

Bartelsman, Eric; Hagsten, Eva; Polder, Michael

Working Paper

Micro Moments Database for Cross-Country Analysis of ICT, Innovation, and Economic Outcomes

Tinbergen Institute Discussion Paper, No. 17-003/IV

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Bartelsman, Eric; Hagsten, Eva; Polder, Michael (2017) : Micro Moments Database for Cross-Country Analysis of ICT, Innovation, and Economic Outcomes, Tinbergen Institute Discussion Paper, No. 17-003/IV, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/149517>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2017-003/VI
Tinbergen Institute Discussion Paper



Micro Moments Database for Cross-Country Analysis of ICT, Innovation, and Economic Outcomes

*Eric Bartelsman*¹

*Eva Hagsten*²

*Michael Polder*³

¹ VU University Amsterdam, Tinbergen Institute and IZA Bonn

² University of Iceland

³ Statistics Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Micro Moments Database for Cross-Country Analysis of ICT, Innovation, and Economic Outcomes*

Eric Bartelsman¹, Eva Hagsten², and Michael Polder³

¹ VU University Amsterdam; Tinbergen Institute; IZA Bonn;

² University of Iceland, ³ Statistics Netherlands

January 11, 2017

Abstract

This paper provides technical documentation to a database built up from firm-level sources titled *Micro moments database* (MMD) that is made available for researchers through Eurostat. The MMD is an internationally harmonized research database of statistical moments collected from linked longitudinal firm-level data in a large selection of EU national statistical offices. The underlying sources for the database are business registers, firm-level surveys on production, usage of Information and Communications Technologies (ICT) and innovative activities, as well as recorded information on trade and worker education, all linked at the firm level. The unit of observation in the MMD represents groups of firms within industries and allows research that bridges micro and macro analysis. The paper delineates the type of research questions that uniquely can be addressed with the MMD, and the advantages and disadvantages of using MMD for questions where alternative datasets are available. The paper next presents the methodology underlying construction of the MMD and provides documentation of the rich set of features. Finally, the paper provides descriptive statistics that highlight the unique character of the data and reviews some of the cross-country analytical work already conducted using the MMD.

Keywords: Innovation, ICT, Productivity, Intangible Investment, linked datasets

*This work has been made possible by Eurostat Grant Agreement 50721.2013.001-2013.082. The micro moments database is available at Eurostat (<http://ec.europa.eu/eurostat/web/microdata/micro-moments-dataset>). We thank participants and a discussant at the Searle Center 2015 for their valuable comments and suggestions.

1 Introduction

There is a long and winding road leading from an innovation to its ultimate impact on society. Most of the twists and turns are fertile ground for economic research, with open questions about incentives and impediments to innovative activity, and the effect of competition and cooperation in going from adoption to private and social impact. The path is flanked by regulations, policies, subsidies and taxes, whose impact on innovation, intended or unintended, also are subject of much research. In order to make progress in understanding the process of innovation, its impact on society, and the role played by policy and economic environment, data is needed to illuminate the area of research. A particular area where data was lacking, was in tracking how firm-level decisions regarding innovative activities, through interactions with customers, suppliers and competitors in the marketplace, lead to aggregate outcomes. For example, analytical work with the MMD can be used to disentangle interactions between firm-level use of information and communication technology (ICT) or firm-level innovative activities on the one hand and aggregate productivity or aggregate employment on the other.

The micro moments database (MMD) is designed to loosen the data constraint for such research. The MMD contains some novel indicators on ICT-use and innovative business practices. More importantly, the MMD provides a facility to identify how changes in national level policies or national economic environment affect the road from innovation to impact. Especially the ability to track both behavioral characteristics and aggregate outcomes in a cross-country setting, with variation in policy stance as well as in timing and direction of policy changes, provides an opportunity for policy effects to be identified. Owing to legal constraints on bringing together single country datasets containing firm-level linked registers and surveys, the MMD makes the tradeoff of gaining a cross-country dimension for analysis at the expense of aggregating up from the firm-level to within-industry groups of representative firms.

The MMD is sourced from firm-level data available at national statistical agencies in European countries. The underlying firm-sources are business registers with the universe of firms, firm-level surveys on production, innovation, ICT-use as well as recorded information on trade and worker skills that are linkable to firms. There are approximately 20 million firms in the underlying business registers, and on the order of 1/2 million firms when these are linked to the firm-level surveys on ICT-use and innovation. The value added covered by firms used for the MMD amounts to well over 50 percent of total industry value added in these countries and the countries included cover more than 80 percent of valued added of the EU-28 plus Norway. In an internationally harmonized manner, the firm-year observations from the national sources are aggregated such that results can be pooled into the cross-country MMD dataset.

The MMD can be linked at the industry-level to other international industry-level datasets, such as STAN or EUKLEMS, yet provides information below the industry-level that can be used to identify producer behavior. Also, the MMD can be merged at the record level to custom aggregations from the firm-level, for example by a researcher with access to confidential US Census business registers and surveys. The MMD includes industry-level aggregates typically published by statistical agencies, such as sums of employment or value added and means of productivity or capital intensity, but also higher moments of the firm-level distributions of underlying variables. Uniquely, the MMD contains moments from multivariate distributions, such as output or growth of firms for different bins of firm-age, size, export intensity, or innovation profiles. Further, by combining variables from different surveys at the

firm level, the MMD contains novel indicators, such as industry average ICT intensity or percentage of ICT-schooled workers in an industry, that go beyond what is available or can be derived from the survey data alone in each country.

The MMD emerged from three subsequent Eurostat-funded projects on microdata linking. In these projects, researchers and statisticians from 14 European National Statistics Offices (NSOs), worked together with academic partners to clean and harmonize the underlying sources.¹ The projects made use of a protocol called distributed micro data analysis (DMD, see Bartelsman et al. 2004) to execute common computer code in each country to generate the MMD. The MMD dataset has been made available for researchers through Eurostat (see table B1).

The paper starts by delineating the types of research questions that uniquely can be addressed with the MMD, and the pros and cons of using MMD for research questions where alternate data sources are available. Next, the paper provides a detailed methodology of the route from national firm-level sources to construction of the MMD. The following section provides documentation of the rich set of features of the MMD, and information on how to get access to the data. Finally, the paper provides descriptive statistics that highlight the unique character of the data and reviews some of the cross-country analytical work already conducted using the MMD. These results are novel in the sense that no cross-country evidence of this type has been available previously. The examples are chosen not only to provide evidence on current academic and policy questions relating to ICT and innovation, but also to showcase the types of analysis that presently can be done using the MMD. Especially in the area of assessing the aggregate 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.

