Cross-country analysis of ICT impact using firm-level data: Micro Moments Database and Research Infrastructure[1]

Eric Bartelsman,Vrije Universiteit Amsterdam
e.j.bartelsman@feweb.vu.nl

Eva Hagsten, Statistics Sweden
eva.hagsten@scb.se

Michael Polder, Statistics Netherlands
jm.polder@cbs.nl

December 21, 2013

ABSTRACT

This paper provides technical documentation to a research database of moments collected from linked longitudinal firm-level datasets in a large selection of EU countries. Further, the paper will describe research infrastructure that has been developed, and the technical possibilities for the research community to conduct their own cross-country firm-level research using the infrastructure. Finally, the paper will provide a variety of findings from our own cross-country analysis of ICT impact, both using the micro-moments database and using the infrastructure for harmonized cross-country micro analysis.

[1] This work has been made possible by Eurostat Grant Agreement 50721.2013.001-2013.082.

1. Introduction

Conducting comparative analysis of enterprise data is a cumbersome endeavour, especially if the desired comparisons cross country boundaries. In a Eurostat-funded project on Linking of Micro-data to Analyse ICT Impact (ESSLait), researchers and statisticians from 14 European National Statistics Offices (NSOs), are providing the research community with two paths towards cross-country analysis.[2] First, using national linked firm-level sources, we have constructed a cross-country industry dataset at a medium-level of disaggregation that includes measures of ICT usage and innovative activity together with measures of business performance and industry dynamics. These measures include typical aggregates, such as sums and means, but also higher moments of distributions of variables of interest, as well as moments from multivariate distributions. The cross-country moment dataset will become available for research use. Secondly, we have developed a research infrastructure for the decentralized execution of program code on national micro-data in multiple countries. This infrastructure will allow custom research projects to tap into confidential firm-level data in multiple countries, and to extract and combine harmonized results for each country. The technical components of the infrastructure are in place, but the 'details' concerning funding of the use of the infrastructure have not been finalized.

The paper presents a broad range of highlights from our cross-country comparisons on adoption, use, and impact of ICT. Our results are novel in the sense that no cross-country evidence of our findings has been available previously. The examples are chosen not only provide evidence on current academic and policy questions relating to ICT, but also to showcase the types of analysis that can be done using the micro moments database and research infrastructure. Especially in the area of assessing the impact of the policy environment, or evaluating the effects of policy changes, having cross-country indicators of both firm-level responses and market outcomes is crucial for identification.

The paper further will provide technical documentation of the research infrastructure put in place by participating NSOs, from harmonization of national sources, through linking and aggregation methods, to safeguards for protecting confidentiality of individual firms. Next, the paper will document the definitions and methods for the cross-country micro-moments datasets that will be made available to qualified researchers. Finally, we will describe the opportunities for external researchers to design analytical modules to be run using the cross-country infrastructure.

The organization of the paper is as follows: We start with an overview of recent research on ICT, categorized by the types of data used. Next, a brief overview of our methodology is given. In section 4, highlights of the analysis done by the ESSLait project are presented. Section 5 presents the details of the research infrastructure and prospects for external

[2] Participating NSOs are: Austria, Denmark, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Norway, Poland, Slovenia, Sweden, United Kingdom. Earlier rounds of the project included the Czech Republic and Romania.

researchers. Section 6 presents the micro-moments database with details of variable definitions and methods. We conclude with some thoughts on remaining challenges.

2. Research on ICT and the Economy

It has now been about 25 years since economists started studying the impact of ICT on the economy or the economics of ICT in general (see e.g. Loveman, 1990). In that period, measurement of processor speeds has shifted from MHz to GHz, internal memory from MB to GB and data storage from MB to TB. A quarter century ago, communications generally went via the plain old telephone (POTS), although facsimile technology (FAX), which had been available for decades, was finally being adopted widely (Peterson, 1995). Since then, we have gone through a bewildering series of acronyms (IP, FTP, WAIS, WWW; ISDN, DSL, FTTH, CDMA, GSM, UMTS, EDGE, LTE, WiMax), with as result that in the developed countries nearly 80 percent of the population uses internet, and that globally nearly 100 percent of the population uses cellular phone (ITU, 2013). Finally, full convergence between information and communications is on its way, with 18 percent of global internet traffic via mobile devices and cloud facilities processing and storing information for remote users.

The potential impact of the ongoing technological advance on the economy seems staggering, and the economics of the activities underlying ICT appear quite different than that governing the old pin factory. Non-economists tend to think that none of the old rules apply (Kelly, 1999; Ford, 2009), but even economists can see that the effects may be far reaching (Shapiro and Varian, 1998; Brynjolfsson and McAfee, 2011). We will describe briefly some of the research that took place over the past 25 years on the economic effects of ICT over the past decades, and will touch on some of the new areas where theory guides us.

We will organize our brief overview of the literature according to the underlying data used. We start with studies using macro-level or industry-level data, with cross-country availability. Next, we will provide an overview of firm-level studies that delve deeper into underlying behavior of firm's adoption and use of ICT and its impact. Most of these studies are conducted on data for a single country.

To understand the impact of technological change on the economy, with a view to the medium and long run, the workhorse method remains growth accounting (e.g. Jorgenson 2008). With this method, the contribution of each factor of production to output is assumed to be proportional to the corresponding share in total input costs. If production increases more than the contribution attributed to measured production factors, this is ascribed to growth in multifactor productivity (MFP), i.e.: technological progress not embodied in production inputs.

Jorgenson and Stiroh (1999) use this method to explain the increase in productivity growth at the macro level in the U.S. after 1995. They find that computer capital played emerged as a source of economic growth and that the increase in labour productivity between 1995-1999 could be explained by computer capital-deepening that resulted from double-digit increases in price-quality of ICT equipment. Oliner and Sichel (2000), using a similar growth accounting model find that the contribution of ICT inputs rose sharply in the 1996-1999 period compared to 1974-1995, and that MFP growth surged 40% over the earlier period.

In the past 10-15 years, measurement of investment in ICT capital goods, and construction of appropriate deflators that take quality improvements into account, has been harmonized across countries (see e.g. Wyckoff, or Schreyer 2000). Using these data along with other measures from official statistics, Colecchia and Schreyer (2002) were able to provide growth accounting for nine OECD countries up to the year 2000. From this work, and others, the view came that the contribution of ICT to growth had accelerated all over, but that the US had higher ICT investment intensities and consequently higher GDP growth.

The growth accounting comparisons have been extended across industries thanks to the efforts in creating EUKLEMS, a harmonized cross-country and industry panel dataset of productive output and inputs in real and nominal terms (O'Mahoney, Timmer and Van Ark, 2007). Analytical work using this database has shown that the US growth was not generated only through rapid increases in productivity in ICT producing industries, but also because of slower uptake of ICT in the industries that use ICT. Especially gains in market services in the US are not matched in EU countries. The most recent results from growth accounting for the aggreagate US and EU economy continue to show that the EU is lagging in growth, mostly from lower uptake of ICT (van Welsum et al., 2013)

In order to understand what the incentives are to invest in ICT, and what the mechanisms are through which ICT use improves productivity, researchers have moved their attention to firm-level data. From the earliest papers showing the relationship between productivity and firm-level use of IT (Brynjolfsson and Hitt, 2000), work has become more adept at estimating the rate of return to investment in ICT capital (e.g. Crepon and Heckel, 2002). In case the return to ICT is higher than for other types of capital, there may be evidence for externalities, over and above the growth accounting contribution from ICT. Alternatively, the above-normal returns may be related to non-measured investments in other complementary inputs (e.g. Bresnahan, Brynjolffson and Hitt, 2002). Further firm-level work explores the effects of ICT on flexibility in other factors (Hempell and Zwick (2008)), the interaction between ICT and outsourcing or offshoring (Abramovsky and Griffith (2006) or Bartel, Lach and Sicherman (2005)), and the relationship between ICT and human capital. Finally, an important new area of research is in the role of management quality and organization choices in adoption of ICT and its impact (Bloom et al. 2012).

While this work is useful in disentangling correlates of firm-level adoption of ICT, or interaction of ICT with other factors in providing in impact on productivity, only having data for firms in a single location makes the identification of causal relationships difficult. In particular, by comparing and contrasting ICT adoption and impact across countries, one can disentangle the role played by the policy and market environment.

Figure 1 presents a micro-to-macro view linking firm-level decisions to interactions in a market to aggregate outcomes. As seen, the firm-level studies are able to look carefully at the drivers of firm-level decisions, providing estimates of production functions, factor input decisions, R&D spending, choice of technology. These decisions however take place in a market environment, entailing interactions with other firms in own-, downstream-, and upstream sectors, as well as an institutional and policy setting. The industry and aggregate studies look at the output and productivity outcomes, but usually assume that the data are generated by a representative firm acting under profit maximization. The mechanisms through which the market environment, including policy changes, affects firm-decisions and subsequently aggregate outcomes can not be traced. The main macro-level research on ICT and growth therefore is not able to identify reasons why ICT investment is lagging in some countries. Clearly, research being able to traverse from firm-level data to aggregate outcomes, through the allocation and selection mechanisms of the market, would be fruitful in gaining a better understanding of the policy drivers of growth.

