weights are then calculated as the number of firms in the population divided by the number of firms in the survey, after the cleaning. All calculations are performed by industry and size class, which are the same used in the first step described above.

2.4.2. Deflation and PPP

MultiProd uses OECD STAN (ISIC4 version, including preliminary data) for obtaining deflators, depreciation rates, capital-labour ratios, as well as LP levels and growth rates at the country, industry year level. As a preparatory step, the STAN database has been checked for possible inconsistencies and, in order to have a balanced panel for each series, missing values at the 2 digit industry level have been filled up with values from industries at a more aggregate level (except LP). The deflators for gross output, value added, intermediates, and gross fixed capital formation are available at the sectoral level; while we use the Consumer Price Index at the country-level for wages. All variables are deflated using their own deflator with the exception of capital and investment, which are deflated using the same deflator.

Every nominal monetary variable is transformed into real 2005 US dollars. To ensure international comparability, exchange rates and PPPs are therefore applied in that year. Since most of the manufacturing sector is usually traded internationally, the series for manufacturing are simply adjusted using the nominal exchange rate (as average over 2005), taken from Eurostat. On the other hand, all other sectors are adjusted using country-level PPPs for 2005 from the Eurostat-OECD PPP Programme.²⁷

Therefore the national firm-level data used to run the MultiProd code, all monetary variables should be in single units of the local currency; for the countries that adopted the euro the variables should be in euros using the fixed official exchange rate at the time of accession. For countries that are not present in the STAN database, the WIOD dataset has been used for obtaining deflators and depreciation rates, whereas the STAN IO tables have been used for labour and intermediates shares; in addition, ad-hoc solutions using directly national data sources have been found to provide missing values for these variables.

2.4.3. Industry conversion: probabilistic matching

As stated before, the program requires a 4 or 3 digit sector variable identifying the main economic activity of the unit, following the ISIC rev. 4 (NACE rev. 2) classification.²⁸ The program can also deal with the dataset being partially or completely classified according to the ISIC rev. 3.1 (NACE

rev. 1.1) classification, in the latter case using an external conversion table provided in the MultiProd package. On the other hand the program does not implement the conversion from country-specific industry classifications into ISIC.

Whenever available the transformation is performed on the BR and then the industry classification information is merged back – for each firm – to the PS. If a Business Register is given, it should contain the same industry variable: same variable name, type and content. If it is not possible to merge BR and PS, then the code implements industry classification changes using the PS only.

Generally, the code assigns a single industry to each firm (the mode, i.e. the most frequently recorded for each firm, and if this is not unique the most recent value). If only one industry variable is specified in the code setting, the code assumes that the same industry classification system is used in all years and there is no break in the classification system. There are two possibilities in such a case: i) if the firms are already classified according to the new system then no transformation is needed, the firms are simply assigned to their most frequent industry; ii) if firms are classified according to the old classification system an external conversion table is used. In this case, the data must follow either NACE Rev 1.1 or ISIC Rev 3.1 because the external conversion tables (obtained from ORBIS) are available only according to these classifications.

In case of a break in the industry classification during the period of analysis, the code back casts the new system onto earlier years. To achieve this, the code first creates a conversion table that exploits either the information of all firms that are active both before and after the break or the firms that report both the old and the new industry classification in the break year. For those firms that appear in the data only prior to the break or those firms whose most frequent industry is classified in the old system, the new industry is assigned using the constructed conversion table and randomly allocating the firms to the possible new industries to match the fractions observed in the conversion table.

2.4.4. Data cleaning

The program carries out a series of basic consistency checks and cleaning of the data:

- **Duplicates.** The program checks for duplicates in terms of firm identifier and year. In case duplicates are found, they are not treated and the program simply stops and warns the user.
- **Units of measurement.** A pop-up question at the very beginning of the code asks whether all monetary values are in local currency and single unit (not thousands).
- **Variables availability.** If key variables are not available in the production survey, the program checks whether they are in the business register; if so it takes them from there. The code also checks the availability of the variables used for MFP calculation and verify whether the robustness measures of MFP (using deflated capital stock) can be calculated. If neither capital nor investment is detected, the program warns the user because MFP cannot be calculated; if the user decides to proceed the output of the program will be based on LP only.
- **Industry variables.** The program checks for the consistency of the industry variables. If the industry variable is available only in the old classification system, it asks whether it is NACE Rev. 1.1 or ISIC Rev 3.1, because the external transformation file works only for those two systems. If the code detects that the industry classification is at 2 digits, it explicitly asks the user whether a more detailed industry classification is available.

- **Birth.** In case the birth year of the firm is missing, the variable is set to be equal to the first year in which the firm reports positive employment (unless it is the first year in the database). The code also corrects for cases in which the reported birth year is more recent than the first year in which the firm reports positive employment. Finally the code allows for left censoring, and the birth is set to missing in case it coincides with the year of censoring.
- **Implausible jumps.** The code treats observations with implausible jumps (consecutive large one-off increases and large one-off decreases or vice versa) in the main input variables of interest (employment, value added, total revenues, wage bill). Specifically, the code drops the observation in case of one-off jumps by factor of 50 (30 in case of labour costs). In the case of employment, the jump factor parameter is 10 and the adjustment is performed only for firms with more than 20 employees. Moreover the code drops the observations in case of large falls after entry; this is done symmetrically for all variables whenever they drop by a factor of 10 in the two years after entry. If a jump is detected, the observation is dropped entirely rather than simply setting the relevant variable to missing because, according to experience on the data, it is difficult to assess whether it is a problem related to the single variable or to the entire record. A very similar strategy is adopted for cleaning the capital stock, once it is computed by the program. The jump parameter takes again a value of 50.²⁹
- **Missing or zeros.** The code drops any observation with missing, zero or negative employment. Moreover it sets FTE employment, wage bill, total revenue, and the capital stock (if available) to missing whenever they take zero or negative values (this is not performed for value added because the program keeps also negative values of VA to characterise the overall distribution in levels of VA/L). The code does not drop records with missing or negative values of these variables because the observation might still contain useful information for other variables (e.g.: if employment in full time equivalents is missing, it might be available in headcounts).

Outlier filtering of final computed variables. The code has also a routine to remove outliers from the main variables of interest. This applies only to internally computed variables (e.g. productivity growth) and is aimed at eliminating any outlier that might affect the results. The focus is on growth rates, hence on the within-firm rather than the cross-sectional variation. This is to avoid throwing out correct and potentially useful information, for instance a very large firm would be dropped if one simply looked at the cross sectional variation in employment. In the baseline case the value is not corrected and the variable is set to missing in that year. For the specific case of productivity variables, the outlier filtering is also applied to the (log-) levels in order to eliminate outliers in the cross-section. This is done mainly to avoid outliers to be included in the group of firms at the frontier.

There are three different methods to identify outliers:

- The first is the “SD” method that defines as outliers the observations that lie beyond a certain number of standard deviations from the mean. This is the default in the code.
- The “Percent” method identifies outliers simply as the observations that lie below the 1st percentile and above the 99th percentile.
- The third is the “Tukey” (1977) or boxplot method, which does not rely on any distributional assumption and identifies outliers as the observations that lie below or above a certain distance from the lower (25th percentile) or upper (75th percentile) quartile, respectively.

In the micro data used for testing the three methods drop a similar share of total observations: 4.3% in case of the SD method, 4.7% in case of the “Percent” method, and 5.2% in case of the “Tukey” method.

2.4.5. Confidentiality

The program can deal with confidentiality in four different manners.

1. First, the program can apply a threshold rule, through an option of the program called *MinElementsInCell*: if this option is specified, the code blanks cells for statistics based on less than X units, where X is specified by the user in the option *MinElementsInCell*.
2. Second, the program can implement a dominance rule if a different option, called *ApplyDominance*, is set to 1. When this is the case, the program blanks cells in which the largest firm or the largest two firms account for more than Y% of the cell’s total, in term of employment (L) or output (GO).
3. Third, if the option called *DisturbPercentiles* is set to 1, the code applies a random disturbance around median and percentiles values (e.g. +/- 0.1% of the value).
4. Finally, a fourth option, called *AveragePercentiles* can be set to 1 whenever percentile values (e.g. the median) cannot be reported for confidentiality reasons. In this case the code replaces them by averages of a certain number of observations around the percentile (specified in the option *AveragePercentilesMinObs*).

2.5. Country coverage

At the time of writing, 18 countries have been successfully included in the MultiProd database (namely, Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Hungary, Italy, Indonesia, Japan, Luxembourg, Netherlands, Norway, New Zealand, Portugal and Sweden). A first version of the output has also been received from China, Costa Rica, and Switzerland. Brazil, Germany, Ireland, Spain, and United Kingdom are currently running the code. The data for each of the countries included so far are collected annually and at firm-level.

For most countries the time period covered by MultiProd spans between early 2000s and 2012. For Austria, Belgium, Chile, Luxembourg and Portugal the time horizon is shorter, whereas for Finland, France, Japan and Norway data are available since at least 1995. Details about time coverage by country are summarised in Table 2. Table 3 details instead the average number of firms and number of employees available each year, restricting to manufacturing and non-financial market services.

As stated before, if information on investment is available, the capital input is measured through the Perpetual Inventory Method (PIM); otherwise, the book value of capital is used. Moreover the labour input can be expressed in headcounts (HC) or full time equivalent (FTE). Table 4 details the availability of the employment and capital variables.

Table 2. Productivity measures used in MultiProd

COUNTRY	YEAR																		
	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
AUS																			
AUT																			
BEL																			
CAN																			
CHL																			
DNK																			
FIN																			
FRA																			
HUN																			
IDN																			
ITA																			
JPN																			
LUX																			
NLD																			
NOR																			
NZL																			
PRT																			
SWE																			

Table 3. Data coverage for manufacturing and non-financial market services sectors

Country	Years	Firms	Employees
Australia	2002-2012	68,499	761,602
Austria	2008-2012	255,701	2,258,626
Belgium	2003-2011	102,574	1,804,465
Canada	2000-2012	509,460	8,058,557
Chile	2005-2012	339,492	5,273,453
Denmark	2000-2012	80,030	1,281,035
Finland	1995-2012	85,038	981,772
France	1995-2012	812,850	11,453,356
Hungary	1998-2012	191,064	1,786,685
Indonesia*	2002-2012	22,907	4,316,496
Italy	2001-2012	312,057	1,893,156
Japan**	1994-2011	27,786	10,552,236
Luxemburg	2003-2012	1,136	105,252
Netherlands	2000-2012	39,375	332,449
Norway	1995-2012	63,593	890,001
New Zealand	2000-2011	90,973	992,208
Portugal	2004-2012	261,580	2,166,037
Sweden	2002-2012	176,652	1,889,764

Note: Manufacturing and services only. Numbers are averages across years. *: Indonesia includes only manufacturing firms with at least 20 employees. **: Japan includes only firms with at least 50 employees; a new run of the code on Census data will be implemented in the near future.

Table 4. Employment and capital data, recording units

COUNTRY	Employment		Capital	
	HC	FTE	PIM	Book Value
AUS				
AUT				
CAN				
BEL				
CHL				
DNK				
FIN				
FRA				
HUN				
IDN				
ITA				
JPN				
LUX				
NLD				
NOR				
NZL				
PRT				
SWE				

Data normally cover the whole population of businesses with at least one employee. Still, Census or administrative data do not always exist and PS data need to be used. One of the big challenges of working with firm-level production surveys is that the selected sample of firms might yield a partial and biased picture of the economy. Whenever available, BRs, which typically contain the whole population of firms, are therefore used to compute a population structure by year-sector-size classes; this structure is then used to re-weight data contained in the PS in order to construct data that are as representative as possible of the whole population of firms and comparable across countries.

To give an idea of the coverage of our sample, Table 5 reports for 2011 the share of firms and employment with respect to both the Business Registry (when available) and to Eurostat data (Annual business demography by size class database). The table is constructed for the manufacturing and non-financial market service sectors, and the data from Eurostat refer to the total number of firms in a country or the total number of firms with at least one employee in accordance with the micro-data used in MultiProd.

Table 5. Representativeness in 2011

Country Year 2011	Share of firms (%)		Share of employment (%)	
	BR	Eurostat	BR	Eurostat
Austria		69		92
Belgium*		70		97
Denmark		102		117
Finland**	98	100	98	100
France		101		108
Hungary		94		99
Italy	11	11	54	54
Netherlands	6	4	56	44
Norway		71		89
Portugal***		92		97
Sweden		94		90

Note: Share of business registers and Eurostat data present in MultiProd. Manufacturing and non-financial market services only. Year 2011. The data come from different sources and the methodology used to treat the data (i.e. cleaning) might differ, hence the comparison must be taken with caution. For instance, in countries in which we compare data with at least one employee, such as Denmark and France, shares higher than 100% are due to the different definitions of firms with at least “one employee” adopted in the two sources. *: A new run of the code on the whole population of firms (from VAT data) will be implemented in the near future. **: Finland is 100% of firms with at least 1 FTE. ***: data for Portugal include the population of active companies, but exclude individual enterprises (i.e., sole proprietors and self-employed); due to data limitations, we can only compare the MultiProd dataset to the total number of firms with at least one employee from Eurostat.

Comparing across different data sources is never easy, but data from Eurostat give a good benchmark to compare our data. The coverage is rather high in most of the countries. For those with a lower coverage the full BR is available, and thus the samples can be re-weighted. For instance Italy has a skewed distribution with a large mass of very small firms which cannot be captured by production surveys. The survey used by MultiProd contains only 11% of the total population of firms (both with respect to the BR and to the data published by Eurostat) but it accounts for 54% of total employment. At the same time we have access to the entire population of firms from the business register, which we use to reweight our sample moments. In the Netherlands the situation is similar, with the only existing survey of firms representing a very small share of firms, but the BR allows us to re-weight those firms ex-post in order to make the reported statistics representative of the total economy.³⁰

For further details on the country coverages and data sources please see the Metadata section in the Appendix.

2.6. Output

The MultiProd output is a collection of files that contain statistics for different variables computed yearly at a detailed sectoral level. The program is flexible to allow for multiple levels of aggregation at which the output is produced and, for each of those levels, the types of aggregated data to be included.

The statistics are never collected at the level of the individual firm. Instead the programme splits firms along various dimensions, into cells, and for each cell collects aggregate yearly data. In addition, these statistics are collected both in level and growth. The dimensions used for the split are the following, which can be specified at different levels of aggregation (see the Appendix for the detailed definition).

1. Sector:
  - STAN A7 aggregation level;
  - STAN A38 aggregation level.
2. Firm-level productivity distribution: splitting firms into productivity quantiles (with productivity defined in various ways, as reported in Section 2.3);
3. Sales distribution: splitting firms into quantiles based on gross output.
4. Size class: splitting firms into groups based on employment levels.
5. Age of the firm: splitting firms into groups based on age.
6. Ownership: independent firm vs. affiliate of a business group, and nationality of the group.
7. Demographics: entrants, exitors, incumbents, etc.

The output is provided also combining the previous dimension together (e.g. the interaction of age and size classes). However note that, although possible, the code never combines more than three dimensions at the same time; the reason is that the number of firms in each cell would become so small to incur into confidentiality problems, especially at high levels of industry disaggregation.

Several statistics are collected:

1. Basic moments: mean, median, standard deviation, and number of non-missing values, for a series of variables.
2. Several measures of aggregate productivity: decomposition of both aggregate LP and/or multi factor productivity (MFP), together with allocative efficiency measures (Olley-Pakes covariance terms).
3. Measures of allocative efficiency based on Hsieh & Klenow 2009, and Petrin & Sivadasan 2013.
4. Distribution characteristics for productivity levels, productivity growth rates, output and wages. These include both non-parametric measures such as percentiles and parameters of the distributions (e.g. pareto).
5. Descriptive statistics of firm characteristics (including growth rate and wage dispersion) by quantiles of the productivity distribution in levels and growth, and by quantiles of the sales distribution.
6. Characteristics (productivity, age, persistence, size, ownership, investments etc.) of firms at the productivity frontier.

7. Employment dynamics by quantiles of the productivity distribution.
8. Estimated parameters from distributed regressions at the firm level.

Finally, the program produces basic tabulations of firms with negative value added and graphs of the sectors' productivity distributions. Additional details on the output produced by the program are contained in the Appendix.

The output produced by the program covers a wide range of topics:

- **Productivity heterogeneity.**
In order to examine the effects of policies and macro shocks on productivity heterogeneity we collect information about the distribution of productivity (using different measures for it, as described before); we aggregate productivity and its variance to the sector level; we decompose productivity dispersion into within- versus between-sector and within- versus between- quantile dispersion.
- **Allocative efficiency.**
The access to firm-level data across multiple countries allows us to conduct an extensive examination of allocative efficiency over time and across countries, applying a number of different methods, among them: the Olley and Pakes (1996) static productivity decomposition, as well as a dynamic version of it (Melitz and Polanec, 2015); measurement of job reallocation; measurement of productivity dispersion in the top and bottom size quantile of firms in each sector and comparison to the productivity dispersion in the whole sector; measurement of misallocation according to the Hsieh and Klenow (2009) procedure with some refinements; a description of the distribution of firm-level distortions in each input and their overall impact on aggregate productivity; analysis of the gap between the value of the marginal product of an input and its marginal cost as in Petrin and Sivadasan (2013); run of distributed firm-level regressions of these measures of misallocation on firm characteristics such as size, age, ownership etc.
- **Granularity.**
The MultiProd project analyses also how much of a country's economic activity is driven by a small number of important firms, and how much of the observed productivity variation is indeed the result of microeconomic variations, collecting statistics at different quantiles of the firm size distribution. The program collects a number of indicators, such as: the market share and the share of employment accounted for by the top decile of firms in terms of gross output; the sectoral level Hirsch-Herfindahl Index (HHI); the decomposition of aggregate productivity in both level and growth between the contribution of the largest firms and that of the other firms.
- **Wage dispersion and its link to productivity.**
Data collected in the MultiProd project are instrumental to understand the evolution of the between-firm component of wage dispersion. In particular the program decomposes the wage dispersion in the within and between contribution both for industry and productivity quantiles; calculates the share of each industry and productivity quantile in the overall within component of wage dispersion; identifies the impact of various policies (e.g. employee protection legislation) on dispersion through distributed regressions.

The next sections will explore in detail each of these blocks, and will provide some examples of the output that can be produced with the MultiProd dataset.

3. Productivity heterogeneity

3.1. Motivation

A startling fact of firm level productivity analysis is the large and persistent differences in both labour productivity and multifactor productivity between firms, even within narrowly defined sectors. The MultiProd project offers a detailed understanding of productivity dispersion by investigating the relationship between microeconomic dispersion and economy-wide dispersion, in order to provide a better illustration of productivity variation within countries. In particular, to document productivity heterogeneity, MultiProd collects information about the distribution of productivity, aggregates productivity and its variance at the 2-digit sector and at the macro-sectoral level, and decomposes productivity dispersion into within- versus between-sectors and within- versus between- productivity quantiles.

3.2. Productivity Dispersion and its Evolution over Time

In order to capture heterogeneity in the data, MultiProd calculates several measures of dispersion for productivity within macro-sectors and 2-digit industries: the standard deviation; the 90-10, 90-50, and 50-10 ratios; as well as the interquartile range (i.e., the difference between the 75th and the 25th percentile).

In particular, the 90-10 productivity ratio is defined as the ratio of the 90th percentile to the 10th percentile of the productivity distribution. It is used widely in the productivity literature to assess the spread of the productivity distribution. The measure is quite intuitive since a 90-10 ratio of X can be interpreted as firms at the top of the productivity distribution, proxied by firms at the 90th percentile, producing X times as much as firms at the bottom of the distribution, proxied by firms at the 10th percentile, given the same amount of inputs.

As an example of the output produced by the code, the next table illustrates the 90-10 ratio for both (log) LP and MFP in 2011.

Table 6. Productivity 90-10 ratio in 2011, by country

COUNTRY Year 2011	Log(LP_VA) 90-10 ratio		Log(MFP) 90-10 ratio	
	Manufacturing	Services	Manufacturing	Services
Australia	1.87	2.05	1.90	2.12
Austria	1.96	2.42		
Belgium	1.60	1.74	1.80	1.78
Chile	3.00	3.53	3.07	3.87
Denmark	1.46	1.96	1.32	1.80
Finland	1.17	1.38	1.19	1.34
France	1.35	1.81	1.40	1.78
Hungary	2.79	3.29	2.54	2.86
Indonesia	3.11	-	3.41	-
Italy	1.66	2.01	1.60	1.88
Japan	1.26	1.38	1.17	1.38
Netherlands	2.00	2.98	2.27	2.27
New Zealand	1.84	2.09	1.92	2.08
Norway	1.73	2.17	1.87	2.08
Portugal	1.88	2.65	1.88	2.75
Sweden	1.45	1.86	1.59	2.45

The table illustrates some important features. First, there is a rather significant productivity dispersion in both manufacturing and services between the top performing and the bottom performing firms, and both in terms of LP and MFP. Second, dispersion is on average higher in services than in manufacturing, whether in terms of LP or MFP. Third, the ratio is particularly high in Chile, Indonesia and Hungary. Finally, in 2011, on average across countries, firms in the top decile of the distribution can produce more than six times as much value added per worker as firms in the bottom decile of the same country's manufacturing sector, and nine times in services. Essentially the same proportion is kept when looking at the 90-10 ratio in terms of MFP: with the same amount of measured inputs, firms at the top of the distribution produce almost seven times the output of firms at the bottom in the manufacturing sector, and almost nine times in the services sector. The large dispersion reflects the heterogeneity of the sample, with developing countries such as Indonesia displaying a very high dispersion.³¹

Correlated to the previous example, with the MultiProd dataset we are also able to explore several firm characteristics and productivity averages by sector, age, size and age-size classes, ownership, and also by quantiles of productivity, productivity growth, and sales distributions.