2 Data for Research on ICT, Innovation

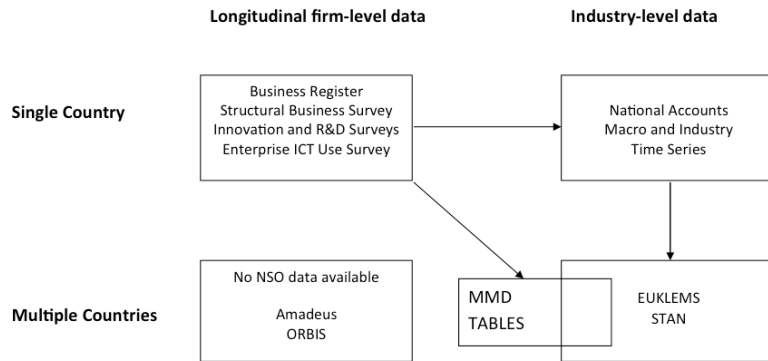
The MMD includes information on traditional firm-level innovative activity, such as R&D spending or product and process innovation as captured by the Community Innovation Survey.² The MMD further combines information on innovative use of ICT and ICT-enabled business processes at the firm level as captured by the EU ICT Usage in Enterprises Survey.³ The MMD also has information on nominal and deflated productive inputs—capital, (skilled) labor, materials—and outputs and information on firm characteristics such as industry, ownership, and age. These are the types of information typically used to study questions regarding ICT adoption, R&D and innovative activity, and economic outcomes such as productivity, factor usage, and production. In describing the research uses of the MMD, we will remain in the broad area of ICT adoption and productivity impact. As will become clear when we present descriptive statistics from the MMD in section 5, the research areas that can be addressed with MMD are much wider, but the strengths of the MMD lie in the ability to cover micro and macro issues in a cross-country setting as described in our examples below on ICT adoption and productivity impact.

¹ Participating NSOs are: Austria (AT), Denmark (DK), Finland (FI), France (FR), Germany (DE), Ireland (IE), Italy (IT), Luxembourg (LU), the Netherlands (NL), Norway (NO), Poland (PL), Slovenia (SI), Sweden (SE), United Kingdom (UK). Earlier rounds of the project included the Czech Republic (CZ) and Romania (RO).

²See <http://ec.europa.eu/eurostat/web/microdata/community-innovation-survey>

³See http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Community_survey_on_ICT_usage_in_enterprises

Figure 1: Data sources for research on ICT and Innovation



From a macro-economic point of view, one of the most important questions in the area of innovation is: what is the (social) impact of innovative activity? Associated questions, such as why firms may lag in their level of innovative activity, and what policy is effective to stimulate innovation, also are important yet difficult questions to answer without appropriate data. The analogue to these questions, but then for adoption, use, and impact of ICT, have been prominent in much empirical literature for the past decades. We will provide a brief overview of this literature according to the underlying data used, as shown in figure 1.

We start discussing studies using single-country macro-level or industry-level data, the top right box, and then move to such data harmonized across countries. 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, the top left box of figure 1. Subsequently, we turn to research questions that are thought to require cross-country firm-level panels. While there are publically accessible cross-country firm-level panels, such as Amadeus or Orbis ⁴, research on innovation and impact is hampered by a limited set of available variables and problems of comparability, coverage, and extensibility to macro analysis.⁵ Cross-country longitudinal panels with harmonized measures of firm-level innovative activity, productive inputs, and economic outcomes (see bottom-left quadrant of figure 1 do not exist. The MMD (bottom row of figure) allows for cross-country analysis, at the expense of compromising on having a firm as the unit of analysis. We conclude this section by arguing how the MMD, possibly combined with informations from cross-country industry panels such as EUKLEMS, or with custom aggregations from firm-level sources, can be used instead to address the main research questions on adoption and impact.

To measure the macro- and industry-level impact of innovation in general and ICT-use in particular the traditional methodology is growth accounting (e.g. Jorgenson et al., 2008). With this method, the contribution of each factor of production to output can be computed. If aggregate production increases more than the contribution attributed to measured production factors, this is ascribed to growth in mul-

⁴Bureau van Dijk

⁵A notable exception is the excellent work on a related question by Bloom et al. (2012a, 2012b) who conduct their own cross-country surveys on management practices to merge with Orbis data.

tifactor productivity (MFP), i.e.: technological progress not embodied in productive 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 emerged as a source of economic growth and that the increase in labor productivity between 1995-1999 could be explained by computer capital-deepening that resulted from double-digit increases in price-quality of ICT equipment. Using the growth accounting framework with detailed sectoral data for the US, Corrado et al. (2007) find that the strong productivity performance was not just related to the contribution of ICT inputs, but likely was instrumental in boosting MFP in the period after 2000.

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. Schreyer 2002). 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 the view emerged 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.

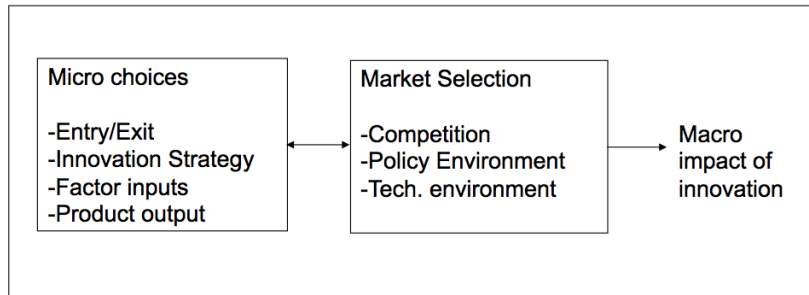
The cross-country growth accounting comparisons have been extended across industries, see the bottom right panel of figure 1, thanks to the efforts in creating EUKLEMS, a harmonized cross-country and industry panel dataset of productive output and inputs in real and nominal terms (O'Mahony and Timmer, 2009).⁶ 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 faster uptake of ICT in the industries that use ICT. Especially productivity gains in market services in the US are not matched in EU countries. The most recent results from growth accounting for the aggregate 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 innovative activity, what policies can affect adoption, and what the mechanisms are through which the such activity improves aggregate productivity, researchers have moved their attention to firm-level data, shown in the top left box of figure 1.⁷ From the earliest papers showing the relationship between productivity and firm-level use of ICT (Brynjolfsson and Hitt, 2000), work has become more adept at estimating the rate of return to investment in ICT capital (e.g. Crépon 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, Brynjolfsson and Hitt, 2002). Strategic interactions between firms and cooperation, for example in ICT standards, or in research joint ventures has also been studied with single-country firm-level data (e.g. Röller et al. 2007). 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 2012), and the relationship between ICT and human capital (Black and Lynch 2001). Finally, an important new area

⁶see www.euklems.net

⁷Much of the firm-level research from the Center for Economic Studies of the US Bureau of the Census (see <http://ideas.repec.org/s/cen/wpaper.html>) draws aggregate conclusions by tracing from the micro to the macro level.