Figure 1. From micro to macro

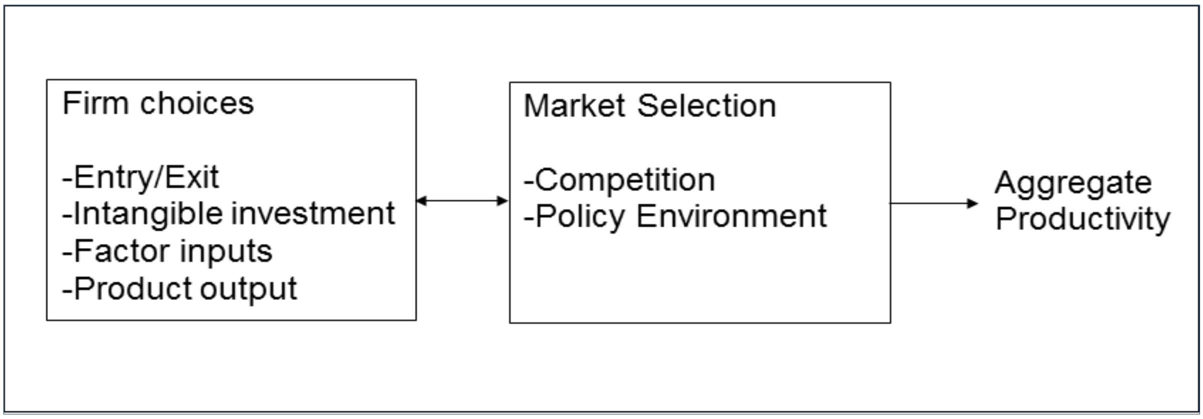
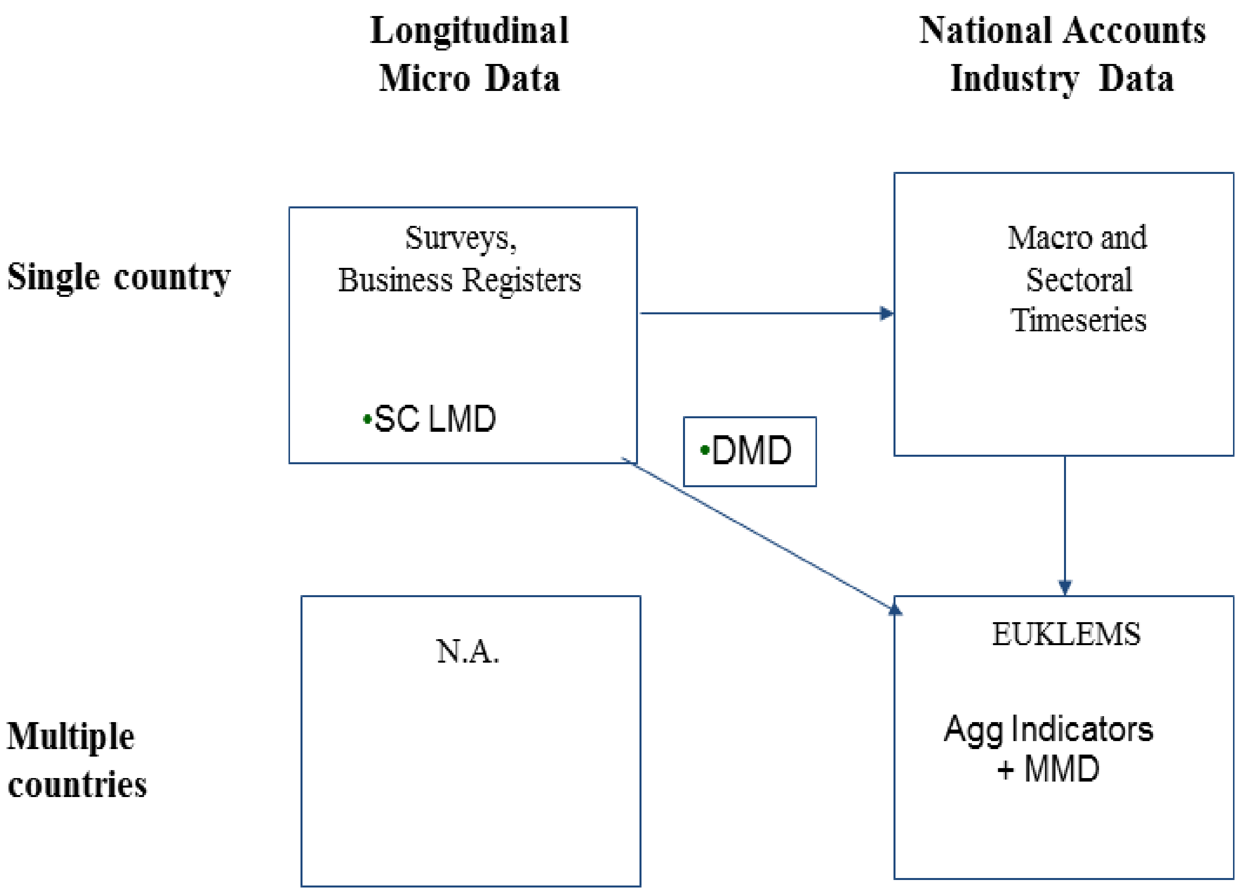


Figure 2 presents an overview of the types of data available for research on firm and industry behavior and outcomes. We discussed above the single-country firm-level studies, and the single country studies using industry data. For a single country, it has become possible to link linked-longitudinal firm-level surveys and study them in conjunction with time-series at the industry and macro level. Much of the work recent coming out of the Center for Economic Studies of the US Bureau of the Census (see http://ideas.repec.org/s/cen/wpaper.html), can draw aggregate conclusions on employment, output and productivity by tracing from the micro to the macro level (e.g. research on implications of computer use on wage and productivity dispersion, Dunne et al. 2004, or disentangling job creation over time, Haltiwanger et al. 2010). However, because the policy environment and institutions often do not vary within a country, it is difficult to identify the effects of changes in policy. The bottom right corner of figure two shows how cross-country industry data, such as that collected and disseminated through the EUKLEMS effort (www.euklems.org), are available. With these data, the drawback of not being able to trace mechanisms through firm decisions is the same as with single country firm-level data. The difficulty in analyzing cross-country firm-level data in conjunction with the corresponding industry and macro timeseries is what has led to the creation of the Distributed Micro Data approach, and the publication of the Micro Moments Database.

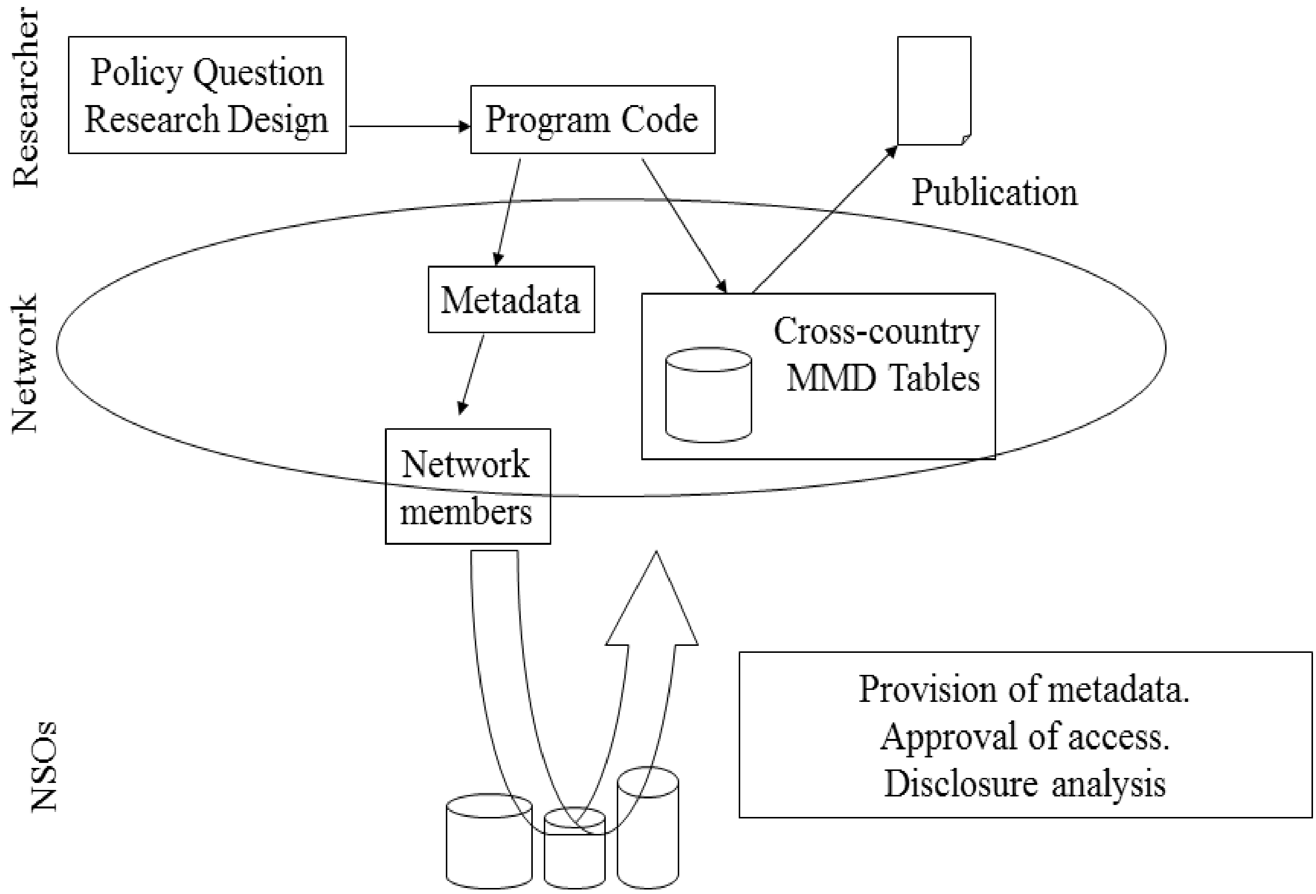
Figure 2. Data for research on productivity



3. Distributed Micro Data Approach

Cross-country comparisons of adoption and impact of ICT at the firm-level may provide the variation needed to disentangle the impacts of policy, environment, factor availability of other drivers. However, researchers in this area are usually faced with a serious dilemma. On the one hand, there are aggregate databases that allow the role of ICT to be investigated in a cross-country setting. While those data are useful for aggregate comparisons, ICT plays an important role in the dynamics of an industry, and its potential positive effect on performance is found to be conditional on additional investments at the firm-level, calling for a micro-oriented approach. But, setting up micro-data research in a multi-country setting is difficult and costly because most of the micro-level information that is collected by national statistical agencies is confidential. This means that the legal framework protecting the data does not allow for direct analysis on a merged cross-country dataset.

Figure 3. Distributed Micro Data Analysis



Within the ESSLait project (ESSnet on Linking of microdata to analyse ICT impact) we adopt the method of Distributed Micro Data Analysis (DMD) as developed by Bartelsman et al. (2004). In this approach, depicted in figure 3, a common protocol is used to extract micro-aggregated information from countries' harmonised firm-level datasets. This involves the assembly of micro-data by participating NSOs, and the running of common software to retrieve the indicators and statistical moments or to conduct statistical analyses. By proceeding in this way, a cross-country dataset containing indicators of underlying distributions and correlations, the MMD tables, can be made public without breaking national rules of confidentiality. Moreover, the infrastructure of NSO data and metadata allows external researchers to hook into the project software using their own analytical add-on modules. The output of these modules can either be stand-alone tables of (non-disclosive) cross-country results, or could generate indicators to augment the publically available cross-country industry dataset.

The first step undertaken in DMD is to query the participating statistical agencies about the availability of their firm level data, see the center of the bottom row of figure 3. Historically, statistical agencies have run firm-level surveys by drawing a sample from their list of all firms. In the early 1990s, Eurostat enacted regulations concerning the definition of the statistical units for business statistics, which recently has been superseded by regulation on a common framework for a business register. For our purposes, all countries already had in place from at least 2001 onwards, a business register containing the universe of enterprises.[3] From the business register (BR), NSOs conduct surveys on production related variables (the Structural Business Survey, or Production Survey--PS), and other topics of interest. For our project, we consider the Community Innovation Survey (CIS), and the ICT Usage Survey (or E-commerce Survey, EC). Evolving over time, some NSOs are reducing their reliance on surveys, and are collecting more and more variables of interest directly from register sources (tax authority, customs, educational registry) and linking the to the BR. In the ESSLait project we explore how the sampling strategies and evolution towards register data affect the ability to conduct longitudinal firm-level studies.[4]

For each year of each of the underlying sources, our project collected information on the unit of observation and how the sampling was conducted. Then we collected names and descriptions of each of the variables, along with the appropriate coding information (units or descriptions of categorical values). Through an iterative process among participants, we culled a list of variables that were comparable across countries and over time, and for which country coverage was reasonably high. This comprises our metadata, which is used by the common software to recode local NSO variables to the common set of project variables, and which can be consulted by outside researchers to design their own program modules (see the middle of center row in figure 3).

The DMD methodology thus allows access to confidential micro data, for the purpose of conducting cross-country analysis. The methodology provides relatively low marginal costs to conduct cross-country analysis, but requires a sizable up from set-up cost. The NSOs need to make an investment in organization of the firm-level data files and in providing the metadata. Next, NSOs need to have a framework to provide access to the data, they need facilities for running the program code and finally they need to conduct disclosure analysis of the output. In the course of the Eurostat-funded projects, much of the required investments in organizing the firm-level data and documenting the metadata have been made. More work remains to be done in providing employer-employee links, linking in trade data, and other sources such as location, transportation, energy use, etc. In countries moving away from separate surveys, and towards register based statistics, the links are part of the ongoing statistical processing. In other countries, and for more historical data, more investment is needed for the NSO to supply a full warehouse for DMD analysis.