Figure 1 and Figure 2 plot the productivity (LogLP_VA and LogMFP_W, respectively) of the 10th, the 50th and 90th percentile of the productivity distribution, normalising the year 2001 to 0.

Figure 1. Trends for top, median and bottom decile of the (log) LP distribution

Note: Log Labour productivity in the 10th, 50th and 90th percentile of the productivity distribution, for manufacturing (left panel) and services (right panel) since 2001. The countries included are AUS, AUT, BEL, CHL, DNK, FIN, FRA, HUN, ITA, JPN, NLD, NOR, NZL SWE. The graphs can be interpreted as the cumulated growth rates of LP within each country and sector over the period. For instance, in 2012 in manufacturing in the 90th quantile of productivity is roughly 19% higher than in 2001. The estimates reported in the graph are those of year dummies in a cross-country regression of log-productivity in the 90th, 50th and 10th percentile of the distribution.

Figure 2. Trends for top, median and bottom decile of (log) MFP distribution

Note: Log-MFP (Wooldridge) in the 10th, 50th and 90th percentile of the productivity distribution, for manufacturing (left panel) and services (right panel) since 2000. The countries included are AUS, AUT, BEL, CHL, DNK, FIN, FRA, HUN, ITA, JPN, NLD, NOR, NZL. The graphs can be interpreted as the cumulated growth rates of MFP within each country and sector over the period. For instance, in 2012 in manufacturing in the 90th quantile of productivity is roughly 24% higher than in 2001. The estimates reported in the graph are those of year dummies in a cross-country regression of log-productivity in the 90th, 50th and 10th percentile of the distribution.

In each figure, the left panel represents productivity dispersion in manufacturing and the right panel represents productivity dispersion in (non-financial) market services. The figures show that there has been an increase in dispersion of productivity over time, especially in the services sector. The negative effect that the Great Recession had on the productivity trends is also evident from the figures. In the manufacturing sector, productivity has grown on average about 1.5% per year between 2001 and 2008 in all quantiles of the distribution. In the non-financial market services sector, instead, productivity has been rather stable up to the crisis in the top half of the distribution, whereas has decreased significantly for the bottom half of the distribution up to 2006. In both manufacturing and non-financial market services the crisis has brought a severe drop in performance for all firms, with those at the bottom of the productivity distribution being particularly affected. For all, but firms at the bottom of the productivity distribution in services, the recovery has come quickly and productivity has returned to positive growth immediately after the crisis.

3.3. Productivity Dispersion Decomposition

3.3.1. Decomposition across Sectors

To better understand the origin of the productivity dispersion, it is possible to decompose the total productivity variance V_t into two components: a within-industry component V_{Ft} and a cross-industry component V_{Xt} . The within-industry component V_{Ft} captures how much a firm's individual productivity differs from the sector (labour-weighted) average. The cross-industry component V_{Xt} captures instead how much sectors vary from each other.

$$V_t = V_{Ft} + V_{Xt} \quad (1)$$

where the within-industry variance V_{Ft} is the average over all sectors j of the square deviation of firms' productivity P_{it} to their sector (weighted) average labour productivity $\bar{P}_{jt}$:

$$V_{Ft} \equiv \frac{1}{L_t} \sum_j \sum_{i \in j} L_{it} (P_{it} - \bar{P}_{jt})^2 = \sum_j \frac{L_{jt}}{L_t} \sum_{i \in j} \frac{L_{it}}{L_{jt}} (P_{it} - \bar{P}_{jt})^2 = \sum_j \frac{L_{jt}}{L_t} \delta_{jt}^2, \quad (2)$$

and the cross-industry component V_{Xt} is the average of the squared deviation of sector j 's average productivity $\bar{P}_{jt}$ to the economy-wide productivity $\bar{P}_t$:

$$V_{Xt} \equiv \sum_j \frac{L_{jt}}{L_t} (\bar{P}_{jt} - \bar{P}_t)^2, \quad (3)$$

where L_{jt}/L_t is the labour share of sector j at time t , and $\delta_{jt}^2 \equiv \sum_{i \in j} \frac{L_{it}}{L_{jt}} (P_{it} - \bar{P}_{jt})^2$ is the labour-weighted industry variance of firm-level labour productivity.³²

The MultiProd project contributes to the literature by offering a detailed decomposition of overall productivity dispersion based on cross-country microeconomic data. This decomposition can help understand how much of the dispersion in productivity comes from microeconomic dispersion within narrowly defined sectors, and how much comes from more aggregate shocks that affect whole sectors. This is achieved by looking at the ratio V_{Ft}/V_t which reflects the importance of microeconomic shocks to aggregate dispersion. The decomposition suggested here is a cross-sectional decomposition of productivity dispersion in a given period t . In the next table we present the ratio V_{Ft}/V_t for LP in 2011.

Table 7. Share of within-sector of variance of LP, 2011

Country Year 2011	% LP Dispersion	
	Manufacturing	Services
Australia	98%	99%
Austria	86%	90%
Belgium	76%	88%
Chile	90%	97%
Denmark	84%	63%
Finland	65%	76%
France	63%	85%
Hungary	79%	99%
Indonesia	79%	-
Italy	82%	65%
Japan	75%	89%
Netherlands	80%	71%
Norway	83%	73%
Portugal	62%	76%
Sweden	53%	74%

The two columns report the share of total LP dispersion accounted for by within-sector dispersion, for manufacturing and services respectively. The results show that on average within-sector dispersion accounts for more than 77% (82%) of the overall LP dispersion observed across firm in manufacturing (services): a large share of dispersion comes from heterogeneity in LP between firms within the same two-digit sector. In other words, a substantial part of productivity heterogeneity does not come from the type of activity that firms engage in, per se, but rather by more intrinsic differences among firms within the same sector of activity in the same country.

In addition to the above decomposition, the overall within industry component can then be further decomposed in the contribution from each industry, to precisely pin down which industry drives productivity dispersion, i.e., which are the industries where dispersion is stronger. Similarly to Carvalho and Gabaix (2013), who investigate the importance of granularity – or microeconomic shocks – in driving macroeconomic fluctuations, one can decompose the within-industry component of the productivity variance into the weighted productivity variance of industries, as shown in the last term of Equation 2.^{33,34}

As an example, we report in the next table the top three contributors to LP variance in 2011 for each country, and for manufacturing and services respectively. Some sectors, such as “Food products, beverage and tobacco”, “Machinery and equipment”, “Wholesale and retail trade, repair of motor vehicles and motorcycle”, and “Transport and storage” regularly appear amongst the sectors characterised by the highest productivity dispersion.³⁵ This suggests that there might be sectoral features of the within sector distribution of firms that might affect the distribution of productivity.

Table 8. Top three sectors in the share of within-sector LP VA variance

Country (2011)	Manufacturing	% Var. Share	Services	% Var. Share
AUS	Food products, beverages and tobacco [CA]	35%	Transportation and storage [H]	26%
	Machinery and equipment n.e.c. [CK]	28%	Legal and accounting activities, etc. [MA]	25%
	Basic pharmaceutical products and pharmaceutical preparations [CF]	13%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	21%
AUT	Machinery and equipment n.e.c. [CK]	17%	REAL ESTATE ACTIVITIES [L]	34%
	Basic metals and fabricated metal products, except machinery and equipment [CH]	15%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	25%
	Wood and paper products, and printing [CC]	12%	Administrative and support service activities [N]	16%
BEL	Chemicals and chemical products [CE]	42%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	35%
	Food products, beverages and tobacco [CA]	30%	Legal and accounting activities, etc. [MA]	22%
	Wood and paper products, and printing [CC]	7%	Transportation and storage [H]	22%
CHL	Wood and paper products, and printing [CC]	54%	REAL ESTATE ACTIVITIES [L]	39%
	Food products, beverages and tobacco [CA]	22%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	36%
	Coke and refined petroleum products [CD]	11%	Transportation and storage [H]	10%
DNK	Machinery and equipment n.e.c. [CK]	28%	Transportation and storage [H]	25%
	Furniture; other manufacturing; repair and installation of machinery and equipment [CM]	24%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	22%
	Food products, beverages and tobacco [CA]	15%	REAL ESTATE ACTIVITIES [L]	13%
FIN	Computer, electronic and optical products [CI]	28%	REAL ESTATE ACTIVITIES [L]	64%
	Wood and paper products, and printing [CC]	17%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	18%
	Basic pharmaceutical products and pharmaceutical preparations [CF]	12%	Telecommunications [JB]	5%
FRA	Basic pharmaceutical products and pharmaceutical preparations [CF]	17%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	22%

Country (2011)	Manufacturing	% Var. Share	Services	% Var. Share
	Computer, electronic and optical products [CI]	17%	Legal and accounting activities, etc. [MA]	22%
	Food products, beverages and tobacco [CA]	12%	Transportation and storage [H]	11%
HUN	Transport equipment [CL]	32%	Administrative and support service activities [N]	36%
	Computer, electronic and optical products [CI]	25%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	23%
	Machinery and equipment n.e.c. [CK]	23%	REAL ESTATE ACTIVITIES [L]	15%
IDN	Transport equipment [CL]	34%		
	Chemicals and chemical products [CE]	18%		
	Food products, beverages and tobacco [CA]	17%		
ITA	Machinery and equipment n.e.c. [CK]	15%	Telecommunications [JB]	23%
	Transport equipment [CL]	13%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	22%
	Basic metals and fabricated metal products, except machinery and equipment [CH]	12%	Transportation and storage [H]	14%
JPN	Coke and refined petroleum products [CD]	25%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	37%
	Chemicals and chemical products [CE]	17%	Legal and accounting activities, etc. [MA]	34%
	Machinery and equipment n.e.c. [CK]	11%	Administrative and support service activities [N]	11%
NLD	Chemicals and chemical products [CE]	34%	Administrative and support service activities [N]	35%
	Food products, beverages and tobacco [CA]	10%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	28%
	Coke and refined petroleum products [CD]	10%	Telecommunications [JB]	12%
NOR	Machinery and equipment n.e.c. [CK]	28%	Telecommunications [JB]	31%
	Electrical equipment [CJ]	14%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	22%
	Transport equipment [CL]	11%	Transportation and storage [H]	16%
PRT	Coke and refined petroleum products [CD]	28%	REAL ESTATE ACTIVITIES [L]	21%
	Food products, beverages and tobacco [CA]	20%	Telecommunications [JB]	21%
	Wood and paper products, and printing [CC]	17%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	18%
SWE	Computer, electronic and optical products [CI]	35%	REAL ESTATE ACTIVITIES [L]	39%
	Wood and paper products, and printing [CC]	14%	Telecommunications [JB]	16%
	Transport equipment [CL]	12%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	15%

Using Equations 1, 2 and 3, the overall variance can be expressed as follows:

$$V_t = \sum_j \frac{L_{jt}}{L_t} \left[\delta_{jt}^2 + (\bar{P}_{jt} - \bar{P}_t)^2 \right] \quad (4)$$

Equation 4 highlights that MultiProd can help assess whether changes over time of aggregate productivity dispersion V can be explained by changes over time in:

- sectors' own dispersions δ_{jt}^2 ,
- a reallocation towards more or less volatile sectors of the economy, which translates into changes in sectors' relative weights in the economy L_{jt}/L_t ,

- the spread between sectors captured by $(P_{jt} - \bar{P}_t)^2$.

3.3.2. Decomposition across Productivity Quantiles

Decomposing productivity dispersion into its various components helps identify factors driving productivity volatility and their evolution over time. The decomposition of Section 3.3.1. (variance across sectors) can also be used to analyse the contribution of different types of firms to overall dispersion. In particular, it is possible to replicate the same decomposition technique but across different quantiles of the productivity distribution, to understand whether productivity dispersion, and its change over time, comes from firms at the top or at the bottom of the distribution.

Equations 1-3 show how the decomposition can be replicated for within and across quantiles of the firm productivity distribution. The variance decomposition of within-quantiles and across-quantiles components, respectively V_{Ft}^Q and V_{Xt}^Q , is given at time t by

$$V_t = V_{Ft}^Q + V_{Xt}^Q. \quad (5)$$

The within-quantiles total volatility at time t is:

$$V_{Ft}^Q \equiv \frac{1}{L_t} \sum_q \sum_{i \in q} L_{iq} (P_{iq} - \bar{P}_{qt})^2 = \sum_q \frac{L_{qt}}{L_t} \sum_{i \in q} \frac{L_{iq}}{L_{qt}} (P_{iq} - \bar{P}_{qt})^2 = \sum_q \frac{L_{qt}}{L_t} \delta_{qt}^2, \quad (6)$$

and the cross-quantiles volatility at time t is:

$$V_{Xt}^Q \equiv \sum_q \frac{L_{qt}}{L_t} (\bar{P}_{qt} - \bar{P}_t)^2, \quad (7)$$

where P_{iq} is the productivity of firm i in quantile q , $\bar{P}_{qt}$ is the weighted average of productivity of firms in quantile q , and δ_{qt}^2 is the weighted quantile variance. The quantiles can be computed for the productivity distribution of the whole economy or individual industries; in the baseline case the analysis has been carried out at the A7 industry level (j subscripts are omitted for simplicity).

This decomposition can help policy-makers better understand the source of productivity dispersion, and in particular help understand whether some parts of the productivity distribution are inherently more dispersed, and how much they contribute to aggregate dispersion.

3.3.3. Entry and Exit

A potential contributor to increases in productivity dispersion across firms could be firm entries and exits. For example, if recent policies within a country enhanced access to credit, a great variety of entrants with greater levels of productivity heterogeneity may be found in contrast to earlier entrants. The greater amount of productivity dispersion amongst entrants could result in greater disparity in earnings across firms and increase the overall earnings dispersion.

An analysis that could complement what was described above would be to look at the evolution of the 90th-50th and 50th-10th productivity differentials of entrants and incumbents, as in Faggio et al. (2010). If productivity dispersion of entrants has increased significantly while little or no change is found amongst stayers, this may provide basic evidence that productivity dispersion is being driven by new firms.

A complementary examination to the above could be to simply examine productivity dispersion (excluding exiting firms and entrants) of a balanced panel of stayers during the sample period.

3.4. Analysing MFP Time-Volatility

As already mentioned, the within-sector component of the variance decomposition highlighted in Section 3.3.1 is close to the concept of fundamental volatility outlined in the follow-up contribution by Carvalho and Gabaix (2013). They define fundamental volatility as the sum of sector-specific time variances σ_j^2 , weighted by sectors' weights in the economy (here using Domar weights).

$$\sigma_{F_t}^2 = \sum_{j=1}^n \left(\frac{S_{jt}}{GDP_t} \right)^2 \sigma_j^2 . \quad (8)$$

Their measure of fundamental volatility is therefore designed to reflect the weighted variance of sectors in the economy, where the variance of each sector is a technical characteristic of the sector. Their objective is to estimate what proportion of GDP time variance can be explained by sector time variances, as reflected in fundamental volatility.

Our variance decomposition highlighted in Section 3.3.1 is instead cross-sectional in nature. However, MultiProd can also improve on their analysis of productivity volatility over time. Due to data limitations, both Gabaix (2011) and Carvalho and Gabaix (2013) have to rely on approximate estimates of the variance over time for firms and sectors, respectively. The MultiProd project can do away with this limitation by being able to estimate sector variances using representative data encompassing many countries (each j would correspond to a particular combination of the A7 industry level and country). In order to obtain an accurate measure of the productivity volatility over time, the MultiProd output could be easily extended to include also the productivity variance of single firms over time, aggregated to a non-disclosive sectoral level (although this would restrict the data to incumbents and possibly lead to survival bias).

4. Allocative efficiency

4.1. Motivation

As sketched in the introduction, a growing body of literature attributes high aggregate MFP not only to the efficiency of technology use and speed of adoption, but also to the efficient allocation of resources across firms. Resource allocation impacts aggregate productivity: a flow of inputs from low- to high-productivity firms is productivity enhancing, whereas factors assigned in an inefficient manner adversely affect aggregate productivity. These important problems have been investigated by two intrinsically interrelated branches of the literature. First, the reallocation literature, which typically focuses on the drivers of resource reallocation, such as creative destruction, entry and exit, increased market share of fast growing firms and so on. Second, the misallocation literature, which typically identifies a specific distortion or a bundle of distortions (policies and/or institutions) and examines the extent to which they adversely impact aggregate productivity.

Regarding the reallocation literature, empirical research finds a constant churning with continuous resource reallocation between heterogeneous firms within sectors (e.g.: Olley and Pakes 1996; Foster et al. 2002; Bartelsman et al. 2004, 2009). The reallocation is not only driven by the exit of shrinking firms and the entry of new firms, but also by the constant upscaling and downscaling of incumbents, a churning that can significantly contribute to aggregate productivity growth. The extent of the contribution depends on how effectively resources are reallocated across firms and sectors. Bartelsman et al. (2004), for example, find that allocative efficiency varies considerably across countries, and Arnold et al. (2011) find substantial differences across sectors and countries. In particular, allocative efficiency explains more than 50% of productivity growth in the US while only representing 4% in Slovenia (Bartelsman et al. 2013). Substantial heterogeneity in resource allocation across countries is also found when a dynamic decomposition method is adopted, as in Bartelsman et al. (2009), who find high dispersion in the estimates that range from positive to negative contributions.³⁶ These results show that countries differ significantly in their ability to allocate resources efficiently, with important consequences for productivity growth.

Regarding the misallocation literature, the empirical evidence generally shows that distortions in the economy can have a quantitatively important effect on aggregate productivity. Indeed, the extent to which resources are misallocated within a country may help explain its poor aggregate productivity performance. Given MultiProd's access to firm-level data across a number of countries, this project is well positioned to conduct an extensive examination of the effects of misallocation over time and across countries.

One of the ways in which the effects of misallocation can be analysed is through a *direct* approach. This branch of the literature starts by choosing an element which is considered a driver of misallocation, obtains a direct measure of this factor, then employs a model of heterogeneous production units to quantify the extent to which these factors impact productivity.³⁷ One of the first studies to contribute to this strand of the literature, Hopenhayn and Rogerson (1993), suggests that misallocation of labour across firms caused by employee dismissal regulations can result in a reduction of MFP by around 5%. Leal (2010) examines another factor of distortion, incomplete tax enforcement, and examines the extent to which this impacts aggregate output and productivity in Mexico. When enforcement is in full force, output and LP are 19% greater with perfect competition and 34% greater with monopolistic competition. In addition, the author suggests that incomplete tax enforcement distorts the economy via three mechanisms: factors are misallocated to low productive entities, low productive firms can more easily enter the market, and capital-labour ratios of the informal sector are reduced. Size contingent policies are also found to lower aggregate output and output per worker by 8.1% and 25.6%, lower average establishment size by 20%, and increase the amount of firms by 23.5% (Guner et al, 2008). In addition,

Buera et al. (2011) finds that financial distortions can largely explain the disparities in output per worker, aggregate MFP, industry level productivity and capital intensity across countries.

Given the difficulty of estimating the difference in importance between factors which are easily measured vs those that are not easily measured, a second method known as the *indirect* approach analyses the effects of the entire bundle of misallocation distortions on MFP. The sources of misallocation are modelled as wedges in the first order conditions of the firm optimisation problem (i.e. on the marginal choices of firms). Here the literature attempts to answer two questions: what determines the magnitude of the effect of idiosyncratic distortions on aggregate productivity and how substantial is the disparity in these wedges (Restuccia and Rogerson, 2013)? To answer the first issue, Restuccia and Rogerson (2008) use a model of heterogeneous establishments and analyse a situation where misallocation from a generic distortion adversely affects aggregate MFP. Their results suggest that output taxes which allocate resources from highly productive to low productive firms can reduce output and estimated MFP by around 30 to 50%.

Hsieh and Klenow (2009) address the second question by identifying and measuring the size of misallocation. Employing micro data on manufacturing entities, they calculate the probable amount of misallocation in China and India while using the US as a benchmark. When hypothetically adjusting the marginal product of capital and labour of China and India to reflect the marginal product of capital and labour from the US, estimated MFP gains are between 30% to 50% and 40% to 60% for China and India, respectively. Moreover, Ziebarth (2012) proposes that resource misallocation was a significant contributor to the observed drop in productivity during the Great Depression. This paper decomposes the shift in industry level productivity into misallocation and changes in efficient productivity similarly to Hsieh and Klenow (2009) to examine the ice and cement industries. The authors find that for the cement industry, within-sector misallocation accounted for at least 50% of the decline in productivity from 1929-1931, 20% from 1931-1933 and 10 to 15% from 1933-1935. In addition, the estimates suggest that within-sector misallocation explains a drop of around 50% in productivity of the ice manufacturing industry from 1929-1935.

Both the reallocation and misallocation literature assume that when factors are assigned in an inefficient manner, aggregate productivity is adversely affected. The two strands of the literature approach this issue from two complementary perspectives and apply different methodological tools.

The extent to which reallocation has occurred is generally analysed with various measures such as job creation and destruction, the OP covariance term and changes in productivity dispersion. In addition, factors which may influence the drivers of reallocation such as technological change, regulations and recessions are explored. The misallocation literature, instead, usually builds on models of heterogeneous production units from Hopenhayn (1992), Hopenhayn and Rogerson (1993), Lucas (1978) or Melitz (2003). The results obtained in this literature are informative regarding the quantitative effects of misallocation on aggregate productivity.

Understanding the drivers of inefficiencies in resource allocation is of primary importance to provide sound policy advice aimed at enhancing countries' aggregate productivity performance. Thanks to its unique collection and use of cross-country harmonised micro data, the MultiProd project enhances our understanding on the contribution of resource allocation for a greater number of countries across a greater number of years and for a wider spectrum of policy issues than any previous micro distributed data project on productivity. One key determinant that can explain cross-country differences in resource allocation is regulation. Some policies which have been found to influence reallocation include entry regulation (Daveri et al. 2016), employment production legislation (Martin and Scarpetta 2012), liberalisation and competition (Melitz and Ottaviano 2008) and so on. Alternatively, differences in reallocation may also be driven by an exogenous shock such as a recession, which may either enhance allocative efficiency through a cleansing

effect (Davis et al. 2012) or scar the economy, thereby hindering an efficient allocation of resources (Barlevy 2002 and Ouyang 2009). MultiProd is in a position to contribute to the literature by providing a micro based harmonised cross-country analysis on the effects of the policy environment and macro shocks on resource allocation and productivity over the cycle.