Figure 2: From micro to macro



of research is in the role of management or firm quality in adoption of ICT and its impact (McElheran 2015; Bloom et al. 2012a, 2012b).

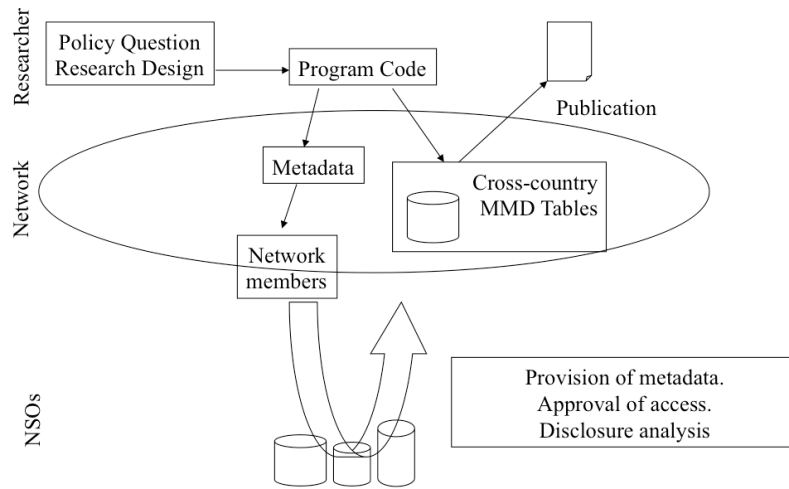
While this work is useful in disentangling correlates of firm-level adoption of ICT, or interaction of ICT with other factors in providing an impact on productivity, only having data for firms in a single location makes the identification of causal effects of policy difficult. In particular, by comparing and contrasting ICT adoption and impact across countries, it becomes possible to disentangle the role played by the policy and market environment by tracking from micro decisions to macro outcomes.

Figure 2 presents a micro-to-macro view, adopted from Bartelsman and Doms (2000), 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, affect 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 particular to each countries economic environment, would be fruitful in gaining a better understanding of the policy drivers of growth.

3 Distributed Micro Data Analysis

Conducting 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. One solution is for researchers to conduct their own firm-level surveys in multiple countries and link the results to available firm-level sources such as Orbis. Bloom et al. (2012a, 2012b) successfully have taken this expensive route to study the impact of firm-level management quality. The approach we describe requires an investment in coordination across national statistics agencies, but subsequently

Figure 3: DMD: Distributed Micro Data Analysis



allows a wide variety of topics to be studied at relatively low marginal costs to the researcher.

Within the ESSLait project (ESSnet on Linking of microdata to analyze ICT impact) we implemented 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' harmonized 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 available legally. Moreover, the infrastructure of NSO data and metadata makes it technically feasible for 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 publicly available cross-country industry datasets.

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 register 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.⁸ 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 (IS), and the ICT Usage Survey (E-commerce Survey, EC). Some NSOs are reducing their reliance on surveys and are collecting more and

⁸An 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.)

more variables of interest directly from register sources (tax authority, customs, educational registry) and linking them to the BR or PS. In the ESSLait project we explore how the sampling strategies and evolution towards register data affect the ability to conduct longitudinal firm-level studies.

For every 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 that is used by the common software to recode local NSO variables to the common set of project variables, and which could be consulted by researchers to design their own program modules (see the middle of center row in figure 3). The tables in Appendix A provide information on the variables and coverage across countries for the underlying firm-level production survey (PS), survey of ICT usage (EC), and Community Innovation Survey (IS) datasets.

The DMD methodology thus provides a way to combine information from single-country confidential micro data for the purpose of cross-country analysis. The methodology results in relatively low marginal costs to generate the cross-country MMD tables, but requires a sizable 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 (indirect) 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. The projects have resulted in a public use version of the MMD that has fully passed disclosure, and a procedure for access to the full MMD tables through Eurostat.

The DMD methodology is not the only way to allow cross-country analysis of firm-level data. To start, much research has taken place using commercially available sources, such as ORBIS from Bureau van Dijk. These data generally are sourced from Chamber of Commerce or mandatory filings of publicly traded firms. However, the coverage and sources vary significantly across countries and it is not feasible to control for the coverage differences across countries and industries, as is possible with DMD. Further, it is costly to collect own firm-level indicators, as done by Bloom et al. (2012a, 2012b) to augment the commercially available firm-level data.

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 single-country data, but need to travel between NSOs to replicate their work in every 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 register and survey 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 Longitudinal Business Database (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.⁹ This method is promising for longitudinal datasets with many firms over time, but having 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, the DMD protocol provides a way to run analytical modules in multiple countries, but does not give 'real time' access, as remote access does. 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 with this method are taking place within Eurostat.

The alternatives above, while imperfect, do contain actual (or anonymized or synthetic) firm-level data, while the MMD contains summary statistics of groups of firms that convey information about the firm-level distributions. For example, within an industry, there is aggregated firm-level information for firms that do and firms that do not innovate or for firms grouped by firm age. Further, there are weighted and unweighted averages as well as summary moments of single variable and multivariate distributions. Each record in the MMD thus is considered an observation from a 'representative firm' comprised of a group of firms in that industry and disaggregated category, for example for the group of firms in the industry that are in the bottom quartile of the productivity distribution, or for the firms that are foreign owned.

As such, the coefficients of equations that are estimated on the MMD panel dataset may suffer from aggregation bias, relative to actual firm-level estimation. On the other hand, specification errors in modeling individual behavior and estimating these at the firm level may be attenuated through aggregation, resulting in better estimates of the effects of interest. An insightful empirical contribution to the issue by Grunfeld and Griliches (1960) points out the tradeoff between aggregation bias and specification errors. Further, they emphasize that the metric of interest may not be the estimation error, but the power of the statistical exercise to provide an explanation to the question at hand. They provide conditions where the explanation or fit of aggregate outcomes is better with aggregate estimation than from aggregation of micro estimates. Using the insights of Grunfeld and Griliches, one can point to applications for which the MMD is best suited. These include research on how micro investment decisions lead to macro outcomes, such as described in the next section. If one is certain of the proper specification of a structural model and is interested in structural parameters of firm-level behavior, then alternative data sources may be preferable to MMD.