The DMD methodology is not the only way to allow cross-country analysis of firm-level data. To start, much work has been conducted using commercially available sources, such as

---

[3] An **enterprise** is the smallest combination of legally recognized units, either constituting an organizational unit for producing goods or services, or benefiting from a certain degree of autonomy in decision making, especially for the allocation of its current resources. It may be a sole legal unit and carries out one or more activities at one or more locations. (See http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Enterprise).

4 Likely, given the direction NSOs are heading, it will become more difficult to find representative longitudinal panels to study the links between innovation, ICT and productivity at the firm level.

ORBIS from Bureau van Dijk. These data generally are sourced from Chamber of Commerce or mandatory filings of publically traded firms. However, the coverage and sources vary significantly across countries. Further, it is costly to combine these data with other firm-level indicators.

A European FP7 programme called 'Data Without Boundaries' has enhanced transnational access to official micro data through coordination of existing infrastructures and through funding of researchers. However, the researchers cannot 'stack' the data, but need to travel between NSOs to replicate their work in each country. Further, while the project aims at providing metadata to allow researchers to see what firm-level is available in the four participating NSOs, these are not yet delivered.

Another option is to generate public use versions of the data where individual firms have been made anonymous. Providing such synthetic data allows outside users to conduct analytical research, although validity of the analysis cannot be assured. In a recent project, the U.S. Census has created a 'synthetic LBD', or business register that not only allows outside users to run their own analysis, but allows validation of the results against the confidential LBD (http://www2.vrdc.cornell.edu/news/data/lbd-synthetic-data/). This method is promising for longitudinal datasets with many firms over time, but a relatively limited number of variables per firm. An increase in the number of covariates greatly complicates generating synthetic data that can match the expanding quantity of moments and joint moments.

Finally, remote execution of analysis at multiple NSO sites may provide an option for cross country firm-level data analysis. In fact, our research infrastructure provides a way to run analytical modules in multiple countries, but does not give 'real time' access, as remote access does. Instead, in our method, an analytical module first needs to be approved by a scientific board, then needs to be run in each country, with output checked for disclosure in each location. By contrast, one could imagine remote execution taking place through a data terminal in a secure 'remote' datacenter that has access to multiple NSO warehouses. Experiments of this method are taking place within Eurostat.

Section 5 of this paper will provide further details of the program code used for collecting micro moments indicators, and in particular the manner in which the firm-level datasets have been harmonized at each statistical agency. First, we turn to a summary overview of some of the statistical analysis that has been conducted during the Eurostat projects using the underlying firm-level data.

4. Examples of Statistical Analysis

In this section, we will present some descriptive statistics and highlights from analytical work done during the Eurostat funded projects. The purpose is to showcase some of the possible directions that researchers could take using the MMD or the research infrastructure, rather than to test a particular hypothesis regarding ICT and performance. The data underyling this work, at each NSO, includes the ICT usage survey data, Business Register data, data on various economic variables (captured mainly by Structural Business Statistics), data from the

Community Innovation Survey, data on international trade activities, and for a subgroup of countries also information on formal educational achievements. The project generally makes use of data for the period 2000-2010. Some countries also have made available a panel dataset of firm-level input and output data with a longer time-dimension (Long Panel).

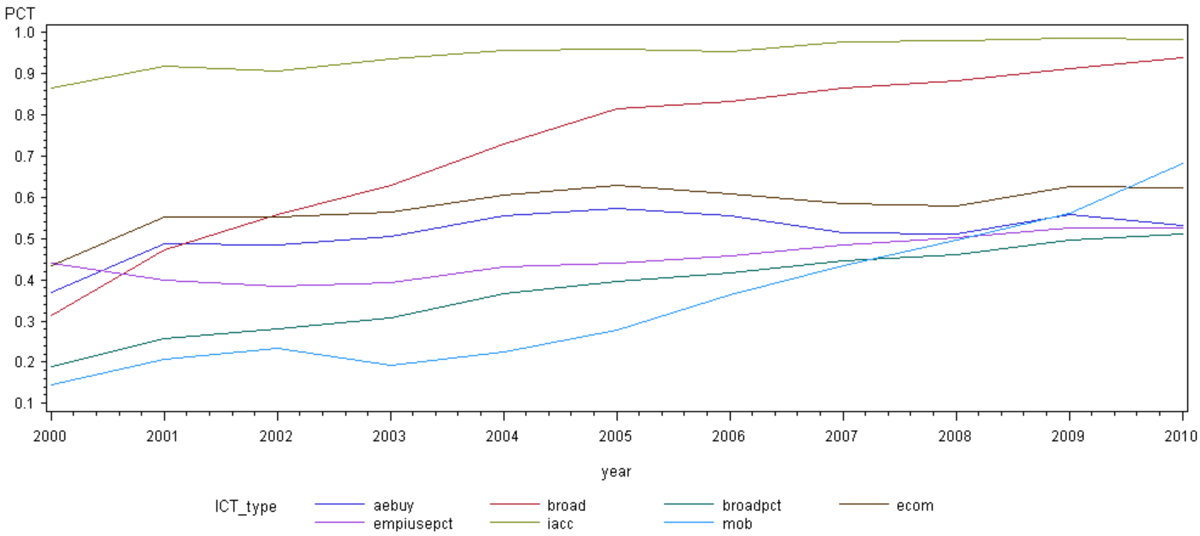
We will first show some stylized facts on ICT use, such as increasing adoption of certain technologies and persistent differences in uptake across countries. We also look at complementarities between ICT and innovative activity at the firm level. We follow with some descriptive statistics on productivity and how it varies between firms that use ICT or engage in innovative activities. Next, we explore some of the differences between weighted and unweighted summary statistics, and how these affect the interpretation of correlations between variables. Finally, we present some results from regressions on innovation using the MMD country/industry panel.

To start, we provide a short descriptive overview of some of the ICT related indicators, derived from the ICT usage survey and other linked sources. One of the variables that provided useful in analysis is the average percentage of workers in a firm with access to broadband enabled workers. This has been trending upwards, as can been seen in Figure 4. Internet use already was saturated in the beginning of the period, while broadband use is nearing saturation, while mobile usage at firms is currently increasing rapidly.

The analytical usefulness of the percentage of broadband enabled workers may come not only through the effects of broadband on the firm, but also because of the selectivity of firms that adopt technology early. As the variable nears saturation, it no longer distinguishes among firms. Therefore we developed an alternative indicator of ICT intensity, aiming to deal with the constantly changing nature of ICT usage. This indicator exploits the fact that we can run common modules on the firm-level data in all countries. We define a composite indicator where new forms of ICT usage can be added, and other forms can be dropped as they reach saturation. For this indicator, separate discrete choice models were estimated at the firm level for different ICT variables (e.g. adoption of ERP, CRM). The idea is that from these estimations, propensities of adoption can be calculated for each firm. As the value of adoption is expected to increase with the existing ICT intensity in a firm, the propensities should be correlated with ICT utilisation. Our suggestion is that a geometric average of the propensities can thus be used as a proxy for the firm-level ICT intensity.

Results for the ICT intensity indicator (ICTi) are reported in figure 5, where 'Intens' indicates whether ICTi > 60% or not. There are some surprises in the ranking based on ICTi, compared to other indicators of ICT usage (Slovenia rather high, Denmark and the Netherlands low). These surprises tend to be less for Intens, although Norway ranks relatively low, and Austria relatively high. In our research, we also find that the rankings are stable and show a slightly increasing pattern over time for most countries. It should be noted that we are in the process of experimenting with specifications for this indicator.

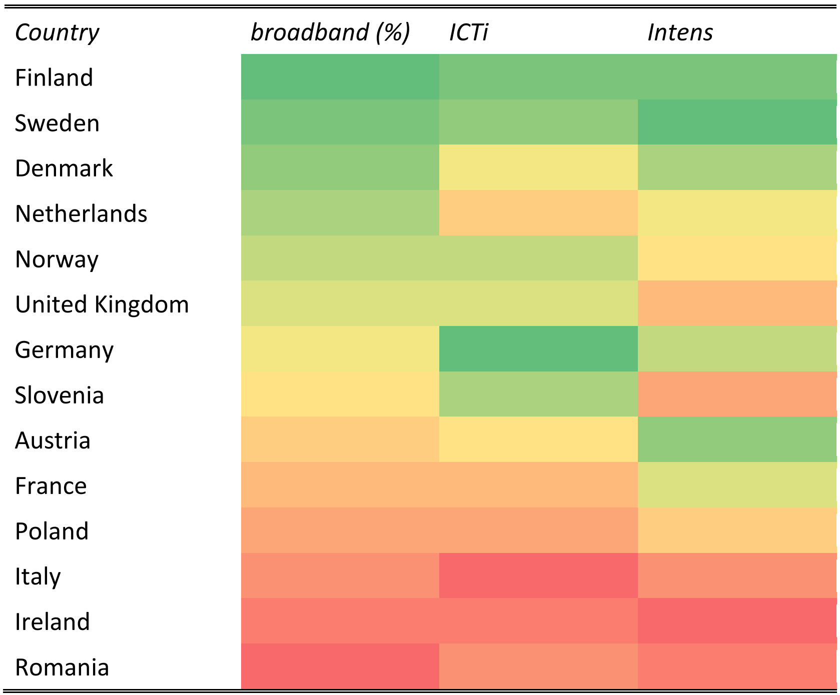
Figure 4. Evolution of ICT usage variables (total economy averages across countries).



Source: ESSLait Micro Moments Database

When looking at any of these indicators across countries and industries, it becomes clear that there is a persistent difference in uptake across countries. Using the MMD, a simple shift-share analysis could be done to parse out how much of the variation comes from differences in industry composition across countries, and how much from differences in country specific uptake.

Figure 5. Ranks of ICT intensity (ICTi and % of broadband enabled workers) by country (2009).



Source: ESSLait Micro Moments Database

The MMD allows one to explore correlates and causes of ICT adoption. The summary statistics tables contain aggregates for industries split by values of a variable from the other datasets, for example as shown in Table 1. Here, we look at the percentage of firms (size-weighted) in 2008 in Finland by broad industry categories that use customer relation software (CRM), split by whether or not they have innovative business practices (from the joint EC-CIS firm sample). Not very surprisingly, the use of CRM is significantly higher for firms that have organization innovation.

Table 1. Percentage of firms with CRM by innovative business practices (yes/no), in Finland (2008).

	% CRM (ORGIN = 0)	% CRM (ORGIN = 1)
MexElec	0.47	0.76
ConsG	0.23	0.72
IntmdG	0.50	0.79
InvesG	0.57	0.72
OtherG	0.51	0.74
Elecom	0.47	0.80
MServ	0,63	0.81
Distr	0.62	0.82
FinBu	0.69	0.78

Source: ESSLait Micro Moments Database

To further explore how different technologies are either complements or substitutes, we provide data on how many firms adopt each each separately, as well as jointly. For example, table 2 provides this data for the adoption of mobile internet and organization innovation in the Netherlands. These data are available by country, industry and year for a variety of combinations of ICT and innovation variables and allow one to track how combinations affect performance.

Table 2. Contingency table for A = mobile internet, B = organizational innovation, for the market services sector in the Netherlands in 2008.