4.2. Measuring Reallocation

In order to examine the effects of policies and macro shocks on reallocation and productivity, a number of methods have been followed in MultiProd.

4.2.1. Olley and Pakes (1996) Static Productivity Decomposition

The Olley and Pakes (OP) decomposition method allows for the estimation of a static measure of allocative efficiency. It is possible to decompose sectoral productivity into the sum of an un-weighted average of firm-level productivities and a covariance term which represents reallocation (also called the OP gap). The latter is a measure of allocative efficiency, since it increases if more productive firms capture a higher share of resources in the sector:

$$P_t = \frac{1}{N_t} \sum_i P_{i,t} + \sum_i (\theta_{i,t} - \bar{\theta}_t)(P_{i,t} - \bar{P}_t). \quad (9)$$

P_t is sectoral level productivity at time t , N represents the number of firms in a sector, $\theta_{i,t}$ is the share of firm i at time t , $\bar{P}_t$ and $\bar{\theta}_t$ are sectoral averages.

It is important to notice that the left-hand side of Equation (9) represents the correct measure of aggregate (sectoral) productivity only if the appropriate weights are used. In the case of value added labour productivity (LP_VA), the weights are simply labour shares, and this is the main measure used in MultiProd. Standard aggregate productivity is also not retrieved when LP is measured in log terms as this would correspond to a geometric average; nevertheless this measure is also adopted given its widespread use in the empirical literature. In the case of MFP, the decomposition is more complicated and the standard output (GO or VA) weights used in the literature are not the correct ones. They have the effect of mechanically inflating the covariance term, as pointed out by Van Biesebroeck (2008). The correct weights are input weights, which are substantially more complicated because they depend on the way MFP is calculated. The MultiProd program implements this type of weights for MFP levels. This also allows for the calculation of a modified OP gap with the correct input weights in the case of MFP, even though there is a caveat in the interpretation since the weights do not sum up to one.

As an example, in the next figure we show the 2011 OP covariance term and the unweighted productivity in the manufacturing sector for the countries included in the MultiProd dataset. The interpretation is the following: the larger OP gap (i.e., the covariance between size and productivity of firms), the higher the overall sectoral productivity with respect to its unweighted mean. Roughly, the figure shows that overall in 2011 Nordic European countries exhibit on average higher OP gap and aggregate productivity as well.

Figure 3. OP covariance and unweighted productivity in 2011, manufacturing sector

Note: Following equation (9), the picture represents for each country the aggregate productivity, decomposed into the OP covariance term and the unweighted productivity. Data on the OP gap are not available for NZL.

Furthermore, we report for some selected countries (namely, BEL, DNK, FRA, SWE) the trend of the rescaled manufacturing OP gap over time. The rescaled OP gap increases at least until 2006 in all countries, then in three (BEL, FRA, SWE) out of four countries displays a decrease in 2007-2009 (essentially during the crisis), and then starts increasing again from 2009 onwards.

Figure 4. OP covariance trend in BEL, DNK, FRA and SWE, manufacturing sector

4.2.2. Dynamic Productivity Decomposition Methods

In addition to static productivity decompositions, dynamic decomposition methods have been widely employed in the literature, in a similar fashion to what was first introduced in Bailey et al. (1992) and augmented by Foster et al. (2001). In this method, industry-level (or aggregate) growth in productivity is decomposed into five elements: *within firm component*, representing within firm growth in productivity; the *between firm component*, representing the part of industrial productivity explained by top performing firms expanding their market shares or by less productive shrinking firms; the *cross effect component*, which accounts for increases in productivity of expanding firms, or drops in productivity of shrinking firms; and the *entry/exit components* that represent the contribution of the productivity of new and exiting firms relative to average productivity in the previous period. The decomposition is defined as follows:

$$\Delta p_t = \sum_{i \in C} \theta_{it-1} \Delta p_{it} + \sum_{i \in C} \Delta \theta_{it} (p_{it-1} - p_{t-1}) + \sum_{i \in C} \Delta \theta_{it} \Delta p_{it} + \sum_{i \in E} \theta_{it} (p_{it} - p_{t-1}) - \sum_{i \in X} \theta_{it-1} (p_{it-1} - p_{t-1}) \quad (10)$$

Where Δ represents the growth over the period; C , E and X refer to incumbents, entrants and exitors; and $\theta_{i,t}$ is the share of firm i at time t . Productivity is measured in logs, and p_{it} is the productivity of firm i at time t while p_{t-1} represents the weighted average productivity level for an industry (or the economy) in the previous time period.

This and analogous methods have proved to be very informative on the specific elements that drive productivity growth, unveiling the important role played by firm heterogeneity and the entry and exit dimensions. It has been widely used in the literature and has improved our understanding of the drivers of productivity growth both within and across countries. However these methods do not come without shortcomings. For instance, as stressed by Melitz and Polanec (2015), standard decomposition methods introduce some biases in the measurement of the contributions of entry and exit due to the wrong choice of the reference productivity level. The MultiProd project therefore adopts a dynamic Olley-Pakes methodology as proposed by the previous authors, which allows a more precise computation of the reallocation terms and is obtained as follows:

$$\Delta p_t = \frac{1}{N^C} \sum_{i \in C} (p_{it} - p_{it-1}) + \Delta \text{cov}_{i \in C}(\theta_{it}, p_{it}) + \left(\sum_{i \in E} \theta_{it} \right) \cdot (p_t^E - p_t^C) + \left(\sum_{i \in X} \theta_{it-1} \right) \cdot (p_{t-1}^C - p_{t-1}^X) \quad (11)$$

where the first term is the change in the unweighted productivity average of incumbents; the second term is the change in the OP covariance term computed for incumbents; and p_t^E , p_t^C , p_{t-1}^X are the weighted productivity averages of respectively entrants, incumbents and exitors computed in the relevant time period and with weights that sum up to one within each group.

Moreover the left hand side of Equation (11) represents standard aggregate productivity growth only in the case of LP with labour weights, analysed in an un-logged specification. This is in general not true in the widely used case of log MFP with output weights. The reallocation terms can also give rise to counterintuitive results, for instance in the extreme case of an ideal frictionless world, where resources instantaneously flow across firms, these terms can still be positive. This is because the change in shares that represent reallocation (for instance in the between component) are weighted by the level of productivity of each firm, so average productivity can increase if one allocates more resources to more productive firms, even when firms equate their input marginal revenue products and marginal costs (the optimal solution in a frictionless world).

Petrin and Levinsohn (2012) have recently tried to address some of these shortcomings and proposed an alternative decomposition that aims at properly accounting for aggregate productivity growth.

As in Hulten (1978), they use Domar weights as the correct weights for MFP measured with gross output (GO), and construct a decomposition method that can be applied to any group of firms. Similar to Jorgenson et al. (2007), this includes a reallocation term that can be split into a component for each primary input of production (for instance labour and capital), which is a useful tool to investigate the importance of policies that specifically affect labour and/or capital markets. Moreover, their setting is more in line with the misallocation literature that will be discussed below. In fact the reallocation terms contribute to aggregate productivity growth only in presence of distortions like markups, taxes, adjustment costs, etc. This approach is particularly suited to study the impact of policy on reallocation and hence on aggregate productivity; it is therefore implemented in the MultiProd project together with a distributed regression analysis to investigate the effect of firm level observables on factor reallocation.³⁸

4.2.3. Job Reallocation

An alternative way to measure reallocation is to look more narrowly at job reallocation. Similar to Davis et al. (1996) and Foster et al. (2014), job creation (JC) and destruction (JD) can be calculated as follows:

$$JC_{qt} = \sum_{f \in Q^+} \left(\frac{X_{fqt}}{X_{qt}} \right) g_{fqt}, \quad (12)$$

$$JD_{qt} = \sum_{f \in Q^-} \left(\frac{X_{fqt}}{X_{qt}} \right) |g_{fqt}|, \quad (13)$$

with $X_{fqt} = \frac{1}{2}(L_{fqt} + L_{fqt-1})$ and $g_{fqt} = \left(\frac{\Delta L_{fqt}}{X_{fqt}} \right)$. Q^+ represents the group of expanding firms

and Q^- refers to firms that are shrinking or exiting. g_{fqt} is the employment growth rate for firm f in group q at time t , and L_{fqt} is the number of workers employed at firm f .

Total work reallocation is the sum of both JC and JD, while Net employment growth (NET) is job creation minus job destruction. In the context of the MultiProd project it is interesting to look at job reallocation by firms belonging to different segments of the productivity (growth) distribution. This has been achieved by calculating the overall net employment growth for groups of firms corresponding to different quantiles of the distribution:

$$NET_{qt} = \frac{1}{X_{qt+1}} \sum_{i=1}^n \Delta L_{fqt+1}, \quad (14)$$

where $X_{qt+1} = \frac{1}{2}(L_{qt+1} + L_{qt})$, and q in this case refers to the group of firms belonging to the

same quantile of the distribution of productivity growth over the two periods. Therefore NET_{qt} is defined with a forward-looking perspective, looking at net employment growth between year t and $t+1$, and it inherently refers to incumbents only, as firms have to be observed in both periods. To avoid problems of sample attrition or in the case of censuses which occur every k years, the statistic can also be computed over a longer time horizon annualising growth rates.

An alternative approach is to construct quantiles based on the productivity level distribution in a single period. In this case, one ranks firms in year t and then takes again a forward-looking view and calculates how much each quantile (of the productivity level distribution) contributes to net employment growth between year t and $t+1$. The problem in this case is how to treat entry and exit. The latter is actually easy to account for because firms that are not present in year $t+1$ will contribute negatively to employment growth by their level of employment in the last year of appearance. On the other hand, entry is more difficult to treat since the level of productivity of the firms is unknown before entry (at time t), and hence those firms cannot be assigned to quantiles. When analysing top and bottom quantiles (e.g. deciles) this might not be a big concern, as entrants tend to feature productivity levels that are just slightly lower than incumbents so they are not very likely to enter at either end of the productivity distribution (Foster et al. 2001 and 2014). Moreover this represents temporary sample attrition, as entrants will appear in the following year. Still, one might want to interpret the figures in any specific year with some care, as in the short run entrants might differentially impact quantiles at the top and at the bottom of the productivity distribution. Two measures have been adopted in the analysis, the first is obtained exactly as in Equation (14), with the only differences that the quantiles are defined on the productivity level at time t , L_{fqt} will include both incumbents and exitors, and L_{fqt+1} will also include exitors simply noting that their employment level at $t+1$ is zero (hence for exitors: $\Delta L_{fqt+1} = -L_{fqt}$). The second measure is very similar but it simply calculates a standard growth rate as follows:

$$NET_{qt} = \frac{1}{L_{qt}} \sum_{i=1}^n \Delta L_{fqt}. \quad (15)$$

In the context of MultiProd a greater source of concern is the fact that production surveys are generally a sample of firms so entry and exit is not precisely observed. For this reason, all of the previous statistics are computed using labour variables coming from the business registers (or similar sources). The sample is still limited to the firms observed in the production surveys (which provide the measures of LP and MFP used for defining the quantiles), but entry and exit are defined using the correct demographic information from the business registers.

4.3. Measuring Misallocation

4.3.1. The Contribution of the Largest Firms to Productivity

As a first step to better picture the extent of misallocation using firm-level data, MultiProd identifies the contribution of the largest firms to aggregate productivity. The rationale is that when frictions are low, resources should flow to the most productive firms in the economy. This would imply that the largest firms would indeed be the most productive. Understanding how productivity of the largest firms compares to that of other firms in their sector can help identify the extent of allocative frictions.

In order to achieve this, MultiProd carries out a cross-sectional analysis of aggregate/sectoral productivity and decomposes it between the contributions of these big players versus the contribution of the rest of firms. In a sector j at period t , the productivity contribution of the largest firms is given by:

$$\bar{P}_{jt}^{S_j^{10}} \equiv \frac{1}{L_{jt}^{S_j^{10}}} \sum_{i \in S_j^{10}} L_{ijt} P_{ijt}, \quad (16)$$

where L_{ijt} is the employment of firm i in sector j at time t and P_{ijt} is its labour productivity, $S_j^{\overline{10}}$ designates the top decile in sector j of firms ranked by sales, and $L_{jt}^{S_j^{\overline{10}}}$ is the total employment of firms in the top decile of the employment size distribution. The analysis can be extended to MFP but the choice of weights to get to a correct measure of aggregate productivity is less straightforward (see Van Biesebroeck 2008). To understand how the largest firms differ from the rest of the sector in terms of productivity, this can be compared to the weighted average productivity in sector j :

$$\overline{P}_{jt} = \frac{1}{L_{jt}} \sum_i L_{ijt} P_{ijt} \quad (17)$$

The ratio $\overline{P}_{jt}^{S_j^{\overline{10}}} / \overline{P}_{jt}$ captures the extent to which productivity in sector j is dominated by the largest firms. In an economy that is characterised by low allocative frictions, resources should flow to the most productive firms. In such an economy, there is a positive correlation between firm size and firm productivity: the average productivity of the largest firms should be higher compared to an economy where reallocation is hampered by frictions.

4.3.2. Productivity Dispersion among Firms of the Top and Bottom Size Quantiles

A second way to unveil the presence of misallocation is to generate descriptive statistics of the dispersion of productivity among the largest firms in a sector. Economic theory predicts that in the absence of any frictions and distortions in the economy, resources should flow to firms until marginal returns are equalised. In such a case, firms of identical size should be equally productive. But in the presence of frictions, which generate misallocation of production factors, the dispersion of measured productivity for firms in the same size class will increase. For this reason the extent of misallocation is positively related to the dispersion of productivity in the top and bottom size quantiles. A high dispersion of productivity in the top size quantile indicates the presence of unproductive firms among the larger firms in a sector. Conversely, a high level of productivity dispersion in the bottom size quantile indicates that productive firms fail to attract resources.

MultiProd provides this potential measure of misallocation by computing productivity dispersion in the top and bottom size quantile of firms in each sector, and comparing it to the productivity dispersion in the whole sector:

$$\frac{\text{prod. dispersion in size quantile } S}{\text{prod. dispersion in whole sector}} = \frac{\frac{1}{L_{jt}^S} \sum_{i \in S} L_{ijt} (P_{ijt} - \overline{P}_{jt}^S)^2}{\frac{1}{L_{jt}} \sum_i L_{ijt} (P_{ijt} - \overline{P}_{jt})^2}, \quad (18)$$

where S is either the top or bottom size quantile in sector j , $\overline{P}_{jt}^S$ is the employment-weighted average labour productivity of firms in the size quantile S of sector j , and $\overline{P}_{jt}$ is the weighted average productivity of sector j . Again, the same calculation can be extended to MFP dispersion, with an appropriate choice of weights. A higher ratio implies that productivity dispersion among the largest firms is high compared to the rest of the sector, indicating the existence of distortions in the economy.

4.3.3. MFP Dispersion and Misallocation

After examining the correlation between size and productivity using a comparison between size classes as above, the next step is to quantify more precisely the extent of the misallocation, and depict its evolution over time and across countries. In order to do that, MultiProd follows the methodology of Hsieh and Klenow (2009, henceforth HK), which makes an explicit connection between misallocation and MFP dispersion.

Dispersion in Revenue and Physical Productivity

Hsieh and Klenow (2009) assume that firms are subjected to idiosyncratic distortions, or wedges, that move firms' choice away from the optimal outcome and therefore generate inefficiencies.³⁹ Their main contribution is to show that wedges induce dispersion of productivity, so that MFP dispersion can be used to quantify the extent of misallocation in an economy.

In a standard framework without wedges, all firms' revenue productivity, called TFPR by Foster et al. (2008), should be equalised in equilibrium for firms of the same sector, even if they have different TFPQ (the standard physical productivity).⁴⁰ Since firms face different wedges from each other, idiosyncratic distortions faced by firms generate a dispersion of marginal choices among firms and therefore generate a dispersion of the resulting TFPR. More precisely, the bigger the wedges, the bigger the dispersion in TFPR, and hence the dispersion captures the extent of the misallocation. Indeed, in the HK framework, aggregate TFP in sector j is given by:

$$TFP_j = \left[\sum_{i \in j} \left(TFPQ_{ij} \cdot \frac{\overline{TFPR}_j}{TFPR_{ij}} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}, \quad (19)$$

where $TFPR_{ij}$ denotes firm i 's revenue productivity and $\overline{TFPR}_j$ is sector j average TFPR. Equation (19) shows how TFPR dispersion affects aggregate sectoral TFP. In the absence of wedges, each firm's $TFPR_{ij}$ is equal to the sector average $\overline{TFPR}_j$ and aggregate TFP in sector j is the ideal (undistorted)

$TFP_j = \left[\sum_{i \in j} TFPQ_i^{\sigma-1} \right]^{\frac{1}{\sigma-1}}$, which is simply a CES function of firm-level productivities. Instead, the existence of wedges creates a divergence of $TFPR_{ij}$ from the sector average $\overline{TFPR}_j$, which in turns *lowers* aggregate TFP in sector j .⁴¹

To understand the severity of misallocation, the first exercise is to graph the distribution of firms TFPR, which can be retrieved directly from the MultiProd data. More precisely, following HK, it is possible to plot for each sector j , the distribution of $\log(TFPR_{ij}/\overline{TFPR}_j)$. MultiProd replicates this exercise on a much wider range of countries. In Figures 5 and 6 we report as example the Danish case both in manufacturing and services.

To get a more precise picture, Tables 9 and 10 display three statistics capturing dispersion of TFPR: standard deviation, the ratio of TFPR between the 75th percentile and the 25th percentile, and the ratio between the 90th and the 10th percentiles of the TFPR distribution. Higher numbers correspond to greater dispersion. Two main things can be noticed: first, dispersion has slightly increased over time and reaches a peak during the Crisis; second, services are more dispersed than manufacturing.

In particular, since the framework implies that TFPR dispersion comes from wedges, higher TFPR dispersion (around the mean) indicates greater misallocation. From this, it can be concluded that the thicker tails of the 2012 TFPR distribution with respect to 2000 represents an increase of misallocation in Denmark.

Figure 5. Distribution of TFPR, manuf. in DNK

Figure 6. Distribution of TFPR, services in DNK

Table 9. Dispersion of TFPR, manuf. in DNK

Year	SD	75-25	90-10
2000	0.417	0.417	0.981
2001	0.420	0.437	0.978
2002	0.404	0.444	0.929
2003	0.400	0.449	0.939
2004	0.373	0.432	0.908
2005	0.402	0.446	0.942
2006	0.401	0.446	0.955
2007	0.403	0.459	0.968
2008	0.400	0.468	0.950
2009	0.449	0.492	1.051
2010	0.438	0.471	1.048
2011	0.432	0.471	1.013
2012	0.412	0.460	0.973

Table 10. Dispersion of TFPR, services in DNK

Year	SD	75-25	90-10
2000	0.628	0.592	1.503
2001	0.601	0.798	1.463
2002	0.548	0.658	1.401
2003	0.527	0.624	1.296
2004	0.515	0.573	1.200
2005	0.524	0.634	1.235
2006	0.542	0.606	1.284
2007	0.615	0.803	1.508
2008	0.541	0.651	1.319
2009	0.600	0.638	1.336
2010	0.547	0.615	1.314
2011	0.551	0.620	1.321
2012	0.555	0.629	1.324

Using the identifying assumptions in HK, it is possible to perform the same exercise on TFPQ, the ideal physical measure of productivity, by plotting the ratio $\log\left(\frac{TFPQ_{ij}}{\overline{TFPQ}_j}\right)$ for firm i in sector j . This yields Figures 7 and 8 (and Tables 11 and 12) for Denmark. Note that, in line with the empirical findings of Foster et al. (2008), TFPQ is more dispersed than TFPR, and has fatter left tails. The intuition is

that more productive firms charge lower prices, that is, there is a negative correlation between physical productivity and prices. This effect dampens differences in TFPQ, so that the dispersion in TFPQ is lower. But their exercise shows that in the presence of wedges, this dampening is imperfect, and the remaining dispersion in TFPQ can be used to quantify the amount of allocative distortions in the economy.

Figure 7. Distribution of TFPQ, manuf. in DNK

Figure 8. Distribution of TFPQ, services in DNK

Table 11. Dispersion of TFPQ, manuf. in DNK

Year	SD	75-25	90-10
2000	0.803	1.106	2.071
2001	0.825	1.086	2.153
2002	0.832	1.096	2.141
2003	0.824	1.122	2.151
2004	0.767	1.056	1.961
2005	0.816	1.083	2.037
2006	0.802	1.098	2.048
2007	0.814	1.085	2.093
2008	0.816	1.089	2.096
2009	0.882	1.160	2.190
2010	0.877	1.162	2.196
2011	0.884	1.192	2.208
2012	0.857	1.154	2.159

Table 12. Dispersion of TFPQ, services in DNK

Year	SD	75-25	90-10
2000	0.833	1.058	2.064
2001	0.892	1.228	2.302
2002	0.837	1.102	2.120
2003	0.811	1.057	2.031
2004	0.797	1.019	2.035
2005	0.829	1.090	2.093
2006	0.847	1.067	2.108
2007	0.983	1.378	2.481
2008	0.871	1.172	2.171
2009	0.948	1.129	2.186
2010	0.873	1.108	2.174
2011	0.859	1.118	2.151
2012	0.886	1.138	2.215

Quantifying Potential Gains from Reallocation

The key insight of the Hsieh and Klenow (2009) method is to show that TFP dispersion can be used as a measure of misallocation, and therefore as a measure of how far an economy is from its efficient equilibrium. Their setting yields a closed form specification of aggregate TFP that depends directly on the dispersion of the distortions. Using sectors aggregate TFP given by Equation (19), it is possible to compute the ratio of actual output to its efficient level (the level that would occur if there were no distortions).