The DMD method for generating comparable cross-country datasets has recently been used for other international projects aimed at studying employment dynamics and competitiveness and trade. The OECD DynEmp project (Crisciolo et al. 2014) has been collecting information on firm entry and exit, job turnover, and high growth firms. The ECB CompNet project (Lopez Garcia et al. 2015) is a collaboration of research departments of Eurosystem Central Banks to combine information of firm performance, wages, financial frictions and trade and generate micro-aggregated tables. While the OECD data is not available for outside researchers, researchers can submit project proposals that may allow access to tables of the CompNet project.

⁹<http://www2.vrdc.cornell.edu/news/data/lbd-synthetic-data/>

4 Micro Moments Database

This 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 dataset of the MMD and instructions on where to download the public use version and how to get research access to the full MMD data at Eurostat are provided in Appendix B, table B1. Table 1 provides an overview of the MMD tables, organized by group. The first group of tables provides meta-information about the underlying datasets, next comes a group of summary statistics for each underlying firm-level survey, PS, EC, and IS, then a group of firm demographics files based on the BR, a group of tables with detailed information on distributions and joint-distributions of variables, and finally tables 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 'Panelinfo' table actually refers to three separate tables in the MMD, namely 'availability', 'panel', and 'attrition'. These tables provide information on the availability of variables and firms in PS, EC and IS surveys, on the properties of the linked panels and on panel attrition in every country, industry and year.

The main body 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 (see table A1). Not only are the summary statistics created for all the firms in the survey, but also for sub-samples of linked firms from multiple surveys, such as PSEC which merges the production survey and the ICT use survey, or PSECIS. Uniquely to MMD, many summary statistics are provided in unweighted and weighted versions that reflect the statistical properties of linked samples. Because the linked surveys are linked to the business register at the firm level, weights could be generated to make each sample representative of the firm universe.

Further, also unique to the MMD, the summary tables split industries into sub-groups, such as size-class, age, or multinational status, ICT intensity, and export status, so that users can compare MMD observations from firms that are ICT intensive or that are not, or young or old (see table 2). 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 about complementarity of the technologies.

The firm-demographics tables provide information on firm entry and exit as well as on gross job creation and destruction. This information is available by country, industry, and year, as well as split by size class, by age, or by size and age. Age related disaggregations are not available in official statistics, but uniquely available in the MMD. The data on firm demographics provided in the MMD should be seen as a complement to the other MMD tables as they are constructed in a harmonized manner across countries and can be readily matched to the other tables of the MMD.

Table 1: Datasets in the MMD

Dataset Name	Tables with details	Description
<i>Informational</i>		
Metadata		Mapping of variable names in national firm-level datasets to MMD variable names
Coverage		Information on number of observations and employment by country-year-industry-size class in linked firm-level datasets (BR, PS, EC, IS, and combinations of sources)
Panelinfo		Three datasets showing availability of variables and panel information in all tables by country
Defl		Deflators used (gross output, value added, materials)
<i>Summary Statistics</i>		
PSSTAT, EC-, IS-	2, A3	Statistics of variables in various surveys, by industry, size-class and further breakdowns
<i>Firm Demography</i>		
Demogr, -age, -agesz		Firm demography from Business Register, by age, by age and size: counts and employment by firm status (exit, entry, continuing, one-year), and sums of positive and negative employment changes; EUKLEMS bottom nodes (euk0, 38 industries)
<i>Firm-level distributions</i>		
ECJoint, IS-, ECIS-		Joint adoption of ICT technologies (combinations of all EC Booleans), innovation variables (combinations of all IS Booleans), selected ICT and innovation variables ^a
PSst, EC-, IS-, PSECIS-	A5	Distributional stats from surveys, and combined sample
PScr, EC-, IS-, PSECIS-	A5	Bivariate distribution from surveys, and combined sample
<i>Industry Dynamics</i>		
Inddyn	A6	Indicators of industry dynamics and reallocation
LP files	A7	Long Panel versions of other tables for selected countries

^aEC and IS Booleans are listed in table A1. Selected crossings of ICT versus innovation variables for ECISJoint are based on BROAD, MOB, ITERP, ADE, SISC, CRM and INPD, INPS, MRKIN, ORGIN, RRDIN, CO, see table A1 for variable descriptions.

The tables with distributional information, the 'st' files, provide means and standard deviations of variables for every 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 every country, industry, year, there are thus 10 moments for each variable, which should be sufficient to identify the distribution family and parameters of the underlying firm-level distribution (see tables A5).

There is a table on industry dynamics (see table A6), 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. Finally, we have 'long panels' of firm-level production statistics for the period 1995-2010, for a selection of countries. These contain the typical indicators used for productivity analysis, but also contain information on ICT.

To provide a further display on the unique character of the MMD, that provides information linkable to country/industry panels, but also provides within-industry information, table 2 shows the disaggregations below industry-level available in each of the main statistical tables of the MMD. For example, the main statistical files show a breakdown of the main indicators by size class of firms within an industry, or for foreign-owned firms versus domestic firms. There is also a disaggregation by ICT intensity of firms. Having breakdowns by type of innovative activity or use of ICT allows for analyses that are not possible in other datasets.

At present, the full MMD tables are available for researchers at Eurostat. A public use dataset, with a selection of indicators from the MMD tables at a level of aggregation that passes disclosure, is available for download from Eurostat website (see table B1). To get access for research purposes to all the MMD tables with full industry and sub-industry detail and no suppressed cells, an application must be made to Eurostat. While the micro-aggregated data in MMD are not at the firm-level, many cells are based on just a few firms and would not pass through the statistical disclosure rules of NSOs. Therefore, at present the full MMD must be accessed at secure Eurostat facilities. More information about the public use data, the MMD, and information about the access procedures can be found on the Eurostat website, using the links given in Appendix B, table B1.