		B		
		0	1	
A	0	711	222	933
	1	413	*234*	*647*
		1124	*456*	*1580*

Source: ESSLait Micro Moments Database

In the MMD, we have built up simple indicators of productivity from the firm-level data. We have labour productivity, defined as deflated sales or value added per worker, and we have Solow residual measures of TFP, which use country specific firm-level proxies for capital service inputs. Using these data, we run a simple panel data production function with fixed effects, with labour productivity as the dependent variable, and capital intensity, human capital intensity, and broadband enabled workers as explanatory variables.

$$\log(lab\_prod_{c,i,t}) = a + b \log(cap/emp_{c,i,t}) + c\ Broadpct_{c,i,t} + Fixed\_Effects(c,i,t) + Error_{c,i,t}$$

Table 3. Regression of labor productivity, MMD panel.
(t-stat in parenthesis. Fixed effects included)

Variable	Average	Weighted Average
BroadPct	0.14 (1.06)	0.67 (3.82)
Capital/Emp	0.10 (1.80)	0.07 (2.73)
R^2	0.13	0.32
Num Obs	500	500

Source: ESSLait Micro Moments Database and own calculations

The column 'Average', denotes that we use the industry average of labor productivity and the explanatory variables, while 'Weighted Average' denotes an industry aggregate, ie weighted average using firm size as a weight. Interestingly, on average broadband does not have a significant effect, which corroborates on average the findings from firm-level regressions in the separate countries. The significant effect in the weighted average regressions indicates that the more productive firms that use broadband intensively are larger. A simple statistical possibility for this result would be a situation where the average productivity of firms does not increase with broadband intensity, but the variance of productivity does. Then, if the firms with high productivity are larger, aggregate productivity will be positively correlated with broadband intensity.

We explore the first part of the story, by seeing how broadband intensity and the variance of productivity relate. Using the MMD data, we have for each country, industry, and year a measure of the cross-sectional variance of firm-level productivity. We can then regress the standard deviation of productivity on broadband and on fixed effects, as shown in table 4.

Table 4. Standard deviation of firm-level productivity distribution regressed on broadband intensity

	Levels	First-differences
γ	0.47	0.28
t-stat	(5.02)	(2.59)
R^2	0.52	0.03
D.F.	1180	1021
Fixed effects	country, industry, time	country, industry, time

Note: Coefficients γ from a regression: $\sigma_{c,i,t} = \alpha + \gamma BBI_{c,i,t} + FE + \varepsilon_{c,i,t}$, with country, industry, time fixed effects, where σ is the standard deviation of the cross-sectional distribution of labour productivity in country c, industry i and time t, and BBI is the broadband intensity. The regression is run in levels and in first differences. D.F. is short for degrees of freedom.

Source: ESSLait Micro Moments Database and own calculations

Similar evidence can be derived from the 'long panel' tables of the MMD. Firms intensive in ICT have shown to be more volatile, although the causality is not completely clear. Either firms willing to take larger risks may be more prone to invest also in advanced ICT systems, or firms facing a more volatile environment may invest in ICTs to reduce adjustment costs. In the 'long-panels' of firm-level production statistics for the period 1995-2010, we classify firms as high or low in ICT, depending on their estimated ICT intensity in the period 2001-2010 and then we look at the time series and cross-sectional distribution of employment and output growth for these two groups of firms.

As seen in Table 5, the dispersion measures are higher for ICT intensive firms (ICT=1), except in France. In the first two columns the average standard deviation of the firm-level time series of labour productivity growth is presented. This is measured at the firm-level using a 5-year moving window. The firm-level dispersion is averaged into an industry series, using firm-size weights. Finally, the industry dispersion is averaged over the period 2003-2007 (thus using underlying firm-level data from 2001-2009). We also have dispersion measures from the firm-level cross sectional distribution, both for productivity growth and output growth. In computing the dispersion, observations are weighted by firm size. In this exercise, the project did not collect the measure for non-ICT intensive firms, only for ICT-intensive firms and the industry as a whole. In all countries, except France, the ICT intensive firms have a higher standard deviation of the cross-sectional distribution of firm-level output and productivity growth.

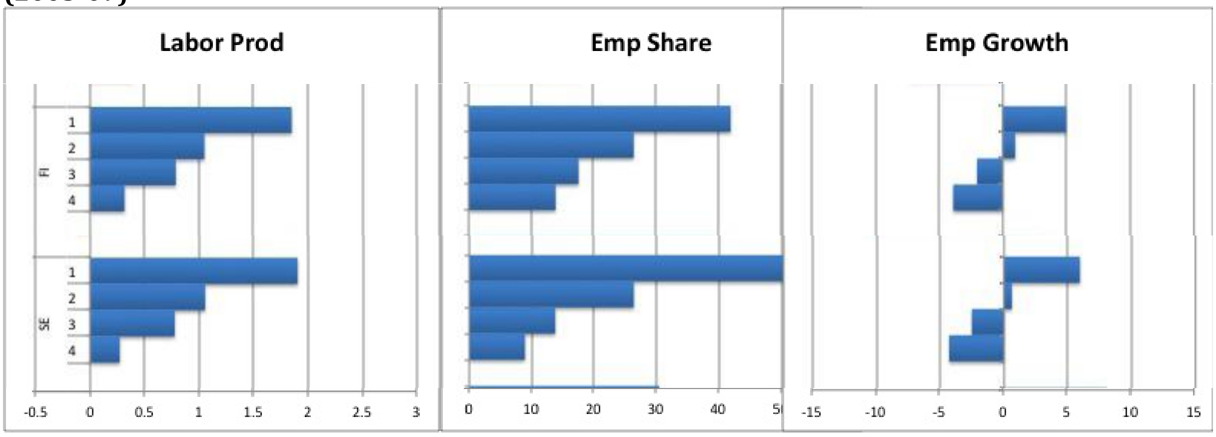
Table 5. Standard deviation of productivity growth (ΔLP) and output growth (Δq).

	Time series				Cross section			
	Productivity growth		Output growth		Productivity growth		Output growth	
Country	ICT=0	ICT=1	ICT=0	ICT=1	ALL	ICT=1	ALL	ICT=1
DK	0.19	0.21	0.24	0.26	0.23	0.24	0.29	0.32
FI	0.19	0.28	0.21	0.31	0.25	0.27	0.3	0.33
FR	0.2	0.18	0.22	0.18	0.21	0.18	0.21	0.19
NL	0.13	0.14	0.11	0.13	0.22	0.24	0.2	0.21
NO	0.18	0.26	0.21	0.29	0.32	0.35	0.33	0.35
SE	0.2	0.26	0.32	0.38	0.33	0.37	0.49	0.52

Note: The table shows the averaged standard deviation of labour productivity and output growth. In the time series columns, the standard deviation of growth is measured at the firm level for a 5-year moving window and averaged across ICT intensive and non-intensive firms in the industry (ICT=1 and ICT=0). The industry and time dispersion measures are then averaged over time and across industries with fixed industry weights. In the columns labelled cross section, the standard deviation of growth for the cross-section of firms in an industry is computed, for ICT intensive firms (ICTi=1) and for all firms (ALL). The industry and time dispersion measures are then averaged over time and across industries with fixed industry weights.
Source: ESSLait Micro Moments Database

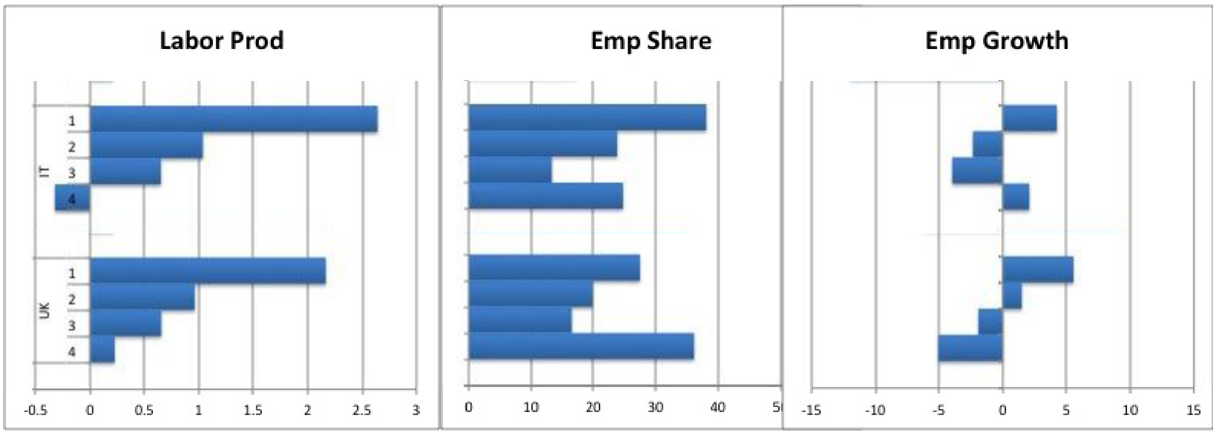
The above has shown that ICT intensity and variance of firm outcomes are correlated. The next part of the story to see why average firm-level impact of ICT may be lower than the aggregate impact, is to see how allocation of resources is related to productivity. The literature of reallocation (e.g. Hsieh and Klenow 2009, or Bartelsman et al. 2013) discusses theoretical arguments for linking productivity variance, or the covariance between productivity and firm size to aggregate productivity. In Figures 6 and 7 we show distributions of productivity, and distributions of employment and employment growth conditional on productivity. The information in the chart is available for all countries and sectors, but here we show information for Finland and Sweden, averaged across industries of the manufacturing sector (excl ICT) and averaged for the period 2003-2009. The first column shows the average labor productivity (value added per worker) of firms in the highest quartile of the productivity distribution relative to the average productivity. In Sweden, the best firms are nearly twice as productivity as the average firm, while firms in the lowest quartile are only about a quarter as productivity as the average firm. The next column shows the distribution of employment across productivity quartiles. In Sweden half of the employment in the sector is in the top quartile of firms. Clearly, the OP-gap measure shown above is related to the width of the distribution of productivity and the covariance with the employment distribution. Finally, the last column shows employment growth of firms by lagged productivity quartile. Here as well, a monotonic relationship with productivity by quartile is good for productivity growth. Next, in Figure 7, we show similar results for Italy and the United Kingdom, where the allocation is not as good for aggregate productivity. In the UK, we see high employment shares in low productivity firms, and in Italy we see that the employment growth is not monotonic over quartiles.

Figure 6. Heterogeneity and resource allocation, Finland and Sweden, Manufacturing excl ICT (2003-09)



Source: ESSLait Micro Moments Database and own calculations

Figure 7. Heterogeneity and resource allocation, Italy and United Kingdom, Manufacturing excl ICT (2003-09)



Source: ESSLait Micro Moments Database and own calculations

We can put the findings together in a set of reduced form regressions using the MMD that is inspired by Crepon et al (1998). First we can link, on average, the effect of ICT use on the probability that a firm innovates. Next, we can find how the predicted probability of innovating and the efficiency of resource allocation affects productivity. In the Crepon model, one needs to first estimate the probability that a firm engages in innovative activity. In industry data, this information is lost in aggregation, and one usually only has information on the share of innovating firms. In our MMD data, we have generated aggregates by industry for firms that innovate and firms that do not. Using the panel dimensions, we can thus estimate a probit equation for the probability of a product innovation, with as explanatory variables the average characteristics of a firm, in this case related to their ICT use. In this equation, all variables are measured as averages across firms. In the productivity equation, by contrast, we use aggregate (or size weighted) averages for productivity and the explanatory variables. Included as an explanatory variable is the covariance between productivity and size, or the Olley-Pakes gap, as described in Bartelsman et al. (2013).