The major difficulty in quantifying misallocation from these wedges is that neither TFPQ nor the idiosyncratic wedges are directly observable, since the data only contain information on firms' revenues and do not contain information on firms' prices from which output could be retrieved. However, HK show that under some assumptions these can be retrieved from the data. In particular they show that a firm's TFPQ can be reconstructed and is proportional to its revenue productivity:

$$TPFQ_i = \kappa_j \frac{(P_i Y_i)^{\frac{\sigma}{\sigma-1}}}{K_i^{\alpha_j} L_i^{1-\alpha_j}}, \quad (20)$$

where κ_j is a sector specific constant that depends on the magnitude of wedges faced by firms in that sector. It cannot be directly observed but washes away when comparing relative productivities.

Since aggregate TFP is decreasing in the amount of TFP dispersion in each sector, the ratio of actual output to its efficient level indicates the extent of misallocation in terms of lost TFP. It is then easy to compute potential TFP gains that could be achieved by allocating resources efficiently within each sector:⁴²

$$\text{Gains}_{jt} = \frac{Y_{jt}^{\text{efficient}}}{Y_{jt}} - 1 \quad (21)$$

While the exercise yields useful insights on the total magnitude of distortions in the economy, the estimates from the HK approach are high because the effect is confounded with any problem of model misspecification or measurement error. A more realistic comparison can be done by comparing these gains to a certain benchmark. Hsieh and Klenow use the gains in the United States in 1997 as a benchmark. Since data for the US are unavailable in MultiProd, we use compute gains relative to a specific year in each country, i.e., the gains that would be achieved by keeping the allocation of resources as in the baseline year in each country. We compute relative TFP gains as follows:

$$\text{Relative TFP gains} = \frac{\text{Gains}_t}{\text{Gains}_{\text{base\_year}}} - 1 \quad (22)$$

The comparison to a benchmark reduces the scope for misspecification because any model misspecification would also affect gains computed in the base year. As an example, in the next figures we plot for the abovementioned countries (Belgium, Denmark, France and Sweden) the TFP gains as in Equation (21), as well as the relative TFP gains as in Equation (22), taking as base year the 2003, both for manufacturing and services. Misallocation has increased over time in all countries (except in Denmark for services); moreover, it is higher in services than in manufacturing. Although in absolute terms misallocation seems to be higher in France and Belgium, Sweden has seen a higher increase over time in services. In the manufacturing sector, on the other hand, the four countries exhibit rather similar trends.

Figure 9. TFP gains

Figure 10. Relative TFP gains, base year 2003

The MultiProd project contributes to the literature by expanding the HK analysis to more than 18 OECD and non-OECD countries, and by testing the robustness of the results to some of the key assumption in the model (e.g. constant return to scale, same elasticity of substitution across industries etc.). Therefore, it can analyse differences in relative TFP gains between countries, and their evolution over time, which can capture geographical and time variations in misallocation that can be linked to policy.

The MultiProd program also allows to examine the TFP gains that are foregone because of the distortions in each input, capital and labour (as in HK 2009 appendix), and hence helps understanding what type of distortions policies should target. Re-writing the model in terms of distortions to the absolute levels of capital and labour, it is possible to compute aggregate TFP in sector j in the case of an efficient allocation of labour:

$$TFP_j^L = \left[\sum_{i \in j} \left(TFPQ_{ij} \cdot \frac{\overline{MPRK}_j}{MRPK_{ij}} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}, \quad (23)$$

where $MRPK_{ij}$ is the marginal revenue product of capital for firm i in sector j . The efficiency gains obtained from eliminating the labour distortions alone, $Gains_j^L$, are simply obtained by comparing TFP_j^L with TFP_j . The efficiency gains in the case of an efficient allocation of capital can be computed in a similar way.

In a model with separate distortions for capital and labour, it is possible to analyse the gap between the value of the marginal product of an input and its marginal cost, analogously to Petrin and

Sivadasan (2013). For instance in the case of labour, the gap for firm i can be defined as: $G_{ijt}^L = |MRPL_{ijt} - w_{ijt}|$, where w_{ijt} is the average wage paid by firm i and $MRPL_{ijt}$ is computed in MultiProd using both the industry-specific labour share from the US, as in Hsieh and Klenow (2009), and the labour elasticity coming from the Wooldridge (2009) production function estimation (as in Petrin and Sivadasan, 2013). The gap captures inefficiencies coming from a misallocation of labour both below and above the optimal level, and it represents the productivity gain from adjusting labour by one unit in the optimal direction, holding everything else constant. Moreover the output gains obtained by firm i by eliminating the entire distortion and moving to the socially optimal labour level L^* , can be approximated with the following triangular loss:

$$T_{ijt}^L = \frac{1}{2} \cdot G_{ijt}^L \cdot |L_{ijt} - L_{ijt}^*| = \frac{1}{2} \cdot G_{ijt}^L \cdot \left| 1 - \frac{(1 + \tau_{ijt}^K)^{\alpha_j(\sigma-1)}}{(1 + \tau_{ijt}^L)^{\alpha_j(\sigma-1)-\sigma}} \right| \cdot L_{ijt} \quad (24)$$

On the one hand this quantity is appealing because it directly represents an output loss and therefore has a direct economic interpretation; on the other hand, its calculation depends heavily on underlying assumptions and therefore needs to be analysed with caution. One example of such assumption is that the output loss due to an inefficient allocation of labour depends on the wedge on capital, $1 + \tau_{ijt}^K$, which in turn depends on the level of the rental price of capital R . This rental price is unobserved and needs to be assumed, and thus the output gap computed depends directly on this key assumption. In any case this quantity has been only employed in distributed regressions at the firm level and it is therefore possible to partially control for these issues (for instance if firms within the same sector face similar rental prices of capital).

Explaining Differences in Misallocation

The HK method does not require specifying what mechanism generates the distortions in the economy, hence its wide applicability. However, from a policy perspective, it is important to better understand what generates the distortions. This can be done both at the firm level through distributed regressions and at the industry level in the ex-post analysis of the collected micro-aggregates. This exercise can be particularly useful in disentangling how much of the distortions are due to idiosyncratic firm characteristics, and how much are due to more structural sector-level factors. In this section we analyse the role of firm characteristics, while we postpone the discussion on the role of policies to Section 4.4.3.

The labour and capital gaps vary at the firm level, and the *MultiProd* program runs distributed regressions to investigate whether firm level observables influence labour and capital inefficiencies, or at least to establish robust conditional correlations. In particular it is possible to investigate whether inefficiencies are more or less severe with the age of the firm, the size class, the ownership of the firm (whether it belongs to a domestic or foreign group), or a demographic event (whether a firm is an entrant or a future exitor). The code runs the following distributed regressions:

$$y_{ijt} = \alpha + \beta_1 \cdot \text{age}_{ijt} + \beta_2 \cdot \text{size\_class}_{ijt} + \beta_3 \cdot \text{ownership}_{ijt} + \beta_4 \cdot \text{demographics}_{ijt} + \lambda_{jt} + \varepsilon_{jct}, \quad (25)$$

where y_{ijt} can be either the labour inefficiency G_{ijt}^L , the capital inefficiency G_{ijt}^K or total efficiency loss due to labour T_{ijt}^L (see Equation 24); and where λ_{jt} are industry-year fixed effects (at the fourth digit). The code first runs univariate regressions with one explanatory factor at a time (and industry-year fixed effects) to look at the effect of each variable separately, and then multivariate regressions with all explanatory variables at the same time, to tease out their relative magnitude. The obtained coefficients are

country specific since the regressions are run separately for each country, but an ex-post meta-analysis allows for the computation of an average effect across countries.

An important part of this analysis is to disentangle how much of the firm-level inefficiencies are idiosyncratic and due to firm characteristics (age, size, ownership structure), and how much are instead more structural inefficiencies largely captured by the year-industry fixed effects. The relative magnitude of these two categories can help policy makers target economic policies either towards certain classes of firms (e.g. young firms) or more structural policies that could target certain sectors.

4.4. Policies, Allocative Efficiency and Productivity

Government regulation across countries is heterogeneous with regards to the type of policies used, the efficiency of their implementation and their effects on reallocation and economic growth. There is a wide range of policies that affect reallocation, for instance credit constraints alleviation measures (Aghion et al., 2007; Klapper et al., 2006), labour market legislation (Haltiwanger et al., 2006; Micco and Pages, 2006), product market regulations and trade liberalisation (Arnold et al., 2008; Bernard et al., 2003, 2011; Melitz, 2003; Melitz and Ottaviano, 2008), bankruptcy (Selowsky and Martin, 1997; Lim and Hahn, 2004), R&D tax credit (Acemoglu et al. 2013). Some policies can create a market environment that increases efficiency of reallocation and positively impact aggregate productivity. At the same time, ineffective regulations can raise the barriers to entry, increase the cost of investment, and lower competitive pressure, thereby reducing reallocation and hindering economic growth and welfare. Given the heterogeneous impact of policies on allocative efficiency, it is therefore important to distinguish the policies that have a positive impact on growth from those that actually hinder the economy. In order to do so it is necessary to examine heterogeneity in policies across countries, identify the link between policies and reallocation, and estimate the contribution of reallocation to aggregate productivity.

4.4.1. Heterogeneity in Policies and Reallocation across countries

In order to analyse the effects of various policies on reallocation and productivity, it is important to start by examining the use and stringency of regulations across countries. Martin and Scarpetta (2012), for example, do this in the case of Employment Protection Legislation (EPL) by plotting the differences in severity of EPLs across countries. This allows one to identify countries with the strictest to the most lenient regulation. Given the number of countries and years encompassed in MultiProd, the project can provide a deeper understanding of the productivity impact of heterogeneous policy implementation across countries, and at all points of the cycle and the growth distribution. The nature of the data is such that, particular attention can be devoted to regulations that affect reallocation but do not necessarily work through the entry and exit channel. Despite receiving somewhat less attention in the reallocation literature, there are plenty of examples of such regulations. Essentially they include any policy that affect the upscaling and downscaling of firms, such as: size contingent policies, policies that affect investment or capital reallocation (e.g. improvements in the functioning of judicial systems and financial markets), policies that reward success (e.g. intellectual property rights and patent systems), manage failures (e.g. bankruptcy) or that re-qualify the resources that are left behind etc. This would constitute an important starting point in determining the effects of policies on reallocation and productivity growth across countries.

Given the variation that exists both in regulation and in the contribution of reallocation to aggregate productivity growth, the following section will investigate the extent to which the two are related. In particular, it will outline whether differences in regulations can explain the amount of heterogeneity in reallocation across industries, countries, and over time. Reallocation can be measured using the methodologies outlined in the previous section, which enable one to analyse changes in reallocation over time. Differences in reallocation can be influenced by technology and market environment. At the same time, however, regulations can impact the extent to which resources are

effectively reallocated within a country. Therefore, understanding the extent to which policies differ across industries and countries may help to identify the link between policy implementation and reallocation.

4.4.2. Link between Policy and Reallocation

Country's framework conditions and the policy environment in which firms operate play an important role in affecting firms' production decisions, in particular the decision to invest in fixed capital, hire and fire workers or engage in risky innovation activities.

In particular, policies that enact costs on the firms' choice of employment levels, e.g. via hiring and firing, can inadvertently impact job creation and destruction, hindering labour reallocation and acting as a drag to upscaling and downscaling (Mortensen and Pissarides, 1994). In addition, Autor et al. (2007) find a negative correlation between EPL severity and entrant rates and job flows across US states. Similar to Arnold et al. (2008), a first attempt at analysing the effects of regulation severity on resource allocation can be implemented by measuring the correlation between the OP covariance term and policy indicators, as follows:

$$y_{jct} = \lambda_j + \lambda_c + \beta * Policy_{jct} + \varepsilon_{jct}, \quad (26)$$

where y_{jct} is the measure of allocative efficiency for industry j in country c at time t , $Policy_{jct}$ is an indicator variable representing industry-level regulation severity, and λ_j and λ_c are sector and country fixed effects. The measure of allocative efficiency does not have to be limited to the OP covariance term but can be anyone of the measures outlined in Section 4.2. Causality can be addressed by adopting a Rajan and Zingales' (1998) approach, since industries might have a differential exposure to the various regulations.

Furthermore, it is likely that differences in the severity of policies across countries could have a differential impact on job creation (JC) and destruction (JD) for firms at the top and at the bottom of the productivity distribution. One could expect that in a country with little regulation on workers' dismissals, the dispersion of net employment growth (*NET*) between firms at the top and those at the bottom would be greater. On the one hand, since the least productive firms tend to shrink faster, the absence of firing regulation may result in greater job destruction (JD) for these firms, given the limited costs associated to shedding jobs. Moreover, given that these firms are more likely to shrink and exit, less regulation may have a more limited effect on incentivising job creation (JC) amongst these worst performing firms. On the other hand the most productive performing firms (i.e. in the top quantile) are more likely to expand faster than firms at the bottom, displaying higher JC and a lower JD. Therefore, less regulation may potentially allow firms at the top to further absorb workers, knowing that they can shed labour at a lower cost in the event of a macro shock. Moreover, JD may increase a little but possibly not substantially since firms are more productive and less likely to shrink and fire workers.

In countries with strict employee dismissal regulations, the differential in *NET* between firms at the top and firms at the bottom may potentially be smaller. Firms at the bottom decile will face greater firing costs and will therefore be less willing to create jobs. At the same time, firms at the top may also be less eager to create jobs, given their inability to reduce labour costs during an economic downturn. Although all firms will be prevented from firing workers, the impact of the policy on reducing job destruction by top performing firms may be limited given that these firms are more likely to expand.

Within the MultiProd project, the impact of policy strictness on the dispersion of *NET*, and thus reallocation can be estimated in the same manner as in Equation (26). However, the reallocation measure y_{qct} is taken from Equations (14) or (15), where the quantiles can be the top and bottom deciles.

4.4.3. The Role of Policies in Explaining Misallocation

As mentioned above, the HK method does not require specifying what mechanism generates the distortions in the economy, hence its wide applicability. However, from a policy perspective, it is important to be able to direct efforts towards particular aspects of the economic framework that generate these inefficiencies. This requires understanding better what generates distortions. Here we focus on the role of structural sector-level factors, and policies in particular, in the context of an ex-post analysis of the collected micro-aggregates.

Busso et al. (2013) suggest that policies can be a major driver of misallocation and policy variation across countries could explain some of the differences in misallocation observed in the data. In particular, they suggest three channels that could generate misallocation: uneven access to financial markets (which prevents productive firms to access capital), imperfect tax collection (which amounts to a subsidy to non-compliant low-productivity firms), and poor enforcement of labour and social security regulations (which incentivises firms to remain small and free-ride on government provided schemes). They suggest that these channels could affect smaller and larger firms differently, and that the magnitude of the effect may differ across countries. Country by country OLS regressions of returns to factors, $\log(TFPR_{ij}/\overline{TFPR_j})$, on size class dummies can reveal in which country misallocation tends to favour small firms, and in which it favours the bigger ones.

Moreover once misallocation is properly quantified, it becomes possible to look into a more precise analysis of the impact of certain policies on misallocation. A similar approach to the analysis of reallocation described in the previous section can be performed here. The extent to which certain policies result in misallocation can be estimated using regression analysis, by estimating the following model:

$$y_{jct} = \lambda_j + \lambda_c + \beta \cdot \text{Policy}_{jct} + \varepsilon_{jct} , \quad (27)$$

where y_{jct} can be either Gains_{jct} , Gains_{jct}^L , or $\text{Gains}_{jct}^{K^*}$, which represent, respectively, potential TFP gains in sector j of country c in the absence of any distortions, in the absence of labour distortions, and in the absence of capital distortions. Moreover λ_j and λ_c are sector- and country fixed effects respectively, and Policy_{jct} captures industry-level regulation severity. Alternatively the dependent variable can be obtained by regressing G_{ijt}^L , G_{ijt}^K or T_{ijt}^L on industry-year fixed effects (λ_{jt}), to retrieve industry averages over time of the labour gap, the capital gap, or the total output loss due to an inefficient allocation of labour. The time-varying industry averages can then be regressed on policy variables to assess their impact on sector-level distortions. One interesting application of this would be to measure not only the direct impact of labour policies (such as employment protection legislation) on labour distortion measures, but also their indirect impact on capital distortions.

4.5. The Great Recession, Reallocation and Productivity Growth

The Great Recession adversely impacted the economies of many countries around the world. From 2007 to 2009, for example average unemployment in the OECD increased by 2.7%, the worst being Spain, which experienced an increase in unemployment of 9.7% (Pissarides, 2013). In addition, Ball (2014) finds that the recession resulted in an average loss of potential output of 8.4% for the 23 OECD countries in his sample, and more than a 30% loss of potential output for Greece, Hungary and Ireland. Furthermore, many EU countries have had persistent output contractions, with the most severely impacted, Greece, with output decreasing even after 6 years post-recession (OECD, 2014). The extent of the economic damage caused by the recession seems to suggest that recessions have a “scarring” effect on an

economy, worsening market distortions and harmfully effecting healthy firms (Barlevy 2002 and Ouyang 2009).

Some empirical and theoretical research however suggest that recessions may be a time for cleansing the economy of unproductive firms, freeing up resources so that they can be more efficiently employed elsewhere. This cleansing hypothesis builds off of the discussion of creative destruction introduced by Schumpeter (1939) and later illustrated by numerous macro models.⁴³ Empirical work by Davis and Haltiwanger (1992 and 1999) finds evidence of increased reallocation in the manufacturing sector during recessions which occurred between the late 1940s and the 1990s. Furthermore, research by Davis, Faberman and Haltiwanger (2006, 2012) finds evidence of cleansing for both service and manufacturing industries.

From a policy perspective it is therefore important to distinguish the impact of the Great Recession. For example, if the recent recession acted as a cleansing mechanism, when designing policies aimed at mitigating the immediate impact of the crisis, policy makers need to take into account the potential detrimental effects for long-term productivity growth. More specifically policies should be targeted at shielding workers from the shocks without undermining the reallocation of resources in the economy (and the exit of the unproductive firms), which is beneficial in the long run. Conversely, if the recession scarred the economy, it is important for policy makers to deal with the short-term effects of the shock so as to bolster the market for future growth (Paci et al. 2012). This could justify shielding firms from economic shocks - e.g. support them to overcome the crisis - in the short term, and not only workers, because the crisis could drive out of the market firms with important skills and capabilities, which are hard to re-build. MultiProd can also help examine the extent to which the Great Recession facilitated or exacerbated reallocation and productivity across countries so as to provide the basis for sound policy advice.

4.5.1. Changes in the MFP Distribution Overtime

If the recent recession had a cleansing effect on the market and encouraged reallocation, one would expect high levels of churning in the economy with firms shrinking at the bottom. Data from the second wave of the DynEmp project seem to provide evidence that goes in this direction. Countries where high growing firms expanded the most (in terms of employment) are also the ones that experienced the greatest amount of shrinking at the bottom of the employment growth distribution. The MultiProd project expands this evidence by including the firm productivity dimension in the picture.

As starting point for this analysis, in Figure 11 we plot the difference between the 90th and 10th percentiles of log-productivity (LP and MFP, respectively) over time, normalised at 0 in 2001; the left panel represents log-productivity dispersion in manufacturing and the right panel represents log-productivity dispersion in (non-financial) market services. The figure shows that productivity dispersion has increased over time in both sectors. Moreover, the dispersion flattens out during the crisis years but increases again as of 2010 (2011 in manufacturing). Therefore, if there has been a cleansing effect, it did not last long: from 2010 dispersion has indeed started to rise again, especially in the services sector. This result is in line with what seen in the context of the misallocation literature; in Figure 9 and 10 we have shown that, for the countries considered, misallocation decreases or flattens out during the crisis years contributing to aggregate productivity growth, but then it starts increasing again. Moreover, this figure relates to the evidence already presented in Figure 1 and 2, in which it was shown that productivity dispersion has increased since 2001, especially in the services sector.

Figure 11. Divergence of (log) productivity over time, manufacturing vs. non-financial market services

Note: The figure plots the estimated year dummies of a regression of log-productivity dispersion (90th and 10th percentiles ratio) within country-sector pairs, using data from the following countries: AUT, BEL, DNK, FIN, FRA, HUN, ITA, JPN, NLD, NOR, NZL, PRT, SWE. The graphs can be interpreted as the cumulated growth rates of productivity within each country and sector over the period. For instance, in 2012 productivity dispersion in manufacturing is roughly 9% higher than in 2001.

In addition, in order to identify potential drivers of dispersion, the MultiProd data allow for productivity dispersion to be further broken down by various firm characteristics, such as old vs young firms; multinational vs non-multinational; foreign vs domestic. A similar analysis can be employed when analysing the effects of the recession on wage inequality, as will be outlined in more detail in Section 6.

4.5.2. Job Creation, Job Destruction, and the Role of Productivity during the Crisis

Another way of analysing whether the recession had a cleansing effect is to examine the extent to which jobs were created and destroyed during the recession. Past macro shocks deemed cleansing resulted in a sharp increase in job destruction with a mild decrease in job creation. For example, during the 2001 recession, the US manufacturing sector experienced a dramatic increase in job destruction, with little or mild decreases in job creation, highlighting the reallocation of resources from low performers to those at the top.

One way to expand and complement the previous evidence is to include the productivity dimension into the picture. MFP and reallocation can be linked by examining differences in net employment growth for firms in the top decile and firms in the bottom decile of the productivity distribution (see Equations 14 and 15), similar to Foster et al (2014). In the event of a macro shock, firms in the top decile should be more resilient and thus experience limited change in job creation and job destruction. Firms at the bottom however are more likely to experience a sharper decrease in job creation and increase in job destruction since they are less productive and more likely to shrink and possibly exit. Increases in the differences in net employment growth amongst these two groups suggest a potential cleansing effect, where resources being shed by firms at the bottom are being reallocated to more productive firms at the top. In addition, one could assume that as the severity of the recession increases, so does the disparity in net employment between low and high productive firms. We can therefore plot the values of net job creation for those at the top and those at the bottom over time, in order to examine the link between productivity and reallocation before, during and after the recession.