Table 2: Disaggregation below Industry Level

	PSSTAT	ECSTAT	ISSTAT
<i>Firm size and growth</i>			
Size	PS	PSEC	PSIS
Age, ageXsize	PS	PSEC	PSIS
HGE, Gazelle	PS	PSEC	PSIS
<i>International Links</i>			
MNC	PS	PSEC	PSIS
foreign-owned	PS	PSEC	PSIS
exporter	PS	PSEC	PSIS
<i>ICT usage</i>			
ICT intensive	PSEC		ECIS
broadband	PSEC		ECIS
ICT specialists, training, outsourcing	PSEC		ECIS
Sharing data	PSEC		ECIS
e-business	PSEC		ECIS
<i>Innovation type</i>			
product	PSIS	ECIS	
process	PSIS	ECIS	
organizational	PSIS	ECIS	
combinations of innovations	PSIS	ECIS	
innovation collaboration	PSIS	ECIS	

Notes: Disaggregations are available for 7 main industries (EUKLEMS 'ALT' classification). Each entry in the table refers to (linked) sample used for the MMD table stated in the column header. Variables available in each MMD table are given in table A1. Variables used to define industry disaggregation are given in table A3.

5 Descriptive statistics and research results from MMD

In this section, we will present some descriptive statistics and highlights from analytical work done to date with the MMD. The purpose is to showcase some of the possible directions that researchers could take using the MMD rather than to test a particular hypothesis regarding ICT, innovation, and economic outcomes. We first show the development over time of use of various information technologies for the total economy, averaged across sectors. Next, we display different analytical measures of ICT intensity, defined at the firm level, across countries. We then disaggregate below an industry, by showing for each industry the ICT intensity for firms that do or do not undertake innovative activity. We also look at complementarities between ICTs and innovative activities at the firm level. The MMD also contains data on higher moments of distributions at the firm level. We show some evidence that the volatility and dispersion of productivity is higher for ICT intensive firms than for others. We also display intriguing descriptive statistics of within-industry joint distributions of productivity and employment (growth). To close, we turn to some analytical work using the MMD that explores impact of ICT on productivity or demand for skilled workers (Bartelsman et al. 2016a, Falk and Hagsten 2015, and Falk and Biagi 2016).

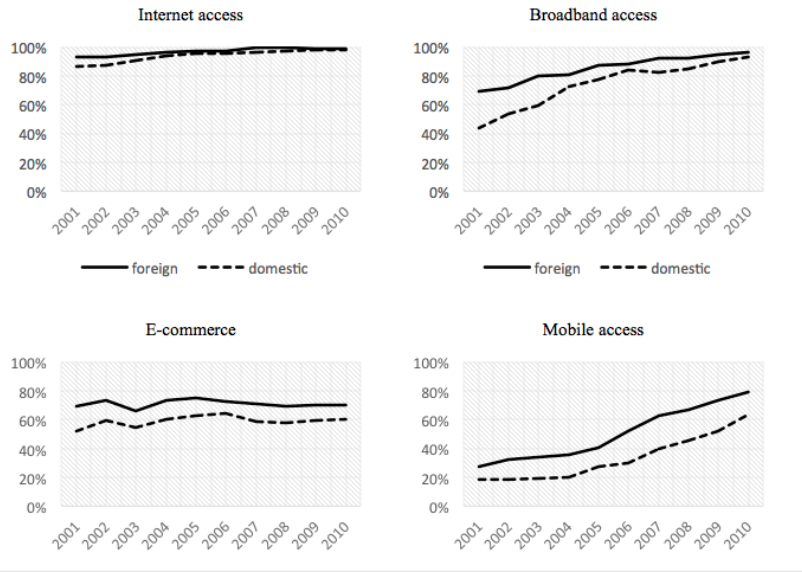
5.1 Descriptive Statistics from MMD

Figure 4 shows the development over time of the penetration of four information technologies, namely internet access, broadband access, mobile internet access and e-commerce, averaged over eleven EU countries, separately for foreign- and domestically-owned firms. The disaggregation of ICT-use across ownership is unique to the MMD and was possible through firm-level linking of the survey with the business register. We see that adoption of internet access was saturated early on, broadband is nearing saturation and mobile access is still increasing. Use of e-commerce does not seem to be increasing much over time. For all measures, penetration among foreign owned firms is higher than for domestic firms.

Traditional indicators of ICT intensity at the industry-level usually involve shares of ICT investment or capital stock relative to total capital (formation) or value added. For firm-level analysis, a drawback of industry indicators is that there is substantial heterogeneity between firms with respect to ICT usage, even within narrowly defined industries. Firm-level information on ICT investment, however, is hard to come by, especially for software which is increasingly important. Collection of such data is rare, and if it exists frequently the data collection method is country-specific, and therefore not suitable for making international comparisons.

An alternative is to use indicators from the ICT survey which are usually Boolean or percentages. An example is our broadband enabled workers indicator, which combines the availability of broadband with the percentage of internet enabled workers. However, rankings of ICT intensity depend on the specific metric chosen to measure ICT intensity. To mitigate the sensitivity with respect to a single indicator, it is therefore advisable also to consider a set or composite of indicators. On the other hand, when using composite indicators one has to make a choice which technologies to include, whether to use weighting when combining them, and if yes, choose the type of weighting. The MMD includes a composite indicator called ICT_i, which is based on a probability model for the adoption of a set of underlying technologies. The idea is that the propensity is a measure of the underlying (latent) value of

Figure 4: Evolution of ICT Penetration (pct of firms)



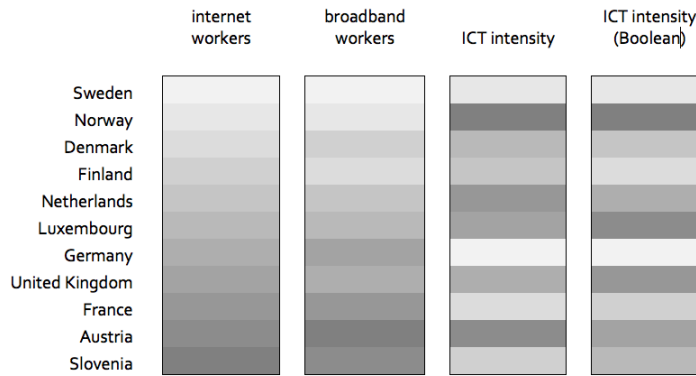
source: MMD ECStat v4.2, Market Services (excl Telecom), averaged across Germany, Denmark, Finland, France, Ireland, Netherlands, Norway, Poland, Sweden, Slovenia, United Kingdom.

adoption, which itself should be a function of the underlying ICT intensity. The ICT_i indicator in the MMD is a geometric average of the propensities of different new technologies. The use of propensities implicitly provides a weighting of the variables, and is smoother than discrete indicators. Another issue in cross-sectional comparisons is that of saturation of certain variables. As levels of adoption reach 100%, indicators become less useful to distinguish between firms (e.g. PCs and internet usage shown in figure 4). The advantage of a composite indicator is that technologies can be dropped or added when they become less or more relevant, so that the indicator remains a good cross-sectional measure of ICT intensity over time. Moreover, since the ICT_i indicator tries to get at the underlying ICT intensity rather than the specific technologies, the resulting rankings should prove to be more robust.