Table 6. Innovation and Productivity Model

	BROADpct	Mobile	Web	E-buy	E-Sell
P(Innov=1)	+	+	+		+

	Pred(Innov)	Human Capital	Capital/labor	OP-Gap	Num Obs
Labor Productivity	**0.13**(3.3)	**0.85**(4.9)	**0.12**(6.1)	**0.32** (6.6)	457

Note: A "+" means positive significant relationship at the 1 per cent level, "-" negative and a blank cell insignificant result. Country, Industry, and Time Fixed Effects included in both equations.
Source: ESSLait Micro Moments Database and own calculations

In this simple exercise, we have made use of the ICT variables as being explanatory for innovative activity of a firm. To use the ICT variables as instruments, we are assuming that they are not themselves caused by the unobserved productivity shock in the productivity equation. Further explorations, using dynamic panel specifications are needed to improve identification of the above model.

5. Research Infrastructure

In this section we describe the technical details of the research infrastructure that has been built at the NSOs during the Eurostat funded data linking projects. In particular, the section will be organized to aid a potential researcher in finding out whether the underlying data can be used to answer a particular research question and how to design the project to make use of the research infrastructure. The research infrastructure will be set up so that external researchers can write code modules that make use of a network of NSO partners. The NSO partners will run the code modules in their country and conduct disclosure analysis of the output. A tentative, non-exhaustive list of potential research areas that can be addressed with the underlying firm-level data include e-business systems, export behaviour, employment dynamics, reallocation, ICT skills and outsourcing, innovation, intangible or knowledge based capital investments, adjustment costs, competition, and financial frictions. At present, the research infrastructure only is accessible to external projects on an experimental basis. For now, modules have been run for the OECD, generating output of their 'DynEmp Express' code, and for the EU-IPTS, generating regression coefficients on the firm-level effects of R&D subsidies. Eurostat is currently working on the institutional and financial details of making the infrastructure available for qualified academics.

Table A3.1 in the appendix shows which years of each of the underlying datasets are available in each of the participating NSOs. The available datasets are called the Business Register (BR), Production Survey (PS), E-commerce Survey (EC), and Community Innovation Survey (IS). At each NSO, the BR is available as a panel, by unique firm identifier and year, while the other files are available in annual cross sections, with the firm identifier as the unique key. Table A3.1 shows a selection of availability of the PS variables, which record firm characteristics and values of productive inputs and output. In the Appendix 2, the list of available variables in each dataset is provided, as well as industry coverage.

The datasets BR, PS, EC, and IS are designed to be as similar as possible to the underlying surveys at the NSO, but considerable harmonization had to take place to make each underlying data set comparable over time and across countries. Each NSO may use their own variable names, but the program code has a translation from the name at the NSO to the variable name available for the common code or for external research modules. Further, the units or response categories for each variable have been harmonized across countries. Difficulties occur with harmonization of the coding of responses following negative responses to 'filter questions' and other types missing values, where NSO have used different strategies.[5]

[5] The missing values following negative filter questions have been recoded to zero, but in some NSOs other types of missing values also had been coded as zero in early years. This makes consistent treatment of missing

The metadata, mapping locations and names of datasets, as well as variable names, allows common code to be run at each site. Two other generic problems have been harmonized across countries. To start, we have harmonized definitions of industry, size classes, and age classes of firms. Each country had to map the industry code in each firm-year observation to our common industry list. The common list we use is based on Nace 1.1, and allows mapping out MMD output to the EUKLEMS dataset at the 30 industry level. Further, we generated 6 more aggregate sectoral groupings that map into the EUKLEMS 'alternative' hierarchy. The major purpose of the alternate hierarchy is to generate ICT producing sector, 'Elecom', that includes detailed industries from both the manufacturing and service aggregates. Table A1.1 and A1.2 gives details. A difficult arises around 2008, when NSOs switch from Nace 1.1 to Nace 2. Our common code includes a method to use the firm-level data, and a possible overlap of the two industry codings in a single year, to generate a concordance file, which is used to convert firms from their Nace 2 code to an appropriate Nace 1.1 codes.

The next bit of harmonization concerns deflation. The Eurostat network has compiled industry-level deflator timeseries for each country for output, value added and intermediate inputs. These have been sourced from the EUKLEMS project and updated to 2010 on the Nace 1.1 basis using available national price series. The program code reads in the common deflator information for each countries and applies them in a uniform manner to the firm-level data.

The common code next reads in the BR and converts any Nace 2 industry codes to the Nace 1.1 classification, using the harmonized concordance method. After this, the common code reads the underlying PS, EC, and IS datasets, renames all the variables to the common project nomenclature and links the datasets to the BR. At this point, firm-level sample weights are generated, using a re-weighting algorithm that compares firms available in a linked dataset with the universe of firms in the BR.

At this point the data are ready to be used both by the project code that creates the MMD, but also by our own analytical modules, or modules written by external users. In each country, various linked datasets are available for the modules to use, namely PSEC, PSIS, and PSECIS, which respectively are the union of firms in the PS and EC in each year, the PSIS, and all three datasets. The external researcher can make use of the information from the sample reweighting, and can call a generic outlier detection/exclusion program for their analysis.

The output of a code module can consist of output datasets, aggregated to industry or other firm-level characteristics to avoid breaking confidentiality, or tables of analytical results (ie regression coefficients and diagnostics). The heterogeneous nature of the output shows why, at present, the external code cannot be submitted via remote execution. The original research design needs to be evaluated for compliance with statistical purposes, which may vary across countries. Next, the output needs to be checked for disclosure. An intermediate form of work

values across countries and over time difficult: our choice has been to make missing Boolean values equal to false if a firm has any true values for other variables in that year.

may be possible, where researchers can collect combined national output in a secure data center, with only the final output being subject to disclosure review.

Writing modules to run on the infrastructure is straightforward. At present, the code is in SAS. The user needs to choose which one or more of the available linked datasets to use, and must refer to the available variables in each dataset. Selection and filtering of the data must take place via the code, iterations between coding and viewing intermediate results is not possible. Finally, the code should where possible combine results into datasets to facilitate combining output from the different countries.

The following section will describe the construction of the micro moments database, MMD and provide an overview for users interested in using the database for research.

6. MMD Summary

The section describes the cross-country tables that comprise the MMD, and provides a brief summary of the methodology used to construct the tables. Detailed contents of each table is provided in Appendix 2. The first set of tables provides information about the underlying datasets, next come summary statistics of each underlying 'survey', PS, EC, and IS, then a set of firm demographics files based on the BR, a set of tables with detailed information on distributions and joint-distributions of variables, and finally a table with industry dynamics indicators.

A dataset called 'metadata', describes the mapping from the underlying firm-level datasets at NSOs to the variable names used in the Eurostat projects. The 'coverage' table provides, by country, information about the industry coverage and the shares of firms and employment of the total industry in each of the surveys. The 'panel' and 'attrition' datasets provides some information on the panel characteristics of the PS and EC surveys in each country. Finally, the 'concordance' file provides for each country the mapping from Nace1.1 to Nace2 constructing using the firm-level data (BR) in each country.

A main part of the MMD is made up of the 'stat' tables, that provide summary statistics of the variables in the PS, EC, and IS surveys, by country, industry, and year. The variables for which summary statistics are generated include the main survey concepts, usually numeric variables, or Booleans, but also derived variables such as productivity or ICT intensity. Not only are the summary statistics created for all the firms in the survey, but also for sum-samples, such as PSEC or PSECIS. Further, the table splits industries into sub-groups, such as size-class, age, or multinational status, ICT intensity, export status, etc, so that users can compare panels of country-industry-time panels built up from firms that are ICT intensive or not, or that export or not. Associated with the 'stat' tables are the 'jointstat' tables, that look at shares of firms in an industry that jointly use two 'technologies' or innovation styles from the EC or IS, such as mobile internet and process innovation. Together with shares of firms using each technology separately, statements can be made up complementarity of the technologies.

The firm-demographics tables provide information on firm entry and exit as well as on gross job creation and destruction by incumbents or through entry and exit. This information is available by country, industry, and year, as well as split by size class, by age, or by size and age. This type of information now is included in Eurostat statistical releases on firm demographics, which should be favoured for official purposes. The data on firm demographics provided in the MMD should be seen as a complement to the other MMD datasets as they are constructed in similar manner across countries and can be readily matched to the other tables.

There is a table on industry dynamics, that collects indicators related to re(allocation) of resources and competition. The indicators include productivity decompositions, covariance of productivity and size, measures of market share churn, and simple versions of the Boone (2008) profit-elasticity measure of competition. The indicators are available by country, industry and time.

The tables with distributional information, the ‘st’ files, provide means and standard deviations of variables for each country, industry, year, but also by quartile of the distribution within each industry. The ‘cr’ files provide joint distributions of two variables by providing means and standard deviations of a variable by quartile group of the other variable. For each country, industry, year, there are thus 10 moments for each variable, which should be sufficient to provide ways to back out the best fitting family of distributions with estimates of its parameters for the underlying firm level distribution.

To start, the MDD tables will be available for researchers through Eurostat. A public release version of the MMD will be published in internet, and will include a selection of indicators for more aggregate industries from the EUKLEMS alternate hierarchy. Use of the data is for research purposes only, and should include a reference to this paper.

7. Conclusions and thoughts for future work

The paper has provided an overview of the MMD and research infrastructure that has been created by a network of National Statistical Offices funder by Eurostat. After a brief summary of research on economic aspects of ICT, and going through some of the statistical results and examples of simple analytical exercises, the paper discusses the how external researchers can access the infrastructure. Further, the paper provides documentation of the MMD that will be available for researchers through Eurostat.

The efforts have led to some analytical results on economic causes and effects of ICT. Much work remains to be done on this topic, especially in understanding why the EU is lagging behind the US in production and use of ICT by firms. The latter will require continued policy effort in the EU, and we hope to have contributed bits of evidence to support this policy effort. More importantly, we hope that availability of the MMD and of the research infrastructure will attract the academic community to take up the challenge to improve our understanding the links between policy, ICT, and economic performance.

In the future, we will strive to expand the datasets available in the research infrastructure, by adding linked employer-employee data, detailed import and export data, as well as more detailed financial information, including links between firms and their lenders. The latter may turn out to be a key in understanding whether the current financial crisis in limiting the ability of firms to invest in risky, ICT-related, project.