Thanks to the data generated by MultiProd, it is possible to further investigate the cleansing effect of the recession using regression analysis. In the spirit of the analysis carried out by Foster et al., it is possible to run the following regression at the country-industry level:

$$\ln\left(\frac{NET_{Top,jqt}}{NET_{Bottom,jqt}}\right) = \lambda_j + \lambda_t + \nu * Cycle_{ct} + \mu * GR_t * Cycle_{ct} + \varepsilon_{jct} , \quad (28)$$

where *Cycle* is the country-year change in the unemployment rate from t to $t+1$, *GR* is a dummy for the Great recession being 1 in the 2007-2009 period and 0 otherwise. As in Foster et al. (2014), *NET* can be calculated considering either the overall growth in employment for both incumbents and firms that exit or only the conditional growth of incumbents (Equation 14). In the previous specification the dependent variable is defined as the ratio of *NET* for firms in the top decile versus firms in the bottom decile of the productivity distribution (in growth or level). An alternative specification using $NET_{Top,jqt}$ as the dependent variable could be run to investigate the role of job creation, which takes place mostly at the top. Finally, the results from the regression analysis could be graphed and compared across countries.

5. Granularity

5.1. Motivation

The traditional view of productivity variation posits that individual firms are subject to uncorrelated idiosyncratic shocks that cancel out if averaged over a large number of firms (akin to a law of large numbers). This view therefore leaves to economy-wide shocks the role of explaining the variations in aggregate productivity. However, identified economy shocks (inflation, wars, changes in policies) are only able to explain a small fraction of the observed fluctuations (Cochrane, 1994), leaving much of the variations to be attributed to unspecified aggregate shocks.

A contrario, Gabaix (2011) proposed what he called the “granular” hypothesis, namely that aggregate fluctuations are instead the results of microeconomic shocks.⁴⁴ Such idiosyncratic shocks, even if they are uncorrelated, may not cancel out on average if sectors are dominated by a small number of large firms. When the distribution of firm size is fat-tailed, the usual law of large numbers argument (i.e., uncorrelated shocks should cancel out on average) breaks down. Microeconomic shocks to large firms may then be sufficient to generate ample fluctuations in the economy. This fact can have important implications for policies aimed at increasing economic resilience, and highlights the importance of studying firm-level data to better understand aggregate outcomes.

The MultiProd project offers new insights on this hypothesis by analysing how much of a country’s economic activity is driven by a small number of important firms, and how much of the observed productivity variation is indeed the result of microeconomic variations.

5.2. Understanding characteristics of the largest firms

Any attempt at understanding the importance of granularity in the dynamics of the economy must encompass an analysis of the influence of the major players, and how much or how little they drive economic activity.

The use of micro-data offers the unique opportunity to better understand the differences between the largest players in the economy and the rest of the economy. In order to achieve this, MultiProd collects statistics by different quantiles of the firm size distribution. The particular focus of the granularity approach is to compare characteristics of the largest firms in the economy to the other firms. It is possible to perform this analysis at different disaggregation levels. One can start from the whole economy and compare the largest players to the rest of firms in the economy. At a finer level, it is possible to perform the same comparison within macro-sectors of activity (A7 classification), again identifying in each macro-sector the largest firms and comparing them to smaller firms in their macro-sectors. Finally, the same analysis can be performed at the finer level of STAN A38 sectors.

A first step in the analysis of the granularity hypothesis consists in looking at the actual weight of the largest firms in the economy, and in general at how concentrated economic activity is. The first measure MultiProd constructs is the market share accounted for by the top decile of firms (ranked according to their sales). The higher the market share, the bigger the weight that these firms have in the economy, and the more concentrated is their economic activity compared to the rest of the firms. That indicator is computed as:

$$\varsigma_j^{10} \equiv \sum_{i \in S_j^{10}} \frac{S_{ij}}{S_j} , \quad (29)$$

where S_j^{10} designates the top decile in sector j of firms ranked by sales.⁴⁵

The second indicator examines at the share of employment that these firms represent, to understand their influence on labour markets:

$$L_j^{10} \equiv \sum_{i \in S_j^{10}} L_{ij} / L_j .$$

As an example in this direction, in the next figure we plot the share of value added and total employment by size quantiles (in terms of sales). In particular, we plot the mean over all countries and years of value added and employment shares for the bottom and the top 10th percentile, and for the 10th to 90th percentile of the gross output distribution. The figure shows that about 79% (78%) of the total value in manufacturing (services) is produced by the firms in the top decile of sales. The rest of the distribution accounts for only the 20% of the value added, with the bottom decile producing a negligible share. Employment is slightly less concentrated: the firms in the top decile of sales employ 66% (66%) of total employment in manufacturing (services). Both manufacturing and services look equally concentrated.

Figure 12. Value added and employment shares by quantiles of sales, manufacturing vs. services

Note: countries included are AUT, BEL, DNK, FIN, HUN, NOR, PRT. The period considered is 1996-2012. Weighted averages across all countries and years.

The third indicator used to assess the extent of granularity is the Herfindahl-Hirschman Index (HHI) at each level of sectoral classification, to capture the aggregate concentration of economic activity within each sector. For a sector j , the index is given by

$$h_j = \sum_i \left(\frac{S_{ij}}{S_j} \right)^2, \quad (30)$$

where S_{ij}/S_j represents the market share of firm i in sector j . The Herfindahl-Hirschman index is

frequently used in the economic literature to assess the degree of concentration of economic activity in a sector (or in the economy as a whole).⁴⁶ The closer the Herfindahl-Hirschman index is to 1, the more concentrated economic activity is.⁴⁷

As an example, we report in the next table the HHI in 2011 for each country of the sample. In most of the cases the HHI is higher in manufacturing than in services.

Table 13. Herfindahl-Hirschman index by country in 2011, manufacturing and services sector

Country (2011)	Manufacturing	Services
AUS	0.007	0.006
AUT	0.004	0.003
BEL	0.006	0.005
CHL	0.027	0.004
DNK	0.009	0.010
FIN	0.050	0.005
HUN	0.016	0.004
IDN	0.005	-
ITA	0.001	0.002
JPN	0.004	0.002
NLD	0.010	0.004
NOR	0.011	0.002
PRT	0.009	0.002

The same Herfindahl-Hirschman index can be computed using labour shares instead of market shares, to give a sense of concentration of labour in a given sector j :

$$h_j = \sum_i \left(\frac{L_{ij}}{L_j} \right)^2. \quad (31)$$

These three measures can then be compared across countries, to understand differences in the distribution of economic activity and link these differences to differences in their policies and business environments. It is also interesting to look at how the weight and role of the largest firms evolved during the Great Recession, and contrast the sectors in which the crisis yielded a more diversified economic structure against those in which economic activity became more concentrated.

A very similar analysis of “concentration” can be made when firms are ranked in terms of productivity instead of size. In particular the first two indicators are obtained also for the firms at the top of the productivity distribution (firms at the “frontier”) to investigate to which extent the two groups of firms overlap, and whether market and labour shares of the most productive firms are close to those of the largest ones.⁴⁸

5.3. The Contribution of the Largest Firms to Productivity

As a second step in the analysis, MultiProd identifies the contribution of the largest firms to aggregate productivity, and to aggregate productivity growth. The rationale is that when frictions are low, resources should flow to the most productive firms in the economy. This implies that the largest firms are indeed the most productive. Understanding how average productivity of the largest firms compares to that of the other firms in their sector helps identifying the extent of allocative frictions.

5.3.1. Productivity Level

It is possible to decompose the aggregate productivity level between the contributions of the largest firms versus the contribution of the rest of firms (at any level of sectoral aggregation). We use LP and labour shares to obtain an exact decomposition of aggregate productivity in sector j into the productivity of each quantile. For any split of firms into quantiles, aggregate LP is decomposed in the following way:

$$P_{jt} = \sum_i \frac{L_{it}}{L_{jt}} P_{it} = \sum_q \frac{L_{qjt}}{L_{jt}} \sum_{i \in q} \frac{L_{it}}{L_{qjt}} P_{it} = \sum_q \frac{L_{qjt}}{L_{jt}} \overline{P_{qjt}} , \quad (32)$$

where $\overline{P_{qjt}}$ is the weighted average LP in quantile q of sector j and $\frac{L_{qjt}}{L_{jt}}$ is the quantile q labour share in sector j . In particular, by splitting firms into the top size-decile and the rest of firms, it is possible to look at the contribution of the largest firms to aggregate productivity:

$$P_{jt} = \frac{\overline{L_{jt}^{S_{10}}}}{L_{jt}} \overline{P_{jt}^{S_{10}}} + \frac{L_{jt}^{S_{90}}}{L_{jt}} \overline{P_{jt}^{S_{90}}} , \quad (33)$$

where $\overline{P_{jt}^{S_{10}}}$ is the aggregate LP of the top size decile in sector j . This contribution captures how much the largest firms contribute to productivity.

In an economy that is characterised by low frictions in reallocation, resources should flow to the most productive firms. In such an economy, there is a positive correlation between firm size and firm productivity: the average productivity of the largest firms should be higher than when reallocation is hampered by frictions.

Is it also possible to compare the unweighted average productivity in the top quantile $\frac{1}{\#S_{10}} \sum_{i \in S_{10}} P_{it}$ (where $\#S_{10} = N/10$ is the number of firms in the top size decile) to average productivity for the whole of sector j given by $\frac{1}{N} \sum_{i \in j} P_{it}$. This identifies how different the largest firms are from the rest of the sector, and whether the most productive firms are indeed in the top size quantile. This analysis can be easily done for both LP and MFP. Also, the previous analysis in terms of weighted average productivity can be performed for MFP, but the weights to obtain the correct measure of aggregate productivity are more complex as discussed above (see Van Biesebroeck 2008).

5.3.2. Productivity Growth

In addition to providing new insights for productivity levels, the MultiProd project sheds light on the contribution of the largest firms to productivity *growth*, by decomposing productivity growth in sector j into the productivity growth of each quantile q .

$$\Delta P_{jt} \equiv \sum_{i \in j} \omega_{ijt} \Delta P_{it} = \sum_{q \in j} \sum_{i \in q} \omega_{ijt} \Delta P_{it} . \quad (34)$$

The choice of the weights depends on the type of productivity analysed. In case of LP, the weights are simply labour shares ($\omega_{ijt} \equiv L_{it} / L_{jt}$), which are also the correct weights to obtain a measure of aggregate productivity growth in the sector. In the case of MFP, the choice of the weights is more complicated and obtaining the correct measure of aggregate productivity growth is not immediate. When MFP is measured in gross output the most appropriate weights are Domar weights $\omega_{ijt} = S_{it}/Y_{jt}$. By using these weights, the previous expression can be rewritten into the weighted sum of the (weighted) average productivity growth $\overline{\Delta P_{qjt}}$ of each quantile q :

$$\sum_{i \in j} \omega_{ijt} \Delta P_{it} = \sum_{i \in j} \frac{S_{it}}{Y_{jt}} \Delta P_{it} = \sum_{q \in j} \frac{Y_{qjt}}{Y_{jt}} \sum_{i \in q} \frac{S_{it}}{Y_{qjt}} \Delta P_{it} = \sum_q \frac{Y_{qjt}}{Y_{jt}} \overline{\Delta P_{qjt}} \quad (35)$$

Note that this decomposition can be performed for any definition of quantiles, or more generally any grouping of firms. In particular, when dividing firms into those that belong to the top size quantile and the others, the previous expression becomes:

$$\sum_{i \in j} \omega_{ijt} \Delta P_{it} = \frac{Y_{jt}^{\overline{S_{10}}}}{Y_{jt}} \overline{\Delta P_{S_{10}jt}} + \frac{Y_{jt}^{S_{90}}}{Y_{jt}} \overline{\Delta P_{S_{90}jt}} \quad , \quad (36)$$

where the first term gives the contribution of firms in the top size decile to productivity growth in sector j . It is possible to compare this contribution across countries for the same sector j , and to look at its evolution over time to understand whether productivity growth has, or not, been driven by the largest firms.⁴⁹

Finally note that in the case of MFP, this is not exactly a decomposition of aggregate productivity growth, which would also include terms capturing the misallocation of input factors and intermediates (see section 3 in Petrin and Levinsohn, 2012 for more details). The two would only coincide in a perfectly competitive frictionless economy, so the exercise performed for MFP is simply an analysis of the average sectoral technical efficiency, and of which group of firms contributes the most to it.

6. Wage Dispersion and its link to productivity

6.1. Motivation

Since the 1970s wage inequality increased for a multitude of countries, in particular the United States and the United Kingdom (Atkinson and Piketty 2009, Machin and Van Reenen 2008). Much of the increases in wage inequality of individuals can be attributed to increases in the wage variance between rather than within establishments, as found by Dunne et al. (2004). In addition to this, Barth et al (2014) show that the growth in the between establishment wage dispersion contributed to around 79% of the growth in the variance of individual wages.⁵⁰ At the same time, there appears to be a positive relationship between productivity and wage dispersion. Dunne et al. (2004) for example find a positive and significant cross-plant correlation between plant-level wages and productivity. Additionally, a positive correlation exists between changes in wages and changes in productivity.

The MultiProd project contributes to this literature by shedding further light on the nature of wage inequality and the components that drive the increase in dispersion in a cross country setting. More specifically the data collected are instrumental to: first, understand the evolution of the between firm component of wage dispersion; second, decompose the wage dispersion in the within and between contribution both for industry and productivity quantiles; third, calculate the share of each industry and productivity quantile in the within component of wage dispersion; fourth, identify the impact of various policies (e.g. employee protection legislation) on dispersion, and finally see whether wage variance is driven by technological change and or capital intensity per worker.⁵¹

6.2. A cross-country description and decompositions of wage dispersion

6.2.1. Graphical illustration of the dispersion in wages

Analysing the evolution of establishment average wage variance can provide insights into potential disparities both over time and across different segments of the productivity distribution. In order to examine this, establishment-level real average wages are estimated for the entire sample period. Next, similarly to Faggio et al. (2010), the 90th, 50th and 10th decile values are computed and normalised to zero in the first year of the sample period. Wage by deciles are then graphed over time to illustrate the cumulative change in these values. This can be undertaken for the economy as a whole, for the main macro sectors, for 2-digit industries, and even further refined by several firm characteristics (e.g.: age, size, productivity segment etc...).

As an example, in the next figures we plot the average increase in the top, median and bottom decile of real wages within country-sectors over time. Wages in the 90th percentile increased over time more than those in the 10th percentile. As a consequence divergence in wages seems to have increased over time on average across all countries, and in both the manufacturing and the services sector.

Figure 13. Top, median and bottom decile of real wages over time

Note: Wages in the 10th, 50th and 90th percentile of the wage distribution, for manufacturing (left panel) and services (right panel) since 2001. The countries included are AUS, AUT, BEL, CHL, DNK, FIN, FRA, HUN, ITA, JPN, NLD, NOR, NZL, SWE. The graphs can be interpreted as the cumulated growth rates of wages within each country and sector over the period. For instance, in 2012 in manufacturing in the 90th quantile of productivity is roughly 10% higher than in 2001. The estimates reported in the graph are those of Year dummies in a cross-country regression of log-wage in the 90th, 50th and 10th percentile of the distribution.

6.2.2. Decomposition across sector and across quantiles

In addition to examining the evolution of wages, it is important to decompose their variance in order to identify where the main contribution of dispersion originates from. Using the same approach as the productivity variance decomposition in Section 3.3, the wage variance can be decomposed into two components: a within-sector component V_{Ft} (Equation (38)), and a cross-sector component V_{Xt} , the average industry wage relative to the average wage of the economy (Equation (39)). Here, the variances have been calculated using employee-weighted averages. The variance decomposition in within- and cross-sector components is obtained as follows:

$$V_t = V_{Ft} + V_{Xt} , \quad (37)$$

where the within-sector component of variance V_{Ft} is:

$$V_{Ft} \equiv \frac{1}{L} \sum_j \sum_{i \in j} L_{it} (w_{it} - \bar{w}_{jt})^2 = \sum_j \frac{L_{jt}}{L_t} \sum_{i \in j} \frac{L_{it}}{L_{jt}} (w_{it} - \bar{w}_{jt})^2 = \sum_j \frac{L_{jt}}{L_t} \delta_{jt}^2 , \quad (38)$$

and the cross-sector component of wage variance V_{Xt} is:

$$V_{Xt} \equiv \sum_j \frac{L_{jt}}{L_t} (\bar{w}_{jt} - \bar{w}_t)^2 . \quad (39)$$

W_{it} is the average wage of firm i at time t , $\overline{W}_{jt}$ is the employment-weighted sector j average of firm-level average wages, $\delta_{jt}^2 \equiv \sum_{i \in j} \frac{L_{it}}{L_{jt}} (W_{it} - \overline{W}_{jt})^2$ is the employment-weighted industry variance, and $\overline{W}_t$ is employment-weighted economy average in period t .

As another example of the output that can be produced with the MultiProd dataset, in the next table we present ratio V_{Ft}/V_t for the 2011, i.e., the share of the total wage variance that comes from within-sector wage variances.

Table 14. Share of within-sector of variance of wages, 2011

Country Year 2011	% Real Wage Dispersion	
	Manufacturing	Services
Australia	88%	75%
Austria	75%	82%
Belgium	62%	73%
Chile	74%	86%
Denmark	88%	72%
Finland	65%	75%
France	71%	76%
Hungary	70%	85%
Indonesia	89%	-
Italy	80%	83%
Japan	79%	80%
Netherlands	64%	93%
Norway	86%	83%
Portugal	67%	76%
Sweden	78%	79%

Note: Share of within-sector variance in total macro-sector wage variance. Firms and sectors weighted by the number of employees.

The two columns report the share of total wage dispersion accounted for by within-sector dispersion, for manufacturing and services respectively. The results show that on average within-sector dispersion accounts for more than 76% (80%) of the overall wage dispersion observed across firm in manufacturing (services): there is still a large amount of heterogeneity in terms of wages between firms within the same two-digit sector. In other words, a substantial part of wage heterogeneity does not come from the fact that firms in certain sectors offer average wages that are very different from those in the rest of the economy, but rather by the fact that firms within the same sector have very different average wages. The decomposition can also be replicated for within and cross-quantiles of the firm productivity distribution, that is, decomposing the share of dispersion that comes from wage dispersion within each productivity quantile, and the share that comes from wage dispersion across productivity quantiles. It can be easily obtained in that case by summing over quantiles and considering, for each quantile q , the average wage of firm i , W_{it} , together with the employment-weighted quantile average of firm-level average wages

$\overline{W}_{qt}$ and the employment-weighted quantile variance $\delta_{qt}^2 \equiv \sum_{i \in q} \frac{L_{it}}{L_{qt}} (W_{it} - \overline{W}_{qt})^2$. The quantiles have been

computed for the productivity distribution of the whole economy or individual industries; in the baseline case the analysis has been carried out at the A7 industry level (j subscripts are omitted for simplicity).

These decompositions are keys to understand whether increases in wage dispersion are mostly coming from heterogeneity across firms within narrowly defined industries or across industries. This shows whether the increase in wage dispersion is a phenomenon caused by an increase in volatility within industries, or a reallocation of labour to industries characterized by a higher dispersion; the within industry component can then be further decomposed in the contribution of each single industry, to precisely pin down where the within-industry increase in wage dispersion is coming from, as the next section will show. In a similar way, the decomposition in terms of quantiles of the productivity distribution can shed light on whether the increase in the wage dispersion took place only for firms at the top of the productivity distribution or in all segments of the distribution.

6.2.3. Industry share of wage volatility to total volatility

In addition to the above decomposition, it is possible to investigate the share of variance by industry. Similarly to Carvalho and Gabaix (2013), one can decompose the within-industry component of the wage variance into the labour-weighted wage variance of industries, as shown in Equation (40). This is an exact decomposition of the within-industry component of variance, which differs from what Carvalho and Gabaix define as fundamental volatility for the weights (not squared in the present case) and the variance (computed on the cross section of firms and not constant over time). This decomposition provides insight into the wage variance by elucidating the proportion for which each industry variance contributes, allowing one to identify the main potential drivers of wage disparity within an economy:

$$V_{Ft} = \sum_j \frac{L_{jt}}{L_t} \delta_{jt}^2, \quad (40)$$

where L_{jt} is the number of employees in sector j at time t , L_t is the total number of workers in the economy, and δ_{jt}^2 is the employment-weighted industry variance of (firm-level average) wages.

In the next table we report for each country the top three contributors to wage variance, for manufacturing and services respectively. The tables show that across countries some sectors, such as Wholesale and retail trade, Legal and accounting in services, as well as Food, beverages and tobacco, regularly appear amongst the sectors characterised by the highest wage dispersion. This result suggests that there might be sectoral features of the within sector distribution of firms that might affect the distribution of wages.