A comparison of ICT indicators across countries is given in figure 5, here 'broadband workers' is based on the percentage of workers with broadband access, 'ICT_i' is the ICT intensity index, and 'Intens' is the percentage of firms in an industry with a non-zero binary intensity indicator. There are some surprises in the ranking based on ICT_i, 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 analytical work with MMD, it has been found that the rankings are stable and the intensities show a slightly increasing pattern over time for most countries. 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 analysis can 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 6 shows an example of using the industry and time dimensions to isolate a country effect. The

Figure 5: Alternate indicators of ICT intensity created for MMD (2009).



Source: MMD, version 4.2. Based on employment and sample re-weighted figures. Light to dark shades means higher to lower ICT intensity. Internet worker: EMPIUSEPCT, Broadband workers: broadpct, ICT intensity: ICTi, ICT intensity (Boolean): Intens. See documentation of ECSTAT table A1 in the Appendix.

figure reveals substantial heterogeneity across countries, with the Nordic countries (minus Denmark) constituting the top-3 at some distance, and Poland, Ireland, and Italy showing lower rates of broadband usage. Correcting for industry structure and time effects increases further the heterogeneity among countries. Overall, the ranking is hardly affected by the correction, but especially just below the middle, it is somewhat more sensitive (e.g. Austria climbs from eleventh place in the uncorrected figures to eighth place in the corrected ones).

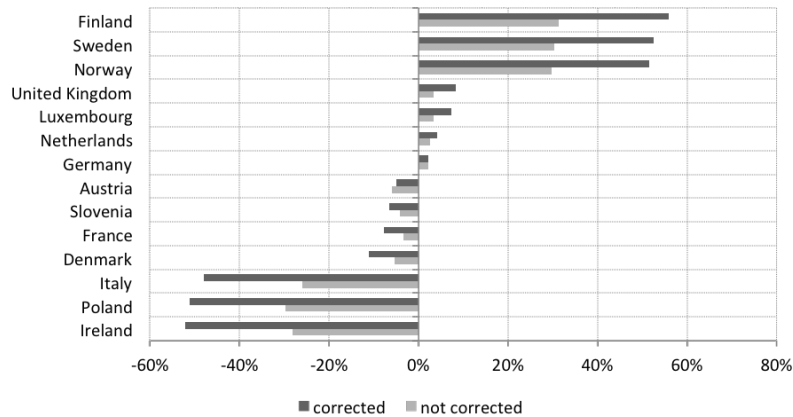
The MMD allows one to explore correlates and causes of ICT adoption or innovative activity. Because underlying surveys are linked at the firm-level, industry data in the MMD can be split by values of firm-level variables, for example as shown in Table 3. Here, we look at the percentage of firms (size-weighted) in 2008 in Finland by broad industry categories that use customer relation software (CRM), split by whether or not those firms have innovative business practices (from the joint EC-IS firm sample). Not very surprisingly, the use of CRM is significantly higher for firms that have organizational innovation.

To further explore how different technologies are either complements or substitutes, the MMD provides counts of how many firms adopt each separately, as well as jointly. For example, table 4 provides this data for the adoption of mobile internet and organizational 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 construct contingency tables and to track how different combinations affect performance.

In the MMD, we have built up simple indicators of productivity from the firm-level data. The data include labor productivity, measured as deflated sales or value added per worker.¹⁰ As described in the introduction, much evidence exists concerning correlations between ICT usage and productivity. Using the 'long panel' tables of the MMD we can look as well at correlations between dispersion in productivity

¹⁰The MMD also include measures of TFP, computed as a Solow residual, using country specific firm-level proxies for capital service inputs.

Figure 6: ICT intensity, corrected for industry/time effects



Source: MMD v4.2 ECStat. Broadband-enabled workers. Sample weighted, country average relative to overall average, corrected and uncorrected for industry structure and year effects.

and ICT. The long panels are derived from unbalanced panels from the Business Surveys in a selection of countries where these data are available from the mid-1990s onwards. In the long-panels 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 productivity and output growth for these two groups of firms. As seen in table 5, the dispersion and volatility 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 labor productivity growth is presented. The volatility is measured at the firm-level using a 5-year moving window. Then the firm-level volatility is averaged into an industry series, using firm-size weights. Finally, the industry volatility 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.¹¹ 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.

In Bartelsman et al. (2016b) the relationship between ICT and dispersion is explored further. In the MMD PSSTAT table, a measure of the cross-sectional dispersion of firm-level productivity is available for every country, industry, and year. The productivity dispersion is regressed on broadband penetration and fixed effects, both in levels and first differences. In levels, but in first-differences as well, broadband intensity has significantly positive effect on productivity dispersion. The main results of the paper are to link employment protection as a causal factor in reducing allocation of resources to innovative, high risk, activities. The paper uses moments from the MMD (and other sources) to calibrate a model and compare simulated data and actual industry data to show how a channel through which employment protection can hold back aggregate productivity.

An issue to be explored further, is the causation of the relationship between intensive ICT use and

¹¹Unfortunately, for the cross-sectional exercise, the measure was not collected for non-ICT intensive firms but only for ICT intensive firms and the industry as a whole.

Table 3: Percentage of firms with CRM

Industry	Industry Disaggregation	
	Org. Innovation = no	Org. Innovation=yes
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: MMD v4.2, table=ECStat, smpl=ECxIS, 2008, Finland

Table 4: Number of firms by Mobile Internet and Organizational Innovation

		Org. Innovation		total
		=no	=yes	
Mobile Internet	=no	711	222	933
	=yes	413	234	647
total		1124	456	1580

source: MMD v4.2, table=ECISJoint, 2008, Market Services, Netherlands

productivity volatility and dispersion. 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 either case, as indicated in figure 2, market selection and competition are an important part of the story of how the firm-level impact of ICT and innovation translates into macro-economic performance. For instance, ICT may improve firm performance but to determine the aggregate impact, the allocation of inputs across firms matters. To increase overall productivity, resources should be allocated to the more productive firms.

The fact that dispersion of productivity is seen to be increasing with ICT, may lead to the possibility that the adoption impact can be leveraged if highly productive ICT intensive firms are larger, or grow more rapidly than firms that are unsuccessful at adopting ICT. As a first step to explore this, we can use the MMD to view reallocation of resources. 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 (also known as the Olley-Pakes cross-term, after Olley and Pakes, 1996). Such indicators, along with information on the churn of market shares, are available in the Inddyn table of the MMD.