References

Abramovsky, L., & Griffith, R. (2006). Outsourcing and Offshoring of Business Services: How Important Is Ict? *Journal of the European Economic Association*, *4*(2-3), 594–601. doi:10.1162/jeea.2006.4.2-3.594

Acemoglu, D., Akcigit, U., Bloom, N., & Kerr, W. (2013). *Innovation, Reallocation And Growth* (Working Paper No. 13-23). Center for Economic Studies, U.S. Census Bureau. Retrieved from http://ideas.repec.org/p/cen/wpaper/13-23.html

Ackerberg, D., Lanier Benkard, C., Berry, S., & Pakes, A. (2007). Chapter 63 Econometric Tools for Analyzing Market Outcomes (Vol. Volume 6, Part 1, pp. 4171–4276). Elsevier. Retrieved from http://www.sciencedirect.com/science/article/B7GX7-4R9GYJT-6/2/2fb39f963628ef9a1b169210160591ab

Akcigit, U., & Kerr, W. R. (2010). Growth Through Heterogeneous Innovations. *National Bureau of Economic Research Working Paper Series*, *No. 16443*. Retrieved from http://www.nber.org/papers/w16443

Andrews, D., & Cingano, F. (2012). *Public Policy and Resource Allocation: Evidence from Firms in OECD Countries* (OECD Economics Department Working Paper No. 996). OECD Publishing. Retrieved from http://ideas.repec.org/p/oec/ecoaaa/996-en.html

Andrews, D., & Serres, A. de. (2012). Intangible Assets, Resource Allocation and Growth: A Framework for Analysis. *OECD Economics Department Working Papers*. Retrieved from http://ideas.repec.org/p/oec/ecoaaa/989-en.html

Arkolakis, C. (2009). A Unified Theory of Firm Selection and Growth. *SSRN eLibrary*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1427119

Bartel, A. P., Lach, S., & Sicherman, N. (2012). Technological Change and the Make-or-Buy Decision. *Journal of Law, Economics, and Organization*, ews035. doi:10.1093/jleo/ews035

Bartelsman, E., Gautier, P., & De Wind, J. (2010). *Employment Protection, Technology Choice, and Worker Allocation* (IZA Discussion Papers No. 4895). Institute for the Study of Labor (IZA). Retrieved from http://econpapers.repec.org/RePEc:iza:izadps:dp4895

Bartelsman, E., Haltiwanger, J., & Scarpetta, S. (2010). Cross-country and within-country differences in the business climate. *International Journal of Industrial Organization*, *28*(4), 368–371. doi:10.1016/j.ijindorg.2010.03.007

Bartelsman, E. J. (2010). Searching for the sources of productivity from macro to micro and back. *Industrial and Corporate Change*, *19*(6), 1891 –1917. doi:10.1093/icc/dtq059

Bartelsman, E. J., & Doms, M. (2000). Understanding Productivity: Lessons from Longitudinal Microdata. *Journal of Economic Literature*, *38*(3), 569–594.

Bartelsman, E. J., Haltiwanger, J. C., & Scarpetta, S. (2009). Measuring and Analyzing Cross-Country Differences in Firm Dynamics. In *Producer Dynamics: New Evidence from Micro Data* (NBER/Chicago., Vol. 68, pp. 15–82).

Bartelsman, E. J., Haltiwanger, J. C., & Scarpetta, S. (2013). Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*, *103*(1), 305–334. doi:http://dx.doi.org/10.1257/aer.103.1.305

Bartelsman, E. J., & Wolf, Z. (2009). Forecasting Aggregate Productivity Using Information from Firm-Level Data. *SSRN Electronic Journal*. doi:10.2139/ssrn.1404564

Black, S. E., & Lynch, L. M. (2001). How to Compete: The Impact of Workplace Practices and Information Technology on Productivity. *Review of Economics and Statistics*, *83*(3), 434–445. doi:10.1162/00346530152480081

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Saporta-Eksten, I., & Reenen, J. V. (2013). *Management in America* (Working Paper No. 13-01). Center for Economic Studies, U.S. Census Bureau. Retrieved from http://ideas.repec.org/p/cen/wpaper/13-01.html

Bloom, N., Garicano, L., Sadun, R., & Reenen, J. V. (2009). *The distinct effects of Information Technology and Communication Technology on firm organization* (Working Paper No. 14975). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14975

Bloom, N., Genakos, C., Sadun, R., & Reenen, J. V. (2012). *Management Practices Across Firms and Countries* (Working Paper No. 17850). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17850

Bloom, N., Sadun, R., & Reenen, J. V. (2012). Americans Do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review*, *102*(1), 167–201. doi:10.1257/aer.102.1.167

Boone, J. (2008). A New Way to Measure Competition. *The Economic Journal*, *118*(531), 1245–1261.

Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics*, *117*(1), 339–376. doi:10.1162/003355302753399526

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economic Perspectives*, *14*(4), 23–48.

Brynjolfsson, E., & McAfee, A. (2011). *Race Against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Digital Frontier Press.

Colecchia, A., & Schreyer, P. (2002). ICT Investment and Economic Growth in the 1990s: Is the United States a Unique Case?: A Comparative Study of Nine OECD Countries. *Review of Economic Dynamics*, *5*(2), 408–442. doi:10.1006/redy.2002.0170

Comin, D., & Mulani, S. (2009). A theory of growth and volatility at the aggregate and firm level. *Journal of Monetary Economics*, *56*(8), 1023–1042. doi:10.1016/j.jmoneco.2009.10.004

Crepon, B., Duguet, E., & Mairessec, J. (1998). Research, Innovation And Productivi[Ty: An Econometric Analysis At The Firm Level. *Economics of Innovation and New Technology*, *7*(2), 115–158. doi:10.1080/10438599800000031

Crépon, B., & Heckel, T. (2002). Computerization in France: An Evaluation Based on Individual Company Data. *Review of Income and Wealth*, *48*(1), 77–98. doi:10.1111/1475-4991.00041

Draca, M., Sadun, R., & Van Reenen, J. (2006, August). Productivity and ICT: a review of the evidence. Monograph. Retrieved September 11, 2013, from http://cep.lse.ac.uk

Dunne, T., Foster, L., Haltiwanger, J., & Troske, K. R. (2004). Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment. *Journal of Labor Economics*, *22*(2), 397–429. doi:10.1086/jole.2004.22.issue-2

Eurostat. (2008). *Information society: ICT impact assessment by linking data from different sources*. Luxemburg: Eurostat.

Ford, M. (2009). *The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future*. CreateSpace Independent Publishing Platform.

Garicano, L., & Heaton, P. (2010). Information Technology, Organization, and Productivity in the Public Sector: Evidence from Police Departments. *Journal of Labor Economics*, *28*(1), 167–201.

Garicano, L., Lelarge, C., & Van Reenen, J. M. (2012). *Firm Size Distortions and the Productivity Distribution: Evidence from France* (CEP Discussion Paper). Centre for Economic Performance, LSE. Retrieved from http://econpapers.repec.org/paper/cepcepdps/dp1128.htm

Gouriéroux, C., Phillips, P. C. B., & Yu, J. (2010). Indirect inference for dynamic panel models. *Journal of Econometrics*, *157*(1), 68–77. doi:10.1016/j.jeconom.2009.10.024

Haltiwanger, J. C., Jarmin, R. S., & Miranda, J. (2010). *Who Creates Jobs? Small vs. Large vs. Young* (Working Paper No. 16300). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w16300

Harrison, R., Jaumandreu, J., Mairesse, J., & Peters, B. (2008). Does Innovation Stimulate Employment? A Firm-Level Analysis Using Comparable Micro-Data from Four European Countries. *National Bureau of Economic Research Working Paper Series*, *No. 14216*. Retrieved from http://www.nber.org/papers/w14216

Hempell, T., & Zwick, T. (2008). New Technology, Work Organisation, and Innovation. *Economics of Innovation and New Technology*, *17*(4), 331–354. doi:10.1080/10438590701279649

Hopenhayn, H. A. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, *60*(5, pag. 1127-1150 .), 1127–1150.

Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, *124*(4), 1403–1448. doi:10.1162/qjec.2009.124.4.1403

Ichniowski, C., Shaw, K., & Prennushi, G. (1997). The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. *The American Economic Review*, *87*(3), 291–313. doi:10.2307/2951347

Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2008). A Retrospective Look at the U.S. Productivity Growth Resurgence. *Journal of Economic Perspectives*, *22*(1), 3–24. doi:10.1257/jep.22.1.3

Jorgenson, D. W., & Stiroh, K. J. (1999). Information Technology and Growth. *The American Economic Review*, *89*(2), 109–115. doi:10.2307/117090

Kelly, K. (1999). *New rules for the new economy: 10 radical strategies for a connected world*. New York, N.Y., U.S.A.: Penguin Books.

Lentz, R., & Mortensen, D. T. (2005). Productivity Growth and Worker Reallocation*. *International Economic Review*, *46*(3), 731–749.

Li, T. (2010). Indirect inference in structural econometric models. *Journal of Econometrics*, *157*(1), 120–128. doi:10.1016/j.jeconom.2009.10.027

Loveman, G. (1990). *An assessment of the productivity impact of information technologies*. MIT.

O'Mahony, M., Timmer, M. P., & Ark, B. V. (2007). *EU KLEMS Growth and Productivity Accounts: An Overview* (p. 19). Retrieved from http://www.euklems.net/data/overview_07i.pdf

Oliner, S. D., & Sichel, D. E. (2000). The Resurgence of Growth in the Late 1990s: Is Information Technology the Story? *The Journal of Economic Perspectives*, *14*(4), 3–22. doi:10.2307/2647073

Peterson, M. J. (1995). The emergence of a mass market for fax machines. *Technology in Society*, *17*(4), 469–482. doi:10.1016/0160-791X(95)00021-I

Pilat, D. (2004). The ICT Productivity Paradox: Insights from Micro Data. *OECD Economic Studies*, *2004*(1), 37–65.

Poschke, M. (2009). Employment protection, firm selection, and growth. *Journal of Monetary Economics*, *56*(8), 1074–1085. doi:10.1016/j.jmoneco.2009.10.013

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, *49*(2), 326–365. doi:10.1257/jel.49.2.326

Van Reenen, J., Bloom, N., Draca, M., Kretschmer, T., Sadun, R., Overman, H., & Schankerman, M. (2010). The economic impact of ICT. *Final Report*.

Van Welsum, D., Overmeer, W., & Van Ark, B. (2013, November). Avoiding the Digital Desert. The Conference Board.

Weintraub, G. Y., Benkard, C. L., & Roy, B. V. (2010). Industry Dynamics: Foundations For Models with an Infinite Number of Firms. *National Bureau of Economic Research Working Paper Series*, *No. 16286*. Retrieved from http://www.nber.org/papers/w16286

Annex 1. Industry Classifications (based on Nace 1.1)

Table A1.1. EUK Industry definitions

TOT	Total Economy
15t37	Manufacturing
15a6	Food, beverages and tobacco
17t9	Clothing
20	Wood and of wood and cork
21a2	Pulp, paper, publishing
21	Pulp, paper and paper
22	Publishing and printing
23t25	Refining, chemicals, and rubber
23a4	Refining and chemicals
25	Rubber and plastics
26	Other non-metallic mineral
27a8	Metals and machinery
27	Basic metals
28	Fabricated metal
29t33	Machinery and equipment
29	Machinery, nec
30t3	Equipment
30a3	Office, accounting and computing machinery; sc. eqpt.
31	Electrical equipment
32	Electronic equipment
34a5	Motor vehicles and transport equipment
34	Motor vehicles, trailers and semi-trailers
35	Transport equipment
36a7	Misc manufacturing
40a1	Electricity, gas and water supply
45	Construction
50t74	Market services
50t5	Trade, hotels, restaurants
50t2	Trade, hotels, restaurants
50	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of fuel
51	Wholesale trade and commission trade, except of motor vehicles and motorcycles
52	Retail trade, except of motor vehicles and motorcycles; repair of household goods
55	Hotels and restaurants
60t4	Transport and communications
60t3	Transport
64	Post and telecommunications
65t7	Banking
70t4	Real estate and bus services
70	Real estate activities
71t4	Business services
71a4	Renting of machinery and equipment; oth. bus. svc.
72	Computer and related activities
73	Research and development
75t99	Social services
75	Public admin and defence; compulsory social security
80	Education
85	Health and social work
90t3	Personal services
90t3x	Personal services excl. media
921t2	Media activities

Table A1.2. ALT Industry definitions (EUKLEMS 'alternate' sectors)

ALT	Description
Elecom	ELECTRICAL MACHINERY, POST AND COMMUNICATION SERVICES
MexElec	TOTAL MANUFACTURING, EXCLUDING ELECTRICAL
ConsG	Consumer manufacturing
IntmdG	Intermediate manufacturing
InvesG	Investment goods, excluding hightech
OtherG	OTHER PRODUCTION
MServ	MARKET SERVICES, EXCLUDING POST AND TELECOMMUNICATIONS
Distr	DISTRIBUTION
FinBu	FINANCE AND BUSINESS, EXCEPT REAL ESTATE
Pers	PERSONAL SERVICES
NonMar	NON-MARKET SERVICES

Annex 2. Description of MMD Tables generated by the coordinated analysis

(V4.2. December 2013)

The coordinated analysis of the ICT Impacts project uses a methodology called 'distributed micro data research' (Bartelsman, Haltiwanger, and Scarpetta, 2004). Under this method, computer code is distributed to each national statistical office to be run on their own meta-data and micro data files. The output files generated by each statistical office and combined into the MMD are described below.