Table 15. Top three sectors in the share of within-sector wage variance, 2011

Country (2011)	Manufacturing	% Var. Share	Services	% Var. Share
AUS	Basic metals and fabricated metal products, except machinery and equipment [CH]	22%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	37%
	Food products, beverages and tobacco [CA]	18%	Administrative and support service activities [N]	30%
	Wood and paper products, and printing [CC]	13%	Legal and accounting activities, etc. [MA]	13%
AUT	Furniture; other manufacturing; repair and installation of machinery and equipment [CM]	17%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	33%
	Basic metals and fabricated metal products, except machinery and equipment [CH]	15%	Legal and accounting activities, etc. [MA]	22%
	Food products, beverages and tobacco [CA]	14%	Transportation and storage [H]	12%
BEL	Food products, beverages and tobacco [CA]	31%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	45%
	Chemicals and chemical products [CE]	27%	Transportation and storage [H]	18%
	Rubber and plastics products, and other non-metallic	12%	Legal and accounting activities, etc. [MA]	14%

Country (2011)	Manufacturing	% Var. Share	Services	% Var. Share
	mineral products [CG]			
CHL	Chemicals and chemical products [CE]	22%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	29%
	Food products, beverages and tobacco [CA]	15%	Advertising and market research; other professional, scientific and technical activities; veterinary activities [MC]	22%
	Basic metals and fabricated metal products, except machinery and equipment [CH]	13%	Telecommunications [JB]	13%
DNK	Food products, beverages and tobacco [CA]	26%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	43%
	Machinery and equipment n.e.c. [CK]	22%	Transportation and storage [H]	15%
	Basic metals and fabricated metal products, except machinery and equipment [CH]	12%	Legal and accounting activities, etc. [MA]	11%
FIN	Computer, electronic and optical products [CI]	23%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	38%
	Wood and paper products, and printing [CC]	21%	Legal and accounting activities, etc. [MA]	18%
	Machinery and equipment n.e.c. [CK]	13%	Transportation and storage [H]	17%
FRA	Food products, beverages and tobacco [CA]	17%	Legal and accounting activities, etc. [MA]	35%
	Transport equipment [CL]	11%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	26%
	Chemicals and chemical products [CE]	10%	Administrative and support service activities [N]	11%
HUN	Basic metals and fabricated metal products, except machinery and equipment [CH]	18%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	28%
	Transport equipment [CL]	15%	Legal and accounting activities, etc. [MA]	18%
	Food products, beverages and tobacco [CA]	13%	IT and other information services [JC]	15%
IDN	Textiles, wearing apparel, leather and related products [CB]	24%		
	Food products, beverages and tobacco [CA]	21%		
	Rubber and plastics products, and other non-metallic mineral products [CG]	16%		
ITA	Machinery and equipment n.e.c. [CK]	19%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	34%
	Basic metals and fabricated metal products, except machinery and equipment [CH]	14%	IT and other information services [JC]	14%
	Rubber and plastics products, and other non-metallic mineral products [CG]	10%	Administrative and support service activities [N]	13%
JPN	Transport equipment [CL]	20%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	37%
	Food products, beverages and tobacco [CA]	16%	Legal and accounting activities, etc. [MA]	36%
	Chemicals and chemical products [CE]	14%	Administrative and support service activities [N]	9%
NLD	Coke and refined petroleum products [CD]	36%	Administrative and support service activities [N]	68%
	Chemicals and chemical products [CE]	18%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	15%
	Food products, beverages and tobacco [CA]	15%	Transportation and storage [H]	8%
NOR	Furniture; other manufacturing; repair and installation of machinery and equipment [CM]	26%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	29%
	Food products, beverages and tobacco [CA]	14%	Legal and accounting activities, etc. [MA]	16%
	Transport equipment [CL]	12%	Transportation and storage [H]	15%
PRT	Food products, beverages and tobacco [CA]	13%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	32%
	Basic metals and fabricated metal products, except machinery and equipment [CH]	13%	Legal and accounting activities, etc. [MA]	15%
	Rubber and plastics products, and other non-metallic mineral products [CG]	11%	Administrative and support service activities [N]	14%
SWE	Basic metals and fabricated metal products, except machinery and equipment [CH]	16%	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	31%
	Machinery and equipment n.e.c. [CK]	14%	Legal and accounting activities, etc. [MA]	23%
	Computer, electronic and optical products [CI]	13%	IT and other information services [JC]	10%

6.3. Policy implications regarding wage and productivity dispersion

Policies and regulations within countries may also impact the shape of the wage and productivity dispersions and the way they have evolved over time. For example, consider a country with little employment protection legislation. In the event of an adverse productivity shock, wage and productivity dispersion are likely to be closely related. During a macro contraction, the output of the least productive firms is likely to decrease quite substantially due to less demand, lowering their productivity. In reaction to this, firms may either exit or fire workers to try to offset a portion of their loss in revenue. As a result, wage and productivity dispersion for the worst performing firms go in parallel, and in this case both decrease. Alternatively firms might decide to hoard labour and reduce the wages of their employees. This will result in an increase in wage dispersion, but also in an increase in productivity dispersion, which is likely to follow suit because the least productive workers are now kept in the market (unless the fall in wages is so strong as to compensate for the lack of the adjustment in employment). The overall increase in dispersion will come from the fact that the best performing firms are likely to be more resilient to the productivity shock, experiencing little or no productivity decline. In addition, the top firms are in a position to absorb resources released from exiting and shrinking firms at the bottom potentially increasing their productivity (they can get the best workers of the worst firms at a lower wage). Hence, countries with less employment regulation are likely to display productivity and wage dispersions that move in parallel over the economic cycle.

For countries with substantial employment protection legislation however changes in productivity and wages may be less synchronised and less receptive to macro shocks. For example, assume that firms at the bottom of the productivity distribution cannot adjust to productivity shocks by reducing labour costs due to labour regulations. Firms can adjust to decreased demand by either reducing output and/or decreasing investment, which is likely to decrease productivity. Since firms cannot reduce pay or fire workers, wage dispersion will not increase; on the other hand productivity dispersion will widen, disconnecting it from the wage distribution. At the top of the distribution, firm productivity will potentially be more resistant to the shock and average productivity may remain relatively unchanged. At the same time, there will be fewer resources for the top performers to absorb since they are trapped in the bottom performers. Also, the most productive firms may be less willing to absorb available employment since regulations restrict them from reducing labour costs in the event of a future contraction, potentially reducing their future productivity growth. Therefore in highly regulated markets, average wage and productivity dispersion may be more disconnected, and may potentially feature a relatively more condensed wage distribution than the one for productivity.

The extent to which wage and productivity dispersion are related within a country therefore may potentially be a significant measure of misallocation. Countries where wage and productivity variance move in the same direction, may indicate market efficiency, where less productive firms shrink and their resources are allocated to firms at the top. Countries which attempt to shield workers during adverse market conditions may have less wage and productivity dispersion. On the one hand, this is of course beneficial to the workers as their jobs and salaries would be protected and shielded by the cycles. Since workers are generally more risk adverse this may be welfare improving in the short run and, in addition, such regulations would also support equity amongst workers during the economic cycle. On the other hand, however, less dispersion in wages and productivity due to substantial regulations may actually inadvertently impact aggregate productivity by distorting the flow of resources from less productive firms to more productivity ones. Despite the focus in the literature on the link from productivity to wages, there could be a reverse link from the wage distribution to the productivity distribution. Too much regulation and a compressed wage distribution may result in less workers incentives and hence less productivity growth. Policies that might be welfare improving in the short-run may have a detrimental impact in the long run.⁵² Greater dispersion in firm performance can result in faster aggregate productivity growth since resources are more effectively reallocated away from poorly performing to highly productive firms.

The data collected through the MultiProd project allow us to shed light on these important questions, and understand how country's specific policies and framework conditions shape the form of the wage distribution, the way it evolves over time and its relationship with the productivity distribution. For instance it is possible to show whether employment protection legislation has an effect on the dispersion of the wage distribution, and whether this in turn has an effect on the productivity distribution and aggregate productivity (and vice versa). Furthermore, the decomposition by segments of the productivity distribution offers a direct way of analysing the relationship between wage and productivity dispersion, possibly offering further evidence on some of the investigated channels, as the next section will show.

6.3.1. Wage dispersion across different segments of the productivity distribution

If economic policies play a role in shaping the wage distribution, one might expect that they will have a differential impact across segments of the productivity distribution. Consider a country with a high minimum wage: one might expect the wage dispersion of the bottom quantile of the productivity distribution to be more compressed in comparison to countries with less regulated markets. On the other hand, the variance at the top quantile could be quite similar, since employees there earn higher wages in the first place. The shape of the overall wage distribution might then be just a composition effect, with the overall compression coming from the firms at the bottom and a similar shape for firms at the top of the distribution.

A preliminary investigation as to whether misallocation distortions exist entails comparing wage variance across quantiles of the productivity distribution and across countries overtime, extending the graphical evidence of Section 6.2.1. This allows for the analysis of the extent to which wage dispersion has evolved overtime at all points of the business cycle and across different segments of the productivity distribution.

In the next figure we report the average wages in the top, bottom and the 4th to 6th decile of LP.

Figure 14. Change in real wages in different parts of the productivity distribution of firms over time

Note: The countries included are AUT, BEL, CHL, DNK, FIN, HUN, ITA, JPN, NLD, NOR. The graphs can be interpreted as the cumulated increase of average wages within each country and sector over the period. For instance, in 2012 in manufacturing the average wage paid by the 10% most productive firms is on average roughly 2,000 US dollars higher than in 2001. The estimates reported in the graph are those of year dummies in a cross-country regression of the average wage in the bottom, top and 4th to 6th decile of the productivity distribution.

The evidence suggests the higher the productivity of a firm, the higher the wages. Moreover, wage inequality between firms with different productivity performance has increased. There are, nonetheless, some important cross-sectoral differences in the magnitude of the gap between wages in the most productive firms and wages in the worst performing firms: among the most productive firms, wages have increased more in the service sector than in the manufacturing one.

A more extensive method is to decompose the within component of wage volatility by quantiles (e.g. deciles) of the productivity distribution. Building off Equation (40) one can use the following:

$$V_{Ft}^Q \equiv \sum_q \frac{L_{qt}}{L_t} \delta_{qt}^2, \quad (41)$$

where L_{qt} is the number of employees in quantile q at time t , L_t is the total number of workers in the economy, whereas δ_{qt}^2 is the employment-weighted quantile variance of firm-level average wages. Decomposing wage dispersion using Equation (41) will allow one to measure the share of each quantile to total wage dispersion. Comparing the shares of the bottom/top quantiles across countries with different EPLs before and after a productivity shock may provide evidence on potential distortions in the economy and on the role that policy plays in the misallocation of resources.

6.4. Potential additional contributions of the project

6.4.1. Increased wage dispersion: technology investment or increases in capital intensity?

Many of the theoretical models examining wage dispersion suggest that disparities are driven by heterogeneous technology usage across firms (Caselli 1999 and Aghion et al. 2002). Such heterogeneity in technological adoption is likely to be reflected in MFP dispersion. Bagger et al. (2012) however suggests that wages are predominately determined by labour quality and capital per worker. In this case, firms that employ more high-skilled workers who use greater amounts of capital pay higher average wages. Despite not having access to data on labour quality, the latter could be identified in the dispersion of capital intensity.⁵³

In order to examine whether wage dispersion is driven more by technological change or increased capital inputs, the MultiProd code could be easily extended to compare changes in wage dispersion across different segments of the capital intensity distribution to wage dispersion across different segments of the MFP distribution. Here capital intensity will be capital per worker and MFP can be calculated with one of the method used in MultiProd.

6.4.2. The causal link between labour productivity and wages

After investigating potential drivers of productivity dispersion, it is important to determine the extent to which wage and productivity dispersion are causally linked. Fundamental theories on labour demand imply that increases in wages should result in greater observed marginal product, through labour adjustments. At the same time, increased wages may increase the incentive to work harder thus increasing LP (Masters 2014). Given the probability of endogeneity, simple OLS estimation is likely to give biased results. To investigate the possibility of a causal effect of productivity on wages and rent sharing a natural experiment or an instrument variable is needed.

Following Card et al. (2014) and Barth et al. (2014) one could employ an instrument variable approach where average sales per worker of a firm in the same narrow defined industry but in a different region is an instrument for sales per worker for a firm.

$$\ln w_{ijr} = \alpha + \beta p_{ijr} + \gamma w_{jr} + \delta s_j + v_{ijr} . \quad (42)$$

Here p_{ijr} is sales per worker in establishment i , industry j and region r ; w_{jr} represents average wage of industry j in region r , i.e. an alternative wage measure; s_j is a vector of skills of a particular industry. The identifying assumption here is that an industry wide shock, conditional on the skill level of the workers in a given industry and local wages, affects sales per worker in the plant but is otherwise unrelated to their average wages. The IV strategy requires spatial information on the location of firms, which are currently not required in MultiProd. Still it could be useful to collect this information (at least for the sub-sample of countries for which it is available) to broaden the scope of MultiProd in a variety of interesting ways.

7. Conclusions

Productivity growth plays a central role in shaping the welfare of societies and the competitiveness of countries. Productivity differences for instance explain a large share of the differences in income per capita across countries. But as firm-level productivity can vary widely, even within narrowly defined industries, analysing industry-average productivity data does not offer a full picture. Countries might display the same average, for example, but this might be characterised by very different underlying distributions. This has important policy implications. For instance, low average productivity in a given country might be due to the presence of too few firms at the top, which would hint at an underlying lack of innovation, or too many firms at the bottom, which would point to weak market selection. These two opposite situations would call for very different policies. Understanding how firm-level productivity patterns translate into aggregate productivity is therefore a key challenge for policy makers.

The MultiProd project provides cross-country harmonised micro-aggregated data of paramount importance for understanding productivity performances across countries. It studies productivity patterns, investigating the extent to which different policy frameworks can shape firm productivity, and examining the way resources are allocated to more productive firms through a process known as allocative efficiency. The analysis aims to be a key input for policy makers as firm-level productivity and allocative efficiency are the engines of future growth. The project uses a distributed microdata approach drawing on the experience of the OECD DynEmp project.

This paper has outlined the past, present and future work of the MultiProd project in light of the current literature, stressing the main contributions of the project, particularly those related to investigating the role of policy and regulations in shaping the allocation of resources across firms and the impact on productivity and other aggregate outcomes. In summary, the MultiProd project aims to:

- Better depict the Schumpeterian process of creative destruction across countries;
- Gauge whether resources are efficiently allocated through the analysis of the entire firm-level productivity distribution, with further refinements by firm size, age, and ownership categories;
- Identify the largest firms and understand how they differ in terms of their weight in the economy, their productivity performance, and their contribution to aggregate productivity growth;
- Identify firms at the ‘frontier’ – the best performers – and understand how they differ across countries, what drives their performance, and how much they contribute to aggregate productivity growth;
- Investigate the within-country difference between the ‘frontier’ and the ‘laggard’ firms, and understand how this difference varies across countries, which are its determinants and how it affects both aggregate productivity growth and allocative efficiency;
- Investigate the cross-country differences in firm-level productivity performance and allocative efficiency before, during and after the financial crisis;
- Investigate the relationship between productivity and wage dispersion, and gauge to what extent heterogeneity in productivity has contributed to wage inequality;
- Examine the effectiveness of various policy frameworks aimed at shaping firm productivity and enhancing resource allocation to more productive firms.

These are just some of the interesting avenues that we plan to address in subsequent work. Last but not least we aim at linking differences in these important features of productivity dynamics and distributions to structural changes, such as digitalisation and globalisation, and countries’ framework conditions and policies.

APPENDIX

Metadata

Table A.1 provides details on the national data sources underlying the final database. In some cases, data come from unified tax databases, which include variables of interest for the whole population of taxable firms.

Table A.1. National data sources

COUNTRY	NATIONAL DATA SOURCES
AUS	Business Longitudinal Database
AUT	"Leistungs- und Strukturstatistik" (PS)
BEL	National Social Security Office and Belfirst
CAN	Statistics Canada's T2-LEAP longitudinal database
CHL	Chilean Tax office database (Servicio de Impuestos Internos de Chile)
DNK	Accounts Statistics for Non-Agricultural Private Sector, General Enterprise Statistics
FIN	Structural Business Statistics data (surveys, corporate taxation records, and Statistics Finland's Business Register)
FRA	FICUS/FARE
HUN	Corporate income tax data (CIT)
IDN	Annual Manufacturing Survey
ITA	ASIA (Business Register), Indagine sulle grandi imprese (SCI), Database Commercio Estero (COE), Balance sheet data for capital companies
JPN	Basic Survey of Japanese Business Structure and Activities
LUX	Register (Repertoire) and Structural Business Survey (SBS)
NLD	Productiestatistieken (PS) + Algemeen Bedrijvenregister (BR)
NOR	Annual Business Accounts Data
NZL	Longitudinal Business Frame (LBF), Linked Employer-Employee Data (LEED), Tax-filed company accounts (IR10), and Annual Enterprise Survey (AES)
PRT	Sistema de Contas Integradas das Empresas, Inquérito Harmonizado às Empresas (IEH), Informação Empresarial Simplificada (IES)
SWE	SBS administrative data (tax data)

Details about sectoral classifications and coverage of the source data are listed in Table A.2. Table A.3 reports instead information on optional variables available, namely: intermediate inputs; birth year of the units; binary variables indicating whether firms belong to a group and whether their owner is foreign.

Table A.2. Sectoral classifications and coverage

COUNTRY	Original Classification	Digit level	Excluded industries w.r.t. MultiProd
AUS	ANZSIC2006	4	None
AUT	(Ö)NACE 2008 (Nace Rev.2)	4	1, from 85 to 90
CAN	North American Industry Classification System (NAICS)	4	None
BEL	NACE-BELL (US SIC also available)	3	None
CHL	CIIU4.CL 2012 (based on ISIC)	4	None
DNK	NACE Rev. 2	4	35 for years 2005-06 and 2008-12, from 85 to 90 for years 2008-12
FIN	NACE Rev 1.1 and 2	3	None
FRA	Nomenclature d'Activités Française (NAF)	4	None
HUN	TEÁOR (NACE Rev 1.1 and 2)	4	12, 19
IDN	ISIC (Indonesian Standard Industry Classification)	3	1, 5, from 35 onwards
ITA	Ateco rev. 1.1 and Rev. 2	3	1
JPN	Japan Standard Industrial Classification (JSIC), rev 8	3	1, 36, 41, 72
LUX	NACE Rev 1.1 and 2	3	1-3, 85-96
NLD	NACE Rev 1.1 and 2	3	1, 5
NOR	NACE Rev 1.1 and 2	3	None
NZL	ANZSIC2006		None
PRT	Nace Rev. 2	3	None
SWE	NACE Rev 1.1 and 2	4	None

Table A.3. Optional variables

COUNTRY	Intermediates	Birth	Group	Foreign
AUS				
AUT				
CAN				
BEL				
CHL				
DNK				
FIN				
FRA				
HUN				
IDN				
ITA				
JPN				
LUX				
NLD				
NOR				
NZL				
PRT				
SWE				

Depending on the country, the employment variables could refer only to the number of salaried employees or to the number of persons engaged (including independent workers, owners, etc.). In Table A.4 we indicate for each country if the variable used as labour input covers total employment or only salaried employees.

Table A.4. Labour input variable.

COUNTRY	L input	
	Persons engaged	Employees
AUS		
AUT		
CAN		
BEL		
CHL		
DNK		
FIN		
FRA		
HUN		
IDN		
ITA		
JPN		
LUX		
NLD		
NOR		
NZL		
PRT		
SWE		

Note: "Employees" are defined as all persons who, by agreement, work for another resident institutional unit and receive remuneration. "Persons engaged" are defined as the total number of persons who work in or for the establishment including working proprietors, active business partners and unpaid family workers, as well as persons working outside the establishment when paid by and under the control of the establishment. Bento and Restuccia (2016) address the issue of different definitions of the employment variable (persons engaged vs. employees, and full-time equivalents vs. headcounts).

*Tables on sectoral, size and age aggregation***Table A.5. Sectoral Aggregation**

STAN A38 aggregation based on ISIC v.4 classification	STAN A7 aggregation
01 to 03 AGRICULTURE, FORESTRY AND FISHING	Agriculture
05 to 09 Mining and quarrying	Mining
10 to 12 Food products, beverages and tobacco	Manufacturing
13 to 15 Textiles, wearing apparel, leather and related products	
16 to 18 Wood and paper products, and printing	
19 Coke and refined petroleum products	
20 Chemicals and chemical products	
21 Basic pharmaceutical products and pharmaceutical preparations	
22 to 23 Rubber and plastics products, and other non-metallic mineral products	
24 to 25 Basic metals and fabricated metal products, except machinery and equipment	
26 Computer, electronic and optical products	
27 Electrical equipment	
28 Machinery and equipment n.e.c.	
29 to 30 Transport equipment	
31 to 33 Furniture; other manufacturing; repair and installation of machinery and equipment	
35 Electricity, gas, steam and air conditioning supply	Utilities
36 to 39 Water supply; sewerage, waste management and remediation activities	Construction
41 to 43 CONSTRUCTION	
45 to 47 Wholesale and retail trade, repair of motor vehicles and motorcycles	Market services
49 to 53 Transportation and storage	
55 to 56 Accommodation and food service activities	
58 to 60 Publishing, audiovisual and broadcasting activities	
61 Telecommunications	
62 to 63 IT and other information services	
64 to 66 FINANCIAL AND INSURANCE ACTIVITIES	<i>Excluded</i>
68 REAL ESTATE ACTIVITIES	Market services (continued)
69 to 71 Legal and accounting activities; activities of head offices; management consultancy activities; architecture and engineering activities; technical testing and analysis	
72 Scientific research and development	
73 to 75 Advertising and market research; other professional, scientific and technical activities; veterinary activities	
77 to 82 Administrative and support service activities	<i>Excluded</i>
84 Public administration and defence	
85 Education	
86 to 88 Human health and social work activities	
90 to 93 Arts, entertainment and recreation	Non Market services
94 to 96 Other service activities	

Table A.6. Size Classes⁵⁴

Size Class Name	Description
Size Class 21	Two size classes with threshold at 5 employees: vMicro (<5), SmallMediumLarge (>=5)
Size Class 2121	Two size classes with threshold at 5 employees but statistics not reporter for the first group: SmallMediumLarge (>=5)
Size Class 22	Two size classes with threshold at 10 employees: Small (<10), MediumLarge (>=10)
Size Class 2222	Two size classes with threshold at 10 employees but statistics not reporter for the first group: MediumLarge (>=5)
Size Class 42	Four size classes with thresholds at 10, 50, and 250 employees: Micro (1-9), Small (10-49), Medium (50-249), Large (250+)
Size Class 4222	Four size classes with thresholds at 10, 50, and 250 employees but statistics not reporter for the first group: Small (10-49), Medium (50-249), Large (250+)
Size Class 52	Five size classes with thresholds at 10, 20, 50, and 250 employees: Micro (1-9), Small (10-19), MediumSmall (20-49), Medium (50-249), Large (250+)
Size Class 71	Seven size classes with thresholds at 5, 10, 20, 50, 100 and 250 employees: vMicro (<5), Micro (5-9), Small (10-19), MediumSmall (20-49), Medium (50-99), MediumLarge (100-249), Large (250+)
Size Class 81	Eight size classes with thresholds at 5, 10, 20, 50, 100, 250 and 500 employees: vMicro (<5), Micro (5-9), Small (10-19), MediumSmall (20-49), Medium (50-99), MediumLarge (100-249), Large (250-499), vLarge (500+)

Table A.7. Age Classes

Size Class Name	Description
Age Class 21	Two age classes with threshold at 5 years: Young (0-5), Old (6+)
Age Class 31	Three age classes with thresholds at 0 and 10 years: Birth (0), Young (1-10), Old (10+)
Age Class 32	Three age classes with thresholds at 2 and 5 years: Startup (0-2), Young (3-5), Old (5+)

Description of the MultiProd output

File names and dimensionality of the output

MultiProd produces a conspicuous number of output files; each file name ends with the list of variables defining what constitutes a cell in the output file. Firms can be aggregated based on firm-specific dimensions, such as sector, age class, size class; or endogenous dimensions, where a given firm's classification will depend on other firms too, such as quantiles of productivity.