As an example, figure 7 show the joint distributions of productivity, and the distributions of employ-

Table 5: Dispersion and Volatility of Productivity and Output Growth

Country	Time Series Volatility				Cross Sectional Dispersion			
	Productivity		Output		Productivity		Output	
	ICT=0	ICT=1	ICT=0	ICT=1	ALL	ICT=1	ALL	ICT=1
DK	.19	.21	.24	.26	.23	.24	.29	.32
FI	.19	.28	.21	.31	.25	.27	.30	.33
FR	.20	.18	.22	.18	.21	.18	.21	.19
NL	.13	.14	.11	.13	.22	.24	.20	.21
NO	.18	.26	.21	.29	.32	.35	.33	.35
SE	.20	.26	.32	.38	.33	.37	.49	.52

Source: MMD v3.4, table=LPSStat. Notes: The table shows the averaged standard deviation of labor productivity and output growth. In the time series columns, the standard deviation of growth (volatility) 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 volatility 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 (dispersion) is computed, for ICT intensive firms (ICT=1) and for all firms (ALL). The industry and time dispersion measures are then averaged over time and across industries with fixed industry weights.

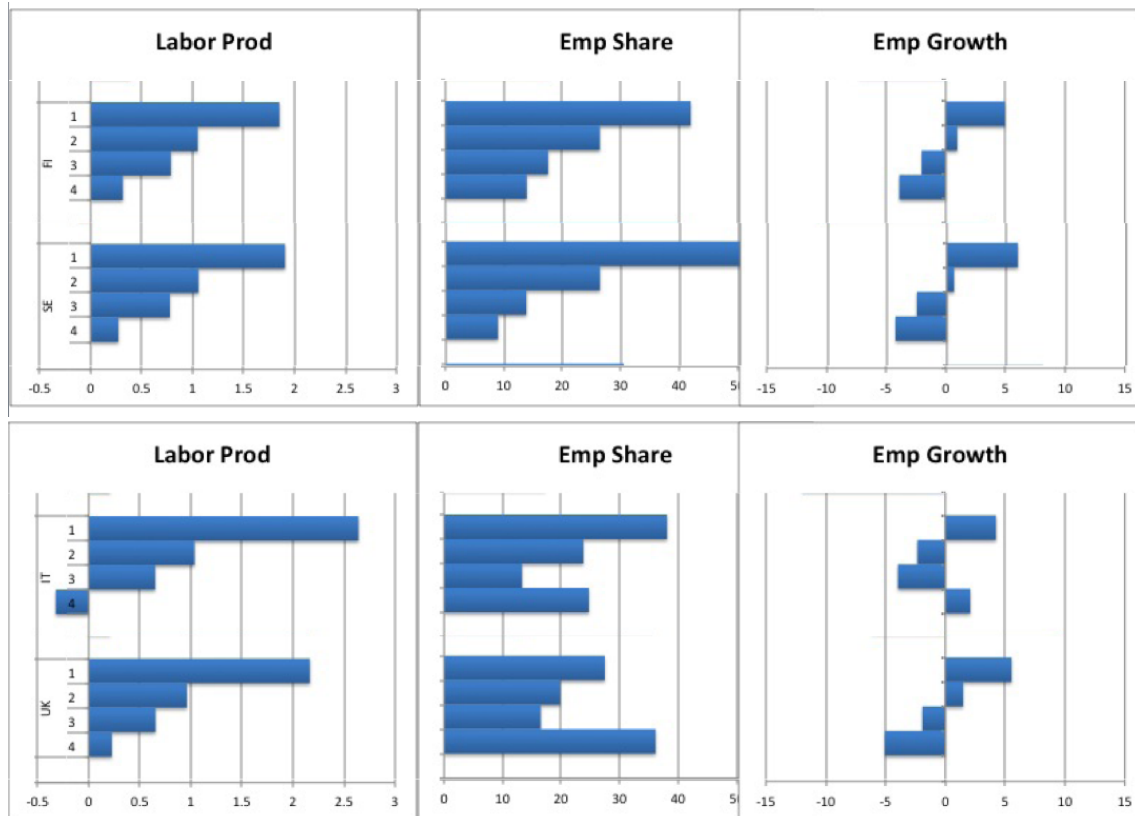
ment and employment growth conditional on productivity. The information in the charts is available for all countries and sectors, but here we show information for Finland, Sweden, Italy and the UK, for the manufacturing sector (excluding ICT) and averaged for the period 2003-2009. The first column shows 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 productive as the average firm, while firms in the lowest quartile are only about a quarter as productive as the average firm. The next column shows the distribution of employment across productivity quartiles. In Sweden half of the employment in manufacturing is in the quartile of firms with the highest productivity. Finally, the last column shows employment growth of firms by lagged productivity quartile. Here as well, a monotonic relationship between productivity and employment growth is good for productivity growth, as is the case for both countries. The story for Italy and the UK is different, as the allocation does not seem as good for aggregate productivity. In the UK, we see high employment shares in low productivity firms, and in Italy employment growth is not monotonic over quartiles.

As an aside, the information on reallocation also can be used to compare pre- and post-crisis developments, as in OECD Future of productivity report (OECD, 2015), reproduced here as figure 8. The figure illustrates that job losses are concentrated in the least productive firms, especially in the (post-)crisis years 2008-2010.

5.2 Analytical Work with MMD

The importance of reallocation to assess ICT impact is explored with the MMD data in Bartelsman et al. (2016a). They run a simple panel data production function with labor productivity as the dependent variable, and capital intensity, the percentage of broadband enabled workers and fixed effects (for country,

Figure 7: Heterogeneity and Resource Allocation



Source: MMD v3.4, 2003–09, Manufacturing excl ICT. Top: Finland and Sweden. Bottom: Italy and UK