Variable names in *Italics* are the unique (combined) keys of the dataset.

OUTPUT Data Files

CONC_Nace2_EUKv2: Concordance to map NACE2 to EUKv2 (=EUK0 in ver2.3)

ind2_BR	NACE2 industry code
eukv2	EUK code (V2.3, EUK0)
emp_br	Employment in split
Pct	Pct of ind2 going to EUKv2
Num	Number of firms in split
Numeuk	Number of EUKv2 mapped to Ind2

MetaData: Information on mapping NSI datasets/variables to ESS names. Ideally, we would get even closer to native structure of datasets in the micro-data laboratories of NSIs. Then, users could write their own programs with requiring work from NSI staff. NSI staff would save zillions in time/effort for their own work as well if this is well organized.

ESSname	Variable name in ESS coding
NSIname	Variable name on NSI dataset
Survey	Source Dataset (in BR, PS, IS, EC)
VarType	char/num (not used or checked)

COVERAGE: Information on linked business register, production survey, E-commerce survey.

YEAR	Year to which data pertain
EUK	Industry classification (EU-KLEMS or ALT definitions, bottom nodes)
SZ_CLS	Size class (1-7, see below for mapping)
N_BR	Number of firms (from Business Register)
N_PS	Number of firms (from Production Survey)
N_EC	Number of firms (from E-Commerce Survey)

N_IS	Number of firms (from CIS Survey)
N_BRPS	Number of firms (merged BR, PS)
N_BREC	Number of firms (merged BR, EC)
N_PSEC	Number of firms (merged PS, EC)
N_BRIS	Number of firms (merged BR,IS)
N_PSIS	Number of firms (merged PS,IS)
N_PSECIS	Number of firms (merged PS,EC,IS)
Emp_BR	Number of employees (from Business Register)
Emp_PS	Number of employees (from Production Survey)
Emp_BRPS	Number of employees (merged BR, PS)
Emp_BREC	Number of employees (merged BR, EC)
Emp_PSEC	Number of employees (merged PS, EC)
Emp_BRIS	Number of employees (merged BR,IS)
Emp_PSIS	Number of employees (merged PS,IS)
Emp_PSECIS	Number of employees (merged PS,EC,IS)
SRC	Coding for different calls to coverage.sas

Dataset varies by industries (EUK) and industries × size_class (ALT)

Coding for SRC = tabulation by EUK, Year.

= tabulation by ALT, Year, SZ_CLS.

DEMOGR: Firm demographics data from Business Register.

YEAR	Year to which data pertain
EUK	Industry classification (EU-KLEMS ALT definition, bottom nodes)
SZ_CLS	Size class
STATUS	Entrant, Exiter, Continuer, or One-year firm
COUNT	Number of firms
EMP	Employment
POS	Aggregate of positive firm-level employment change
NEG	Aggregate of negative firm-level employment change

Coding for STATUS:

CO if firm in year = *t*-1, *t*, *t*+1;
EN in *t*, *t*+1, not t-1;
EX in *t*-1,*t*, not *t*;
OY in *t*.

DEMOGRAGE*: Firm demographics data from Business Register.

YEAR	Year to which data pertain
EUK	Industry classification (EU-KLEMS ALT definition, bottom nodes)
AGECAT	Age Category
STATUS	Entrant, Exiter, Continuer, or One-year firm
COUNT	Number of firms
EMP	Employment
POS	Aggregate of positive firm-level employment change
NEG	Aggregate of negative firm-level employment change

Coding for STATUS:

CO if firm in year = *t*-1, *t*, *t*+1;
EN in *t*, *t*+1, not t-1;
EX in *t*-1,*t*, not *t*;
OY in *t*.

DEMOGRAGESZ*: Firm demographics data from Business Register.

YEAR	Year to which data pertain
EUK	Industry classification (EU-KLEMS ALT definition, bottom nodes)
AGESZ	Combined Age/Size category
STATUS	Entrant, Exiter, Continuer, or One-year firm
COUNT	Number of firms
EMP	Employment
POS	Aggregate of positive firm-level employment change
NEG	Aggregate of negative firm-level employment change

Coding for STATUS:

CO if firm in year = *t*-1, *t*, *t*+1;
EN in *t*, *t*+1, not t-1;
EX in *t*-1,*t*, not *t*;
OY in *t*.

INDDYN: File with variables describing 'industry dynamics'.

YEAR	Year to which data pertain
EUK	Industry classification (EU-KLEMS or ALT definitions)
SRC	=1 for EUK industry hierarchy, =2 for ALT hierarchy
CHURNQ/V	Sum of absolute value of market share changes of firms (Q and/or V)
OPtyp	OP cross term: difference between aggregate (ie weighted average) and average productivity
Ityp	Aggregate 'inputs' for type

Ptyp	Average productivity
Styp	Standard deviation of productivity distribution
Wtyp	Aggregate, ie weighted average, productivity
Nobs_op	Number of firms used in above calculations
Ptyp_xx	Average productivity of firms xx (in CO, EX, EN)
Ityp_xx	Aggregate inputs of firms xx (in CO, EX, EN)
Ctyp_xx	Productivity contribution of firms xx
Styp_xx	Standard deviation of productivity of firms xx

Coding for typ:

LPQ	Labor productivity based on deflated Sales
LPV	Labor productivity based on deflated Value Added
TFP	TFP (Value added with capital and labor)
MFP	MFP (Gross output, with cap, lab, and materials)

Coding for xx:

CO	Continuing firms
EN	Entrants
EX	Exiters

PSSTAT: Summary Statistics for PS variables

ECSTAT: Summary Statistics for E-commerce variables

ISSTAT: Summary Statistics for CIS-innovation variables

YEAR	Year to which data pertain
EUK	Industry classification (EUK or ALT definitions)
SMPL	Sample used (PS, EC, IS, PSIS, PSEC, PSECIS, ECIS)
SUBVAL	Value of Subname, to split industry, (see below for use)
SUBNAME	Variables used to split industry in tabulation (see below)
sumvars	Variables that can be aggregated through summation (see below)
avgvars	Variables that can be aggregated through averaging (see below)
UWT_Pct	For each ‘avgvar’ variable used: the weighted average value across firms, using firm size (employment) as weight.
RWT_Pct	For each ‘sumvars’ and ‘avgvars’ variable used from: the weighted average value across firms, using re-weight procedure to compute weight
RUWT_Bool	For each ‘avgvar’ variable used: the weighted average value across firms, using re-weight procedure to compute weight
NOBS	Number of observations in SMPL,EUK,YEAR,SUBVAL grouping

Organisation of STAT-files:

TABLE	Sumvars; Avgvars	Industry	Samples	Subnames
ECSTAT	---; EC Booleans, EC Pct	EUK	EC, PSEC	---
	---; EC Booleans, EC Pct	ALT	EC, PSEC, ECIS, PSECIS	---
“”	---; EC Booleans, EC Pct	ALT	EC	SZ_CLS
“”	---; EC Booleans, EC Pct	ALT	PSEC	MNC, FRGN_OWN, EXPORT, BUSORG, INTENS, HGE, GZL
“”	---; EC Booleans, EC Pct	ALT	ECIS	CO, INPD, INPS, ORGIN
ISSTAT	IS Numeric; IS Booleans, IS Pct	EUK	IS, PSIS	---
“”	IS Numeric; IS Booleans, IS Pct	ALT	IS, PSIS, ECIS, PSECIS	---
“”	IS Numeric; IS Booleans, IS Pct	ALT	IS	SZ_CLS
“”	IS Numeric; IS Booleans, IS Pct	ALT	PSIS	MNC, FRGN_OWN, EXPORT, HE, GZL
“”	IS Numeric; IS Booleans, IS Pct	ALT	ECIS	INTENS, BROADCAT, BUSORG, ITOUT, LINK
“”	IS Numeric; IS Booleans, IS Pct	ALT	IS	CO, INPD, INPS, ORGIN
PSSTAT	PS Numeric; Productivity vars, HK vars	EUK	PS	---
“”	PS Numeric; Productivity vars, HK vars	ALT	PS, PSEC, PSIS, PSECIS	---
“”	PS Numeric; Productivity vars, HK vars	ALT	PS	SZ_CLS, MNC, FRGN_OWN, EXPORT
“”	PS Numeric; Productivity vars, HK vars	ALT	PSEC	BROADCAT, INTENS, BUSORG, ITOUT, LINK
“”	PS Numeric; Productivity vars, HK vars	ALT	PSIC	CO, INPD, INPS, ORGIN, INNOV

Coding for SUBNAMES

SZ_CLS	Size class (see below for coding)
AGECAT	Age Category (see below for coding)
AGESZ	Age/Size category (see below for coding)
HGE	Dummy for ‘High growth enterprise’ from fast-grow filter (0,1;1=yes, see below)
GZL	Dummy for ‘Gazelle’ from fastgrow filter (0,1;1=yes, see below)
MNC	Multinational Dummy (0,1; 1=multinational firm)
FRGN_OWN	Foreign Ownership Dummy (0,1; 1=owned by foreign firm)
EXPORT	Dummy for exporting firm (0,1; 1=yes)
BROADCAT	Broadband category (see below)
LINK	Number of Electronic Linkages with customers/suppliers (see below; 0, 1, 2, 3)
CO	Innovation cooperation (0,1;1=yes)

INPD New goods or services (0,1;1=yes)
INPS Process Innovation (0,1;1=yes)
ORGIN Organizational Innovation (0,1;1=yes)

HGE
0 = Not a high growth enterprise
1 = More than 10 employees in t0, more than 10 pct annual growth for 3 years

GZL
Not a Gazelle
HGE=1 and younger than 5 years old in year 3

SZ_CLS
1 <=emp <20
20 <=emp <50
50 <=emp < 250
250 <=emp

AGECAT
Age < 3
3 <= Age < 6
6 <= Age < 9
9 <= Age < 12
12 <= Age < 15
Age >= 15

AGESZ
Age < 5 and Emp < 50
Age < 5 and Emp >= 50
Age >= 5 and Emp < 50
Age >= 5 and Emp >= 50

BROADCAT
0 if BROAD=0;
1 if BROAD=1 and BROADPCT<40
2 if BROAD=1 and BROADPCT>=40 and BROADPCT<=90
3 if BROAD=1 and BROADPCT>=90

LINK
0 if max(of SISAINV SISAACC SISAPROD SISADIST SIPUINV SIPUACC)=0
1 if max(of SISAINV SIPUINV)=1
3 if LINK=1 and sum (of SISAACC SISAPROD SISADIST SIPUACC)=4
2 if (LINK=1 or LINK=3) and sum(of SISAACC SISAPROD SISADIST SIPUACC)<4