To give an idea of the output produced the table below details the firm-specific variables that can be used, sometimes in combination, to define a cell.

Table A.8. Firm-specific variables used to define a cell

Variable	Description	Maximum number of categories
ind_a7	Macro-sector classification based on a coarser <u>OECD STAN A*10 classification</u>	7
ind_a38	Sectoral classification based on the <u>OECD STAN A*38 classification</u>	38
group_nat	Whether firm belongs to a domestic group or not	2
ageclass21	Age class: Young (0-5 years), Old (6+ years)	2
sizeclass22	Size classes: Small (<10 employees), MediumLarge (10 employees or more)	2
sizeclass52	Size classes: Micro (1-9 employees), Small (10-19), MediumSmall (20-49), Medium (50-249) and Large (250+)	5
sizeclass5222	Same as sizeclass52, but no data is collected on Micro firms (9 employees or less)	4
sizeclass81	Size classes: vMicro (<5 employees), Micro (5-9), Small (10-19), MediumSmall (20-49), Medium (50-99), MediumLarge (100-249), Large (250-499) and vLarge (500+)	8
demographics_L	Demographic of the firm: incumbent, entrant, future exitor, blip firm (appears one year only)—only used for measures of job creation	4

In addition to these groups that depend on firm-specific characteristics, some statistics are also computed based on firms' position relative to the other firms in their sectors, based on a given variable. Firms are then grouped into quantile classes corresponding to their relative positions.

Four variables are used to create such quantile classes: three measures of productivity (LogLP_VA, LogMFP_SW and LogMFP_W) both in levels and growth— as well as gross output GO.⁵⁵ These quantile classes are either cut off at the 10th and 90th percentiles (e.g. in order to compare the top and bottom deciles of the productivity distribution), or at the 10th, 40th, 60th and 90th percentiles, in order to include a “median” (40-60) group of firms in the comparison.

General Syntax of Variable Names

Most variables are of the form: [QUANTITY]_[statistic], where *QUANTITY* is the economic quantity being measured and *statistic* is the type of statistic being collected. For example: quantity L stands for the number of employees in a given firm, so L_av measures the average number of employees by firm.

Table A.9. List of statistics computed for a given quantity

Type of statistics	Description
_av	Mean
_sd	Standard deviation
_pXY	XY th percentile of the variable. In particular, p50 is the median.
_N	Number of observations (see below).
_pareto_b ; _pareto_se	Estimated Pareto coefficient (and associated standard error) for the distribution of the variable.

Importantly, for any quantity, a variable `[quantity]_N` recording the number of firms on which the other statistics of *measure* were computed. The number of observations used for calculations is unique to each measure, as it is possible that different data are present or missing in the original dataset.

There are some exceptions to the general syntax. Some variables, such as the correlations or weighted variables, do not follow the general syntax `[quantity]_[statistic]`. For further details on the exceptions please refer to the output description included in the package sent to run the code.

Basic Firm-Level Quantities

The following table details the basic data and their description. This pertains to firm-level data, which are being aggregated into different statistics.

Table A.10. List of Basic Quantities

Quantity	Description
L	Number of employee (headcount)
L_FTE	Number of employees (full time equivalent)
G0	Gross Output
VA	Value Added
W	Real wage
age	Age
K	Capital (either book values or using PIM)
K_Rob	Capital robustness check (book values, only if both K and I are available)
I	Investment
II	Intermediates
LP_VA	Labour Productivity (Value added per employee)
LP_VA_FTE	Labour Productivity (Value added per employee), using FTE instead of headcount if available
MFP_W	Multi-Factor Productivity, computed using Wooldridge (2009) instrumented estimation
MFP_SW	Multi-Factor Productivity, computed as a Solow residual, using external, industry specific labour shares from OECD STAN (the cross-country-year median, one value for each industry)
MFP_SW_Rob	Same as above, but using the alternative measure of capital K_Rob (if available)

Variables Derived from Basic Quantities

Based on the basic quantities listed in in the previous table, a number of additional variables are computed.

Logged variables

The most common variable transformation is Log, which gives the logged value of a quantity. For instance `LogG0_av` represents the average of logged gross output in that cell.

Time transformations

Other transformations typically involve a time-dimension, e.g. using values of a given quantity in past or future periods. These other transformations are often compound with Log, since log-difference represents (approximate) growth rates.

Below is a list of the main transformations.

Table A.11. Time transformations

Transformation	Description
L1	Lagged value (last period)
D1	Backward-looking first-difference: $x_t - x_{(t-1)}$
F1	Forward-looking first-difference: $x_{(t+1)} - x_t$
A1	The lead of the variable
M2	2-period moving average (forward-looking)
B5	Average over previous years (up to 5 years)
C5	Average over future years (up to 5 years)
P2	Squared value.
N1	Only applicable to L. N1L is the net job creation between current and future period.

Note that transformed variables are computed at the firm-level, *before* being aggregated into a statistic. For instance, LogLP_VA_av is the average of the logged labour productivity, not the log of the average labour productivity.

Details of the MultiProd code

Details on the Computation of Weights

The main issues that are worth highlighting in the calculation of the weights are the following:

- The code always checks that the number of firms in the survey is at most as large as in the population; otherwise it sets the weight to 1. This is a non-zero probability event, in particular in case of external population breakdowns (e.g.: separate run on the BR only) together with a break in the industry classification, or in case the population structure is missing in some years. The latter case is possible when the population is available only in some Census years (e.g. for Japan or USA), the program then obtains interpolated values in the missing years and a firm might change size group between Census and non-Census years. The former case, despite being unlikely in general, is still possible because the program adopts a random assignment of firms to new industries in case of a change in the classification, hence firms can be assigned to different industries in the separate run on the PS and on the BR.
- In case a VA-based MFP is calculated, the weighting is implemented on value added (rather than log VA). In this way, negative or zero VA observations are not re-weighted. The rationale for this choice is that the probability of observing negative VA cases should be a priori the same in the population and in the sample. Then missing variables for specific measures of productivity are taken into account (e.g. missing capital for MFP). There is no need to take weights on the levels for any other log variable because negative or zero values are dropped in data cleaning. For the

specific case of MFP_W the number of firms in the sample is obtained by counting the non-missing observations for the variable itself and adding the observations with negative or zero VA.⁵⁶

- The weights used for calculating annual growth rates or 3/5 year annualised growth rates are based on specific population structures obtained to map the availability of firms over a 2, 3, 5 year horizon in the population. But note that: a) this has not been implemented for lags; b) for countries that have the census every X year (e.g.: JPN), the annual growth rates or 3 year annualized growth rates are going to take a zero value (apart from years that span two consecutive censuses e.g. 5 year annualised growth rates for the USA). In this particular case (when all values are zero or missing), the code sets all weights to missing and the weighted results will not be reported. The alternative would be to set the weight to 1, but this would be very confusing because un-weighted results would be reported together with weighted results.

The weighted means and percentiles do not depend on the type of STATA® weights that are used. In principle *pweights* should be used but these types of weights do not allow the user to calculate standard deviations. Thus we use the *awweights* option, which does what one would expect from *pweights* and exactly the same thing as the *svy* commands.⁵⁷

NOTES

-
- ¹ See Hsieh and Klenow (2010) for a more recent review of the literature.
- ² See Parente and Prescott (1994), Comin and Hobijn (2010), Schmitz (2005) and Bloom et al (2013).
- ³ See Restuccia and Rogerson (2013) for a review on the misallocation literature, and to Foster et al. (2001, 2002, 2014) for examples of work in the reallocation literature.
- ⁴ The time period is country specific depending on data availability and might be limited to more recent years for some countries.
- ⁵ The OECD currently follows the distributed microdata approach in three ongoing projects: MultiProd, DynEmp and MicroBeRD. The DynEmp (Dynamics of Employment) project provides harmonised micro-aggregated data to analyse employment dynamics (find out more: www.oecd.org/sti/dynemp) and MicroBeRD provides information on R&D activity in firms from official business R&D surveys (find out more: www.oecd.org/sti/rdtax).
- ⁶ For more details on the evolution of between-firm wage inequality and its relationship with productivity dispersion, see Berlingieri, Blanchenay and Criscuolo (2017).
- ⁷ For example a small contribution of the frontier to aggregate productivity growth could be explained by two different mechanisms that would require very different policies. On the one hand, it could be due to the fact that firms at the frontier are too small compared to other countries, hence the analysis should be devoted to understanding the reasons why firms do not grow and reach the needed size, which might be due to the presence of size contingent policies. On the other hand, a small contribution of the frontier could be related to lower productivity growth compared to other countries, thereby requiring policies such as tax credits which incentivise R&D spending and technology adoption.
- ⁸ The MultiProd Stata code and its help files are distributed to network member countries. The package includes the instructions on how to run the code and a description of the main output files generated by the program.
- ⁹ With “granularity” we refer to the extent to which economic activity in general, and aggregate productivity in particular, is driven by a small number of large firms. When large firms represent a disproportionate share of the economy, indeed, aggregate fluctuations may be governed by idiosyncratic shocks to these large firms. This hypothesis - called the “granular hypothesis” and proposed in Gabaix (2011) – suggests that aggregate fluctuations can be traced back to micro shocks hitting a small number of large firms. See further details in section 5.
- ¹⁰ The MultiProd project builds directly on the experience gained in the recent distributed microdata project “DynEmp”, which relied on business registers to investigate employment dynamics and firm demographics and showed the important role of young firms in job creation in 18 countries and over the 2000s using a single tested STATA ado file (Criscuolo et al., 2014a and 2014b).
- ¹¹ The statistics are generally collected at the firm level. Generally, countries are asked to run the code at the level of the firm (enterprise), given that in the vast majority of countries in the MultiProd network production information are collected at the level of the enterprise, rather than at the level of the plant. But note that the code can be run also at the plant level.
- ¹² Note also that in some countries administrative data, productions surveys, and business register are merged together to supply a unique data source (e.g. Norway). Administrative data (e.g. tax records) typically

---

cover the entire population of firms and provide information on production outputs and inputs, hence they are used as unique data source (e.g. France).

- 13 However, when the first year in which the unit appears with non-zero employment coincides with the first year of the business register or the year of left-censoring, then the birth year is left as missing.
- 14 To ensure international comparability, exchange rates and PPPs are applied in 2005. See section 2.4.2. for further details.
- 15 This section draws heavily on Gal (2013). For an overview of productivity measures at the aggregate and industry-level see OECD (2001), many of the issues discussed there apply also at the firm level.
- 16 For this reason, this measure of productivity is not currently included in the output of the code to limit the number of variables requested to countries. In any case note that if one considers within-sector patterns, this is less of a problem, since firms in the same sector tend to have similar intensities of intermediate input usage. Revenue based labour productivity has been used, for instance, in Foster et al (2001), where the authors mention that the results in terms of their decompositions and relative productivity are very similar across productivity measures.
- 17 Note that usually all MFP measures assume that firms are price takers on both output and input markets and maximise profits. Also note that assumptions about returns to scale are not too critical from an empirical point of view; according to the experiments on productivity measurement performed by Van Biesebroeck (2007), all approaches described score relatively well irrespective of such assumptions.
- 18 For the original idea, see Caves et al. (1982) and for applications, see Griffith et al. (2009) and Arnold et al. (2008).
- 19 The external aggregate labour and intermediate shares are taken from the OECD STAN database. Whenever the labour and intermediates' shares are missing (or their sum larger than 1), the value is taken from sectors at a more aggregated level, as for deflators (see below). Note that the aggregate labour share might be underestimated because the income of self-employed workers is missing. The share of labour input costs published in the STD Productivity database is adjusted for self-employed income but unfortunately the database does not have the country-industry-time coverage needed. Adjustments like the one suggested in the OECD Productivity Manual (OECD, 2001) have not been carried out because this might end up overestimating the labour share, and the micro-data are more likely to underestimate it as well (e.g., temporary workers from employment agencies are not always included).
- 20 The firm level labour and intermediate shares are winsorised at the top and bottom 1% to remove outliers (i.e. all data points below the 1th or above the 99th percentile are set respectively to the 1th and 99th percentile values). In case the resulting capital share is negative, it is set to missing and productivity is not computed (incidence is very low: only 0.1% of observations in the micro data used for testing).
- 21 Endogeneity arises because as productivity changes, optimizing firms react by adjusting their inputs (right hand side variables), and productivity also directly affects value added or gross output (the left hand side variable).
- 22 Another strand of production function estimators comes from the dynamic panel data literature (e.g. Blundell and Bond, 2000), but they have lost popularity compared to the semi-parametric approaches discussed here.
- 23 Their method also controls for selection bias. Selection can cause a problem since firms with higher capital stock will more easily weather negative productivity shocks and are more likely to stay in the sample than others.

- 24 In particular, the Akerberg et al. (2006) critique refers to a co-linearity problem: the joint inclusion of the nonparametric, polynomial terms of the variable input (e.g. labour) together with its structural coefficient in the production function makes the latter potentially unidentified.
- 25 Conversely to Olley and Pakes (1996), rented capital is not considered to compute the total capital stock because this information is not available in all countries. Consequently the capital stock might be underestimated, and particularly for small firms.
- 26 Note that while employment from BR (if available) is used to define the size classes for the weighting, employment from PS is used for all other calculations. A consequence of this choice is that when the `sizeclasses` are defined for the output statistics (which are based on PS), the weights may not be entirely correct in case of inconsistencies between the BR and PS employment measures, or in case the population breakdown is obtained through a separate run. Despite this shortcoming, this choice is better than defining the `sizeclasses` for the output statistics on BR (with PS employment for calculations), because in that case there would be an inconsistency between the `sizeclasses` and actual firms contained in the class.
- 27 For more information, see www.oecd.org/std/ppp/manual.
- 28 NACE and ISIC have a one-to-one correspondence at 2-digit level.
- 29 Note that the cleaning is robust to missing observations (gaps) in the Production Survey because it simply treats consecutive observations as consecutive years. This implies that the jumps may be defined over a time period longer than 1 year, which could make the jump more plausible. This is the reason why the jump parameters are set to fairly high and hence more conservative values.
- 30 In the Netherlands the coverage of the survey changes year by year as in some benchmark years Statistics Netherlands surveys a larger number of firms; for instance, in 2009 the share increases to 7.3%.
- 31 The table display the average dispersion within 2-digit sectors. While performing the same exercise within 4-digit sectors would partially reduce the dispersion (not available to us due to confidentiality), the comparison across countries at the 2-digit is nevertheless informative. Moreover the microdata used for MultiProd contain, or are representative for, the whole population of firms with at least one employee, which naturally imply a higher level of dispersion than other more selected samples.
- 32 Note that this is the variance decomposition of labour productivity. It can be generalised to MFP but the choice of the appropriate weights becomes less straightforward. In the literature it is common to use output weights (GO or VA depending on how MFP is estimated) but the resulting weighted average does not correspond to the precise measure of aggregate productivity. Moreover the standard Domar weights used to decompose GO MFP productivity growth do not yield an exact decomposition. Van Biesebroeck (2008) shows that to do so one would need more complex input weights.
- 33 This is an exact decomposition of the within-industry component of variance, which differs from what they define as fundamental volatility for the weights (not squared in the present case) and the variance (computed on the cross section of firms and not constant over time).
- 34 For a definition of “granularity” see footnote 9 in the Introduction; for further details see Section 5.
- 35 Among the sectors with lowest productivity dispersion we often find: “Textiles, wearing apparel, leather and related products”; “Electrical equipment”; and “Scientific research and development”.
- 36 Among the others, see also: Baily et al. (1992), Griliches and Regev (1995), Foster et al. (2001), and Baldwin and Gu (2002).

Various factor distortions have been analysed include, taxes, cost of doing business, credit constraints, size contingent policies, trade barriers and so on. For a review of this literature refer to Restuccia and Rogerson (2013).

Given the nature of the data, it will not be possible to precisely account for the entry and exit components for those countries that rely on production surveys. The approach proposed by Petrin and Levinsohn (2012) puts less emphasis on this dimension, an extra reason why their method is adopted within the MultiProd project.

In this indirect approach it is not necessary to specify exactly which mechanism generates these wedges.

In the remainder of this subsection, we use TFP instead of MFP to be consistent with the existing literature, and in particular with the Hsieh and Klenow (2009) paper on which it relies heavily.

This is a consequence of Jensen's inequality applied to CES aggregation.

Note that in the HK's setting the gains from reallocation are only coming from *within* sectors, not across sectors. All figures at the macro-sector level (e.g. manufacturing) are obtained as weighted averages of 4 digit industries.

See for instance Caballero and Hammour (1994), Hall et al (1995), and Gomes et al. (1997).

Note that there is a considerable literature, mainly theoretical, preceding Gabaix on proposing mechanisms that generate macroeconomics shocks from purely microeconomic causes (for instance, Jovanovic 1987, Durlauf 1993, Nirei 2006).

Alternatively, one could perform the opposite exercise and analyse what share of firms account for a certain level of market share (e.g. how many firms make up 25% of the market). This indicator is computed

as $t_j^{25} \equiv \frac{F_j^{sales^{-1}}(0.25)}{S_j}$ where F_j^{-1} is the inverse of the cumulative distribution of firm sales in sector j. Again, a smaller number of firms accounting for 25% of the market share would indicate a more concentrated economic activity, with all the consequences that this can bring.

A cross-country harmonised database of Herfindahl-Hirschman indices currently does not exist; this would be an extra contribution of the project that could be a useful input for future research.

Imagine for instance a sector with 4 firms. If the firms have equal market share of 25%, the corresponding Herfindahl-Hirschman index of the sector is $h_j = 1/4$. If instead the market is concentrated in one firm that has a market share of almost 100% while the three other firms have almost none, the corresponding Herfindahl-Hirschman index will be very close to 1. Note however that this measure assumes an autarchic economy (i.e. abstract from imports and exports).

More generally MultiProd will also collect a series of statistics for the "frontier" firms. See Andrews, Criscuolo and Gal (2015 and 2016) for more details on the literature and the type of analysis that can be performed on the subject.

An analysis of productivity growth over time however comes with a caveat. Either the identity of big players is calculated in every period, so that definitions of quantiles are time-specific, or it is possible to fix the identity of firms quantiles at a specific time and follow the contribution of those firms over time, which restricts the analysis to incumbents and might lead to problems for exiting firms.