IS Numerics

RRDINX	Expenditure in intramural R&D (in national currency)
RRDEXX	Purchase of extramural R&D (in national currency)
RMACX	Expenditure in acquisition of machinery (in national currency)
RTOT	Total of these four innovation expenditure categories (in national currency)

PS Numerics

NV	Nominal value added (in national currency)
NQ	Nominal gross output (in national currency)
E	Full-time employment
PAY	Total wage bill (in national currency)
NM	Nominal expenditures on intermediates (in nominal currency)
K	Capital services measure

EC Booleans

BROAD	Firm has broadband
AEBUY	Firm orders through computer networks (websites or EDI)
AESELL	Firm sells through computer networks (websites or EDI)
IACC	Firm has internet
WEB	Firm has website
MOB	Firm has mobile access to internet
ITERP	Enterprise Resource Planning
SISC	Sharing Electronic Data
CRM	Use of CRM
ADEGOV	Use of ADE for sending/receiving from Govt
ADENGOV	Other ADE uses (see below)
ADEINV	Other ADE uses *or* INVOICE
INVOICE	Electronic Invoicing
SISAPU	Combined Supply Chain (see below)
ECOM	Either AEBUY *or* AESELL

IS Booleans

INPD	Introduced onto the market a new or significantly improved good /service
INPS	Introduced new or significantly improved process
MRKIN	Did the enterprise introduce a market innovation
RRDIN	Engagement in intramural R&D
EXPMKT	Enterprise sells in foreign market(s)
ENTGP	Enterprise is part of a group
NEWMKT	Did the enterprise introduce a product new to the market

FUND	Did enterprise receive external funding for R&D
CO	Cooperation arrangements on innovation activities
ORGIN	Innovative business practices
FUNNAT	National R&D Funding
FUNEU	EU R&D Funding
FUNRTD	R&D Funding through RTD projects

EC Pct

BROADPCT	% of workers with acces to broadband
AEBVALPCT	% of orders through internet
AESVALPCT	% of sales through computer networks (websites or EDI)
ECPCT	% E-Commerce (buy+sell)
EMPIUSEPCT	% of workers with access to internet
MOBPCT	% of workers with mobile access to internet

IS Pct

TURNMAR	% of turnover in new or improved products that were new to the market

Productivity variables

LPV	value added based labour productivity
LPQ	gross output based labour productivity
TFP	gross output based productivity with inputs capital, labour and intermediates
MFP	value added based productivity with inputs capital and labour

HK variables

HKPCT	percentage of workers with higher formal education
HKITPCT	percentage of workers with higher formal education in ICT or related fields
HKNITPCT	1 - HKITPCT

Derived ECStat variable definitions (see also readEC.sas of code distribution; right-hand side variables refer to Eurostat transmission file names)

ECPCT sum of e-sales and e-purchases shares (AESVALPCT + AEBVALPCT);
CRM max(of CRMSTR CRMAN);
SISC max(of SISU SICU);
SISAPU max(of SISAINV SISACC SISAPROD SISADIST SIPUINV SIPUACC);

LINK

0 = max(of SISAINV SISAACC SISAPROD SISADIST SIPUINV SIPUACC) = 0;
1 = max(of SISAINV SIPUINV) = 1;
3 = LINK=1 and sum(of SISAACC SISAPROD SISADIST SIPUACC) = 4;
2 = (LINK = 1 or LINK = 3) and sum (of SISAACC SISAPROD SISADIST SIPUACC) < 4;

BROADCAT

if BROAD=0 then BROADCAT=0;
if BROAD=1 and BROADPCT<.40 then BROADCAT=1;
if BROAD=1 and BROADPCT>=.40 and BROADPCT<=.90 then BROADCAT=2;
if BROAD=1 and BROADPCT>=.90 then BROADCAT=3;

ITOUT Combined 3-way for Boolean variables ITSP x ITSPT x XFSP, (1…8)

ITOUT=8 - (ITSP*4+ITSPT*2+XFSP)

BUSORG Combined 3-way for Boolean variables ITERP, CRM, SISC, (1…8)

BUSORG=8 - (ITERP*4+CRM*2+SISC)

Derived EC variables defined in ESSprod.sas, %intens macro, not in readEC.sas)

INTENS Boolean for computed ICT intensity (0, 1 if ICTi > 0.6)
ICTi Geometric mean of latent probability estimates for CRM, ITERP, SISC, ADE (version 1), or MOB, ECOM, SISC, ADENGOV.

Derived ISStat variable definitions (see also readIS.sas)

INPD = max(of inpdgd inpdsv);
INPS = max(of inpspd inpslg inpssu);
MRKIN = max(of mktdgp mktpdp mktpdl mktpri mktmet);
ORGIN = max(of orgbup orgwkp orgexr);
EXPMKT = max(of mareur maroth);
FUND = max(of funloc fungmt funeu funrtd);
INNOV Combined 3-way for Boolean variable INDP x INPS x ORGIN, (1…8)

INNOV = 8 - (INPD*4+INPS2+ORGIN)

ECJOINT: Summary Statistics for combined EC Booleans.

ISJOINT: Summary Statistics for combined IS Booleans.

ECISJOINT: Summary Statistics for combined EC & IS Booleans.

YEAR	Year to which data pertain
EUK	Industry classification (EUK or ALT definitions)
SMPL	Sample used (EC, IS, ECIS)
SUBVAL	Value of Subname, to split industry, (see below for use)
SUBNAME	Variables used to split industry in tabulation (see below for variables used)
jointbool	Two booleans that are combined (b1 and b2)
UWT_Pct	For each 'jointbool' variable used: the weighted average value across firms, using firm size (employment) as weight.
RWT_Pct	For each 'jointbool' variable used from: the weighted average value across firms, using re-weight procedure to compute weight
RUWT_Bool	For each 'jointbool' variable used: the weighted average value across firms, using re-weight procedure to compute weight
NOBS	Number of observations in SMPL,EUK,YEAR,SUBVAL grouping

ECJoint uses combinations of EC Booleans plus derived EC Boolean variable INTENS
ISJoint uses combinations of IS Booleans
ECISJoint uses combinations of BROAD, SISAPU, ECOM, ADENGOV, INPD, INPS, MRKIN, ORGIN, RDENG, CO

PSst: File with moments of distributions of variables in PS;

PSECst: File with moments of distributions of variables in merged PS and EC;

PSISst: File with moments of distributions of variables in merged PS and IS;

PSECISst: File with moments of distributions of variables in merged PS, EC and IS;

YEAR	Year to which data pertain
EUK	Industry classification (EU-KLEMS or ALT definitions)
VNAME	Name of variable whose moments are computed (includes variables in first column of table 1, below)
QRT	Quartile of the distribution (1=lowest; 4=highest, 0=overall mean)
MEAN	Mean of variable VNAME in quartile=QRT of distribution
STD	Standard deviation of variable in quartile
NOBS	Number of firms in quartile

Coding of VNAME:

PSst:	{de, dq, dv,dw, dbq, dbv (firm growth rates), lagged productivity}
PSECst:	{de, dkl, dq, dv, dw, d(Prod), w, k/l, Prod, hkvars, broadpct, lagged broadpct}
PSISst:	{de, dkl, dq, dv, dw, d(Prod), w, k/l, Prod, hkvars, turnmar, lagged turnmar}
PSECISst:	{de, dkl, dq, dv, dw, d(Prod), w, k/l, Prod, hkvars, broadpct turnmar, lagged broadpct turnmar}

PScr: File with moments of joint distribution of two variables in PS;

PSECcr: File with moments of joint distribution of two variables in merged PS and EC;

PSIScr: File with moments of joint distribution of two variables in merged PS and IS;

PSECIScr: File with moments of joint distribution of two variables in merged PS, EC, and IS.

YEAR	Year to which data pertain
EUK	Industry classification (EU-KLEMS or ALT definitions)
QNAME	Name of variable used for quartile distribution
VNAME	Name of variable whose moments are computed. (see below for combinations of Qname and Vname included in analysis)
QRT	Quartile of the distribution (1=lowest; 4=highest)
MEAN	Mean of variable VNAME in quartile=QRT of distribution of QNAME
STD	Standard deviation of variable in quartile
PCC	Pearson correlation coefficient between VNAME and YNAME in quartile
NOBS	Number of firms in quartile

VNAME × QNAME:

PScr:	{de, dq, dv,dw, dbq, dbv} × {lagged productivity}
PSECcr:	{de, dkl, dq, dv, dw, d(Prod)} × {lagged broadpct} + {w, k/l, Prod, hkvars} × {broadpct}
PSIScr:	{de, dkl, dq, dv, dw, d(Prod)} × {lagged turnmar} + {w, k/l, Prod, hkvars} × {turnmar}
PSECIScr:	{de, dkl, dq, dv, dw, d(Prod)} × {lagged broadpct turnmar} + {w, k/l, Prod, hkvars} × {broadpct turnmar}

Annex 3. PS Survey Meta Data

Table A3.1. Availability of PS data based on 2010 Metadata Survey and Output from Code Version 4.2 (Y=yes, N=no)

ESSLait name	Description	AT	DE	DK	FI	FR	IE	IT	LU	NL	NO	PL	SE	SI	UK
NV	nominal value added (in national currency)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
NQ	nominal gross output (in national currency)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
E	full-time employment	N	Y	Y	Y	N	Y	Y	N	N	N	Y	Y	N	N
PAY	total wage bill (in national currency)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
NM	nominal expenditures on intermediates (in national currency)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
K	capital services measure (in national currency)	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	N	Y
HKPCT	pct workers with post upper secondary education	Y	N	Y	Y	N	Y	N	N	N	Y	N	Y	Y	Y
HKITPCT	pct workers with post upper secondary IT education	Y	N	Y	Y	N	Y	N	N	N	Y	N	Y	N	Y
HKNITPCT	pct workers with post upper secondary non-IT education	Y	N	Y	Y	N	Y	N	N	N	Y	N	Y	N	Y
EXPORT	firm exports of goods and services	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
NX	firm exports of goods and services (in national currency)	Y*	N	Y	Y*	Y	Y	Y*	Y	Y*	Y	Y	Y	Y	Y
Wgt_PS	sample weight on business survey	Y	N	N	N	N	N	N	N	Y	N	N	N	N	Y
EMP_BR	number of employees given in Business Register	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
FRGN_OWN	dummy for foreign ownership	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y
MNC	dummy for multinational corporation	Y	N	Y	N	Y	N	N	N	Y	N	Y	Y	Y	N
AGE	age of firm in given year	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
BIRTH	year of birth (first year of activity)	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Link between Business Register and Employer-Employee Register	N	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	N
	Long panel data set (at least 16 years 1995-2010)	N	N	Y	Y	Y	N	N	N	Y	Y	N	Y	Y	N

Note: All the variables are available for 2001-2010 except in Germany and Ireland where the series start in 2002 and in Luxembourg and Slovenia where data are available from 2003. An asterisk (*) denotes firms with information on exports of goods only.

Table A3.2. Coverage for Community Innovation Survey (CIS), by country and year

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
AT	x				x		x		x		x
DE											
DK							x		x		x
FI	x		x		x		x		x		x
FR					x				x		x
IE							x		x		x
IT	x				x		x		x		x
LU	x		x		x		x		x		x
NL	x		x		x		x		x		x
NO			x		x		x		x		x
PL	x				x		x		x		x
SE					x		x		x		x
SI			x		x		x		x		x
UK	x				x		x		x		x

Note: In case a certain question was included in the R&D survey, it might be provided directly from that survey.