-
- 50 This measure is calculated using a sample of workers who stay in the same establishment one year to the next, thereby holding fixed time-invariant unobservable and observable characteristics of both firms and individuals. Additional methods and datasets are used to calculate the between establishment contribution with values ranging from 86% to 67.5%. For the full details refer to Barth et al. (2014).
- 51 For a recent paper on this topic based on MultiProd data, see Berlingieri, Blanchenay and Criscuolo (2017).
- 52 A first hint in this direction is offered by the results of the DynEmp project (see Criscuolo et al., 2014). Looking at job creation and destruction, the results show that there is a clear relationship between the dispersion of the distribution and the performance of the top firm: in order to sustain high growth at the top there must be higher shrinkage at the bottom.
- 53 Some countries have measures of skill at the firm level. Hence a more direct analysis could be carried out for the sub-set of countries that have this information.
- 54 The table describes the size classes used in the main output files, more classes are defined in the code.
- 55 See below for a more exhaustive list of variables and their description.
- 56 Still, the calculation of MFP_W requires information on lagged values so it faces the same problem of annualised growth rate variables (the weights might be too large because even in the population the firms might not be present for two consecutive years – see below).
- 57 For further information refer to: <http://www.stata.com/support/faqs/statistics/weights-and-summary-statistics/>

REFERENCES

- Acemoglu, D., U. Akcigit, N. Bloom, and W. R. Kerr (2013), “Innovation, Reallocation and Growth”, *NBER Working Paper*, No. 18993, National Bureau of Economic Research.
- Akerberg, D., K. Caves, and G. Frazer (2006), “Structural Identification of Production Functions”, *MPRA Paper*, No. 38349, University Library of Munich, Germany.
- Aghion, P., T. Fally, and S. Scarpetta (2007), “Credit Constraints as a Barrier to the Entry and Post-Entry Growth of Firms”, *Economic Policy*, Vol. 22, No. 52, pp. 731–79.
- Aghion, P., P. Howitt, and G. Violante (2002), “Technological Progress, Luck and the Variability of Earnings”, *NYU Working Papers*.
- Andrews, D., C. Criscuolo and P. Gal (2015), “Frontier firms, technology diffusion and public policy: micro evidence from OECD countries”, *OECD Productivity Working Papers*, No. 2, OECD Publishing, Paris, DOI: <http://dx.doi.org/10.1787/5jrql2q2jj7b-en>.
- Andrews, D., C. Criscuolo, and P. N. Gal (2016), “The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy”, *OECD Productivity Working Papers*, No. 5, OECD Publishing, Paris, DOI: <http://dx.doi.org/10.1787/63629cc9-en>.
- Arnold, J., G. Nicoletti, and S. Scarpetta (2008), “Regulation, Allocative Efficiency and Productivity in OECD Countries”, *OECD Economics Department Working Papers*, No. 616, OECD Publishing, Paris, DOI: <http://dx.doi.org/10.1787/241447806226>.
- Arnold, J., S. Scarpetta, and G. Nicoletti (2011), “Regulation, Resource Reallocation and Productivity Growth”, *EIB Papers*, Vol. 16, No. 1, pp. 90–115.
- Atkinson, A., and T. Piketty (2009), *Top incomes over the twentieth century: a contrast between continental european and english-speaking countries*, OUP Oxford.
- Bagger, J., B. J. Christensen, and D. T. Mortensen (2012), “Wage and Labor Productivity Dispersion: The Roles of Total Factor Productivity, Labor Quality, Capital Intensity, and Rent Sharing”, Working paper, <http://www.eea-esem.com/files/papers/eea-esem/2013/3032/BCMpaper20120614.pdf>.
- Baily, M. N., C. Hulten, D. Campbell, T. Bresnahan, and R. E. Caves (1992), “Productivity Dynamics in Manufacturing Plants”, *Brookings Papers on Economic Activity. Microeconomics*, pp. 187–267.
- Baily, M. N. and R. M. Solow (2001), “International Productivity Comparisons Built from the Firm Level”, *Journal of Economic Perspectives*, Vol. 15, No. 3, pp. 151–172.
- Baldwin, J. R., and W. Gu (2002), “Plant turnover and productivity growth in Canadian manufacturing”, *OECD Science, Technology and Industry Working Papers*, No. 2002/02, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/745662258480>.
- Ball, L. M. (2014), “Long-Term Damage from the Great Recession in OECD Countries”, *NBER Working Paper*, No. 20185.
- Barlevy, G. (2002), “The Sullyying Effect of Recessions”, *The Review of Economic Studies*, Vol. 69, No. 1, pp. 65–96.

- Barnett, A., A. Chiu, J. Franklin, and M. Sebastia-Barriel (2014), “The Productivity Puzzle: A Firm-Level Investigation into Employment Behaviour and Resource Allocation over the Crisis”, *Bank of England Working Paper*, No. 495.
- Bartelsman, E. J. (2004), “The analysis of microdata from an international perspective”, OECD Statistics Directorate, STD/CSTAT(2004)12.
- Bartelsman, E. J., and M. Doms (2000), “Understanding Productivity: Lessons from Longitudinal Microdata”, *Journal of Economic Literature*, Vol. 38, No. 3, pp. 569–94.
- Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013), “Cross-Country Differences in Productivity: The Role of Allocation and Selection”, *The American Economic Review* 103 (1): 305–34.
- Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2009), “Measuring and Analyzing Cross-Country Differences in Firm Dynamics”, In *Producer Dynamics: New Evidence from Micro Data*, University of Chicago Press, pp. 15-76.
- Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2004), *Microeconomic Evidence of Creative Destruction in Industrial and Developing Countries*. Tinbergen Institute Discussion Paper 04-114/3. Tinbergen Institute.
- Bartelsman, E., S. Scarpetta, and F. Schivardi (2005), “Comparative Analysis of Firm Demographics and Survival: Evidence from Micro-level Sources in OECD Countries”, *Industrial and Corporate Change*, Vol. 14, No. 3, pp. 365–391.
- Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2014), “It’s Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the U.S.”, *NBER Working Paper*, No. 20447.
- Bento, P., and D. Restuccia (2016), “Misallocation, Establishment Size, and Productivity”, *NBER Working Paper*, No. 22809.
- Berlingieri, G., P. Blanchenay, and C. Criscuolo (2017), “The Great Divergence(s)”, *OECD Science, Technology and Industry Policy Papers*, No. 39, OECD Publishing, Paris.
- Bernard, A. B., J. Eaton, J. Bradford Jensen, and S. Kortum (2003), “Plants and Productivity in International Trade”, *American Economic Review*, Vol. 93, No. 4, pp. 1268–1290.
- Bernard, A. B., S. J. Redding, and P. K. Schott (2011), “Multiproduct Firms and Trade Liberalization”, *The Quarterly Journal of Economics*, Vol. 126, No. 3, pp. 1271–1318.
- Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013), “Does Management Matter? Evidence from India”, *The Quarterly Journal of Economics*, Vol. 128, No. 1, pp. 1–51.
- Bloom, N., and J. Van Reenen (2007), “Measuring and Explaining Management Practices Across Firms and Countries”, *The Quarterly Journal of Economics*, Vol. 122, No. 4, pp. 1351–1408.
- Blundell, R., and S. Bond (2000), “GMM Estimation with Persistent Panel Data: An Application to Production Functions”, *Econometric Reviews*, Vol. 19, No. 3, pp. 321–40.
- Bond, S., and M. Söderbom (2005), “Adjustment Costs and the Identification of Cobb Douglas Production Functions”, *IFS Working Papers*, No. 05/04.
- Buera, F.J., J.P. Kaboski, and Y. Shin (2011), “Finance and Development: A Tale of Two Sectors”, *The American Economic Review*, Vol. 101, No.5, pp. 1964–2002.
- Busso, M., L. Madrigal and C. Pagés (2013), “Productivity and Resource Misallocation in Latin America”, *The BE Journal of Macroeconomics*, 2013, Vol. 13, No. 1, pp. 903–932.
- Caballero, R. J., and M. L. Hammour (1994), “The Cleansing Effect of Recessions”, *The American Economic Review*, Vol. 84, No. 5, pp. 1350–68.

- Card, D., F. Devicienti, and A. Maida (2014), “Rent-Sharing, Holdup, and Wages: Evidence from Matched Panel Data”, *The Review of Economic Studies*, Vol. 81, No. 1, pp. 84–111.
- Card, D., J. Heining, and P. Kline (2013), “Workplace Heterogeneity and the Rise of West German Wage Inequality”, *The Quarterly Journal of Economics*, Vol. 128, No.3, pp. 967–1015.
- Carvalho, V. M., and X. Gabaix (2013), “The Great Diversification and Its Undoing”, *American Economic Review*, Vol. 103, No. 5, pp. 1697–1727.
- Caselli, F. (1999), “Technological Revolutions”, *American Economic Review*, Vol. 89, No. 1, pp. 78–102.
- Caselli, F. (2005), “Accounting for Cross-Country Income Differences”, *Handbook of Economic Growth*, Vol. 1, No. 9, pp. 679–741.
- Caves, D. W., L. R. Christensen, and W. Erwin Diewert (1982), “The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity”, *Econometrica: Journal of the Econometric Society*, Vol. 50, No. 6, pp. 1393–1414.
- Cochrane, J. H. (1994), “Shocks”, *NBER Working Paper*, No. 4698.
- Comin, D., and B. Hobijn (2006), “An Exploration of Technology Diffusion”, *NBER Working Paper*, No. 12314.
- Criscuolo, C., P. N. Gal, and C. Menon (2015), “DynEmp: A Routine for Distributed Microdata Analysis of Business Dynamics”, *Stata Journal*, Vol. 15, No. 1, pp. 247–274.
- Criscuolo, C., P. N. Gal, and C. Menon (2014), “The Dynamics of Employment Growth: New Evidence from 18 Countries”, *OECD Science, Technology and Industry Policy Papers*, No. 14, OECD Publishing, Paris, DOI: <http://dx.doi.org/10.1787/5jz417hj6hg6-en>.
- Daveri, F., R. Lecat, and M. L. Parisi (2016), “Service Deregulation, Competition and the Performance of French and Italian Firms”, *Scottish Journal of Political Economy*, Vol. 63, No. 3, pp. 278–302.
- Davis, Steven J, R. Jason Faberman, and John Haltiwanger (2006), “The Flow Approach to Labor Markets: New Data Sources and Micro–Macro Links”, *Journal of Economic Perspectives*, Vol. 20, No. 3, pp. 3–26.
- Davis, S. J, J. Haltiwanger, and S. Schuh (1996), *Job Creation and Destruction*. Cambridge MA: MIT Press.
- Davis, S. J., R. J. Faberman, and J. Haltiwanger (2012), “Labor Market Flows in the Cross Section and over Time”, *Journal of Monetary Economics*, Vol. 59, No. 1, pp. 1–18.
- Davis, S. J., and J. Haltiwanger (1999), “On the Driving Forces behind Cyclical Movements in Employment and Job Reallocation”, *American Economic Review*, Vol. 89, No. 5, pp. 1234–1258.
- Davis, S. J., and J. Haltiwanger (1992), “Gross Job Creation, Gross Job Destruction, and Employment Reallocation”, *The Quarterly Journal of Economics*, Vol. 107, No. 3, pp. 819–63.
- Di Giovanni, J., and A. A. Levchenko (2012), “Country Size, International Trade, and Aggregate Fluctuations in Granular Economies”, *Journal of Political Economy*, Vol. 120, No. 6, pp. 1083–1132.
- Duarte, M., and D Restuccia (2014), “Relative Prices and Sectoral Productivity”, *University of Toronto Working Paper*, No. 530.
- Dunne, T, L Foster, J Haltiwanger, and K R. Troske (2004), “Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment”, *Journal of Labor Economics*, Vol. 22, No. 2, pp. 397–429.

- Durlauf, S. (1993), “Non Ergodic Economic Growth”, *Review of Economic Studies*, Vol. 60, No. 2, pp. 349–366.
- Faggio, G., K. G. Salvanes, and J. Van Reenen (2010), “The Evolution of Inequality in Productivity and Wages: Panel Data Evidence”, *Industrial and Corporate Change*, Vol. 19, No. 6, pp. 1919–1951.
- Foster, L., C. Grim, and J. Haltiwanger (2014), *Reallocation in the Great Recession: Cleansing or Not?* Working Paper 20427. National Bureau of Economic Research. Working Paper No. 20427.
- Foster, L., J. Haltiwanger, and C. J. Krizan (2002), “The Link Between Aggregate and Micro Productivity Growth: Evidence from Retail Trade”, *NBER Working Paper*, No. 9120.
- Foster, L., J. C. Haltiwanger, and C. J. Krizan (2001), “Aggregate Productivity Growth. Lessons from Microeconomic Evidence”, In *New Developments in Productivity Analysis*, University of Chicago Press, pp. 303–72.
- Gabaix, X. (2011), “The Granular Origins of Aggregate Fluctuations”, *Econometrica*, Vol. 79, No. 3, pp. 733–72.
- Gal, P. N. (2013), “Measuring Total Factor Productivity at the Firm Level Using OECD-ORBIS”, *OECD Economics Department Working Papers*, No. 1049, OECD Publishing, Paris, DOI: <http://dx.doi.org/10.1787/5k46dsb25ls6-en>.
- Gomes, J., J. Greenwood, and S. Rebelo (2001), “Equilibrium Unemployment”, *Journal of Monetary Economics*, Vol. 48, No. 1, pp. 109–52.
- Griffith, R., S. Redding, and H. Simpson (2009), “Technological Catch-Up and Geographic Proximity”, *Journal of Regional Science*, Vol. 49, No. 4, pp. 689–720.
- Griliches, Z., and H. Regev (1995), “Productivity and firm turnover in Israeli industry: 1979–1988”, *Journal of Econometrics*, Vol. 65, No. 1, pp. 175–203.
- Guner, N., G. Ventura, and Y. Xu (2008), “Macroeconomic Implications of Size-Dependent Policies”, *Review of Economic Dynamics*, Vol. 11, No. 4, pp. 721–44.
- Hall, R. E. (2015), “Quantifying the Lasting Harm to the U.S. Economy from the Financial Crisis”, *NBER Macroeconomics Annual*, Vol. 29, No. 1, pp. 71–128.
- Hall, R. E., H. Farber, and J. Haltiwanger (1995), “Lost Jobs”, *Brookings Papers on Economic Activity*, Vol. 1995, No. 1, pp. 221–273.
- Hall, R. E., and C. I. Jones (1999), “Why Do Some Countries Produce So Much More Output Per Worker than Others?” *The Quarterly Journal of Economics*, Vol. 114, No. 1, pp. 83–116.
- Haltiwanger, J., S. Scarpetta, and H. Schweiger (2006), “Assessing Job Flows across Countries: The Role of Industry, Firm Size and Regulations”, *NBER Working Paper*, No. 13920.
- Hopenhayn, H. A. (1992), “Entry, Exit, and Firm Dynamics in Long Run Equilibrium”, *Econometrica*, Vol. 60, No. 5, pp. 1127–50.
- Hopenhayn, H., and R. Rogerson (1993), “Job Turnover and Policy Evaluation: A General Equilibrium Analysis”, *Journal of Political Economy*, Vol. 101, No. 5, pp. 915–938.
- Hsieh, C., and P. J. Klenow (2010), “Development Accounting”, *American Economic Journal: Macroeconomics*, Vol. 2, No. 1, pp. 207–223.
- Hsieh, C., and P. J. Klenow (2009), “Misallocation and Manufacturing TFP in China and India”, *The Quarterly Journal of Economics*, Vol. 124, No. 4, pp. 1403–1448.
- Hulten, C. R. (2001), “Total Factor Productivity. A Short Biography”, In *New Developments in Productivity Analysis*, University of Chicago Press, pp. 1–54.

- Hulten, C. R. (1978), “Growth Accounting with Intermediate Inputs”, *The Review of Economic Studies*, Vol. 45, No. 3, pp. 511–518.
- Jones, C. (2013), “Misallocation, Economic Growth, and Input-Output Economics”, In D. Acemoglu, M. Arellano, and E. Dekel, *Advances in Economics and Econometrics*, Volume II, Cambridge University Press.
- Jorgenson, D.W., M. S. Ho, J. D. Samuels, and K. J. Stiroh (2007), “Industry Origins of the American Productivity Resurgence”, *Economic Systems Research*, Vol. 19, No. 3, pp. 229–52.
- Jovanovic, B. (1987), “Micro Shocks and Aggregate Risk”, *Quarterly Journal of Economics*, Vol. 102, No. 2, pp. 395–409.
- Klapper, L., L. Laeven, and R. Rajan (2006), “Entry Regulation as a Barrier to Entrepreneurship”, *Journal of Financial Economics*, Vol. 82, No. 3, pp. 591–629.
- Klenow, P., and A. Rodriguez-Clare (1997), “The Neoclassical Revival in Growth Economics: Has It Gone Too Far?”, *NBER Macroeconomics Annual*, Vol. 12, pp. 73–103.
- Leal Ordóñez, J. C. (2014), “Tax Collection, the Informal Sector, and Productivity”, *Review of Economic Dynamics*, Vol. 17, No. 2, pp. 262–286.
- Levinsohn, J., and A. Petrin (2003), “Estimating Production Functions Using Inputs to Control for Unobservables”, *The Review of Economic Studies*, Vol. 70, No. 2, pp. 317–341.
- Lim Y., C. Hee Hahn (2004), “Bankruptcy Policy Reform and Total Factor Productivity Dynamics in Korea: Evidence from Microdata”, In *Growth and Productivity in East Asia, NBER-East Asia Seminar on Economics, Volume 13*, University of Chicago Press, pp. 297–326.
- Lucas Jr, R. E. (1978), “On the Size Distribution of Business Firms”, *The Bell Journal of Economics*, Vol. 9, No. 2, pp. 508–23.
- Machin, S, and J. Van Reenen (2008), *The New Palgrave Dictionary of Economics*. Edited by Steven N. Durlauf and Lawrence E. Blume. 2nd ed. Wage Inequality. Basingstoke: Nature Publishing Group.
- Martin, J P., and S Scarpetta (2012), “Setting It Right: Employment Protection, Labour Reallocation and Productivity”, *De Economist*, Vol. 160, No. 2, pp. 89–116.
- Masters, A. (2013), “On the Firms’ Component of Wage Dispersion”, Working paper, <http://www.albany.edu/~am755146/papers/pdweb.pdf>.
- Melitz, M. J. (2003), “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”, *Econometrica*, Vol. 71, No. 6, pp. 1695–1725.
- Melitz, M. J., and G. IP Ottaviano (2008), “Market Size, Trade, and Productivity”, *The Review of Economic Studies*, Vol. 75, No. 1, pp. 295–316.
- Melitz, M.J., and S. Polanec (2015), “Dynamic Olley-Pakes Productivity Decomposition with Entry and Exit”, *RAND Journal of Economics*, Vol. 46, No. 2, pp. 362–375.
- Mortensen, D. T., and C. A. Pissarides (1994), “Job Creation and Job Destruction in the Theory of Unemployment”, *The Review of Economic Studies*, Vol. 61, No. 3, pp. 397–415.
- Mueller, S. (2008), “Capital Stock Approximation Using Firm Level Panel Data”, *Jahrbücher für Nationalökonomie und Statistik*, Vol. 228, No. 4, pp. 357–371.
- Nirei, M. (2006), “Threshold Behavior and Aggregate Critical Fluctuations”, *Journal of Economic Theory*, Vol. 127, No. 1, pp. 309–322.
- OECD (2014a), “Gross Domestic Product (GDP : GDP per Head, US \$, Current Prices, Current PPPs”, <http://stats.oecd.org/index.aspx?queryid=558>.

- OECD (2014b), “Gross Domestic Product (GDP) : GDP, Volume – Annual Growth Rates in Percentage”, <http://stats.oecd.org/index.aspx?queryid=26646>.
- OECD (2001), *Measuring Productivity - OECD Manual: Measurement of Aggregate and Industry-level Productivity Growth*, OECD Publishing, Paris, DOI: <http://dx.doi.org/10.1787/9789264194519-en>
- Olley, G. S., and A. Pakes (1996), “The Dynamics of Productivity in the Telecommunications Equipment Industry”, *Econometrica*, Vol. 64, No. 6, pp. 1263–1297.
- Ouyang, M. (2009), “The Scarring Effect of Recessions”, *Journal of Monetary Economics*, Vol. 56, No. 2, pp. 184–99.
- Paci, P., A. Revenga, and B. Rijkers (2012), “Coping with Crises: Policies to Protect Employment and Earnings”, *The World Bank Research Observer*, Vol. 27, No. 1, pp. 106–141.
- Parente, S. L., and E. C. Prescott (1994), “Barriers to Technology Adoption and Development”, *Journal of Political Economy*, Vol. 102, No. 2, pp. 298–321.
- Petrin, A., and J. Levinsohn (2012), “Measuring Aggregate Productivity Growth Using Plant-Level Data”, *The RAND Journal of Economics*, Vol. 43, No. 4, pp. 705–25.
- Petrin, A., and J. Sivadasan (2013), “Estimating Lost Output From Allocative Inefficiency, with an Application to Chile and Firing Costs”, *The Review of Economics and Statistics*, Vol. 95, No. 1, pp. 286–301.
- Pissarides, C. A. (2013), “Unemployment in the Great Recession”, *Economica*, Vol. 80, No. 319, pp. 385–403.
- Prescott, E. C. (1998), “Lawrence R. Klein Lecture 1997: Needed: A Theory of Total Factor Productivity”, *International Economic Review*, Vol. 11, No. 4, pp. 525–51.
- Rajan, R. G., and L. Zingales (1998), “Which Capitalism? Lessons From the East Asian Crisis”, *Journal of Applied Corporate Finance*, Vol. 11, No. 3, pp. 40–48.
- Restuccia, D., and R. Rogerson (2008), “Policy Distortions and Aggregate Productivity with Heterogeneous Establishments”, *Review of Economic Dynamics*, Vol. 11, No. 4, pp. 707–20.
- Restuccia, D., and R. Rogerson (2013), “Misallocation and Productivity”, *Review of Economic Dynamics*, Vol. 16, No. 1, pp. 1–10.
- Schmitz Jr., J.A. (2005), “What Determines Productivity? Lessons from the Dramatic Recovery of the US and Canadian Iron Ore Industries Following Their Early 1980s Crisis”, *Journal of Political Economy*, Vol. 113, No. 3, pp. 582–625.
- Schumpeter, J. Alois (1939), *Business Cycles*, Volume I, Cambridge University Press.
- Selowsky M. and R. Martin (1997), “Policy Performance and Output Growth in the Transition Economies” *The American Economic Review*, Vol. 87, No. 2, pp. 349–353.
- Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2015), “Firming Up Inequality”, *NBER Working Paper*, No. 21199.
- Syversen, C. (2014), “The Importance of Measuring Dispersion in Firm-Level Outcomes”, *IZA World of Labor, Institute for the Study of Labor*, 1–10.
- Syversen, C. (2011), “What Determines Productivity?” *Journal of Economic Literature*, Vol. 49, No. 2, pp. 326–65.
- Syversen, C. (2007), “Prices, Spatial Competition and Heterogeneous Producers: An Empirical Test”, *The Journal of Industrial Economics*, Vol. 55, No. 2, pp. 197–222.

- Syverson, C. (2004) “Product Substitutability and Productivity Dispersion”, *Review of Economics and Statistics*, Vol. 86, No. 2, pp. 534–550.
- Van Biesebroeck, J. (2008), “Aggregating and Decomposing Productivity”, *Review of Business and Economics*, Vol. 53, No. 2, pp. 122–46.
- Van Biesebroeck, J. (2007), “Robustness of Productivity Estimates”, *The Journal of Industrial Economics*, Vol. 55, No. 3, pp. 529–69.
- Wooldridge, J. M. (2009), “On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables”, *Economics Letters*, Vol. 104, No. 3, pp. 112–114.
- Ziebarth, N. (2012), “Misallocation and Productivity during the Great Depression”, Working paper, http://faculty.chicagobooth.edu/workshops/macro/past/pdf/productivity_dispersion.pdf