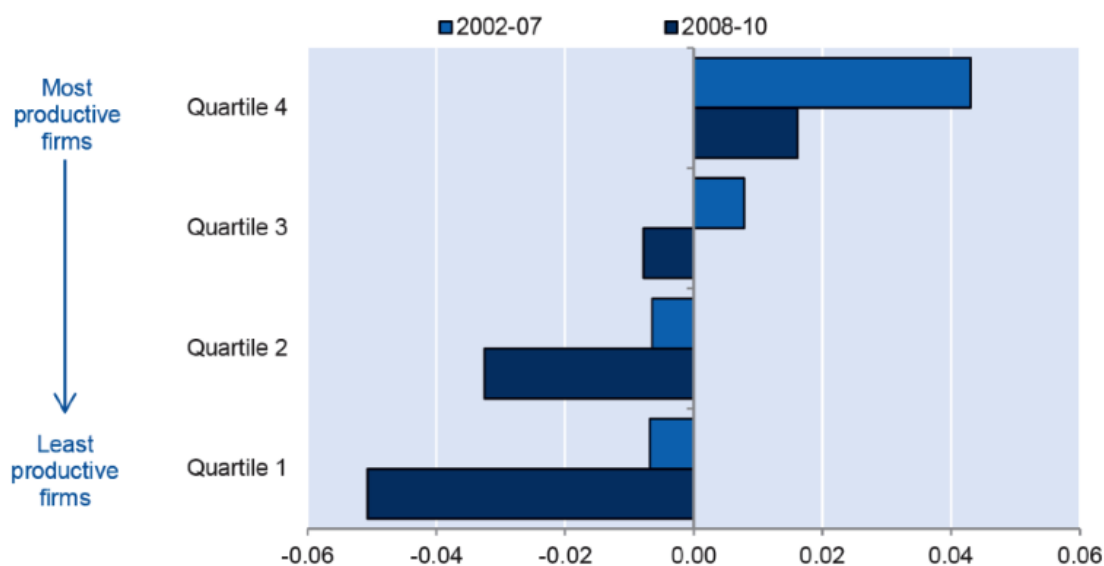
industry and time) as explanatory variables.

The production function is estimated with two different industry level aggregates that are available in the MMD, namely the average and a weighted average. The latter is comparable to typical industry-level production function estimation, where total industry output divided by total industry employment (labor productivity) is regressed on other industry totals. The former is more comparable to estimation with firm-level data, where the coefficients on capital intensity and broadband are estimated for the average firm.

The results in Bartelsman et al. (2016a) show that, on average broadband does not have a significant effect, which corroborates the findings from firm-level regressions in the separate countries where many countries did not show significantly positive effects, as reported in Eurostat (2013). The significant effect in the weighted average regressions indicates that the more productive firms that use broadband intensively are larger. This is consistent with the narrative that selection effects favor successful ICT adopting firms.

The MMD has also been used to study the impact of other indicators of ICT use. To provide evidence on the effects of e-commerce, Falk and Hagsten (2015) use the MMD to run regressions of productivity on the proportion of firms conducting sales over internet (or other electronic networks), both in long-difference (two year window) form and using GMM. Penetration rates of e-commerce for selling is

Figure 8: Average employment growth across the firm MFP distribution; deviation from 2002-10 average



Source: OECD (2015), based on MMD

much lower than penetration of internet or broadband, and varies more strongly across countries and sectors, suggesting the possibility that early adopters face large potential gains. In the study, it is shown that growth in e-commerce penetration is significantly correlated with productivity growth, but that this effect is stronger for small firms at low penetration rates and absent for large firms at higher penetration rates.

A next example of analysis with MMD shows that using micro moments data can provide results to interesting macro questions while still being able to take into account micro behavior. In Bartelsman et al. (2016a), the framework inspired by Crépon et al. (1998) is used to fit the pieces together. The framework, called CDM after the authors, is used to simultaneously estimate adoption of innovative activity and the productivity effect of innovation. In this setting, Bartelsman et al. estimate on average, the effect ICT-use has on the probability that a firm innovates, and how the predicted probability of innovating and the efficiency of resource allocation affects aggregate productivity. In the CDM model, one needs to first estimate the probability that a firm engages in innovative activity. In typical industry data, this information is lost in aggregation, and one usually only has information on the share of innovating firms. In the MMD data, industry data is split for firms that innovate and firms that do not. Using the country, industry and time panel dimensions, Bartelsman et al. can thus estimate a (weighted) 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. They find that various measures of ICT correlate positively with the probability of innovating. In the productivity equation, they use both averages and aggregates (or size weighted average) for productivity and the explanatory variables, as described above. They show that on average predicted innovative activity does not affect productivity significantly, while aggregate

productivity is affected significantly.

Another example is from a study of the relative demand for skilled employees (with post upper secondary education) of Falk and Biagi (2016). They utilise information on ICT usage, education, production and capital in firms from the MMD for seven European countries over the years 2001-2010. They regress the proportion of highly skilled employees on the use of different types of ICT and internet. They find that these technologies are positively and significantly related to the proportion of highly skilled employees, especially so for the manufacturing firms. Their results also indicate that ICT is a complement to higher skills rather than a substitute.

6 Conclusions and future work

The paper has introduced the micro moments database for cross-country research into ICT, innovation and economic outcomes that is available for researchers through Eurostat. The MMD bridges a gap between single-country linked longitudinal micro data and cross-country industry panel data. As shown in the literature review, cross-country comparisons of adoption and impact of ICT at the firm-level would be ideal to provide the variation needed to disentangle the impacts of policy, environment, and factor availability. However, such data are not readily available. The MMD provides enough information below the industry level to identify behavioral responses, yet is aggregated enough to allow merging the data across countries to improve identification and micro-to-macro analysis.

The paper describes the harmonized methods through which the data are collected, and provides documentation to the MMD, and its many novel features. Information is provided on the public use version of the MMD and procedures to access the data for research use at Eurostat. The paper next provides some summary statistics of the MMD and results from ongoing and completed research using the MMD in order to highlight the dimensions in which the MMD is unique. This section hopefully will provide inspiration to researchers about the questions that uniquely can be answered with these data.

The efforts so far have led to some analytical findings on economic causes and effects of ICT use. Much work remains to be done on this topic, especially in understanding why firms in countries across the EU differ so much in their adoption of ICT, even when correcting for industry patterns. The MMD opens up the possibility of exploring this through connecting information on productive inputs and outputs across countries to information on economic environment and policy. We hope that availability of the MMD will attract the academic community to take up the challenge to improve our understanding of the links between policy, innovation and ICT, and economic performance.

In the future, we will strive to expand the datasets available in the MMD, 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 financial crisis and its aftermath are limiting the ability of firms to invest in risky, ICT-related, projects.

References

- [1] ABRAMOVSKY, L. AND R. GRIFFITH (2006): “Outsourcing and Offshoring of Business Services: How Important Is ICT?” *Journal of the European Economic Association*, 4, 594–601.
- [2] ANDREWS, D. AND F. CINGANO (2012): “Public Policy and Resource Allocation: Evidence from Firms in OECD Countries,” OECD Economics Department Working Paper 996, OECD Publishing.
- [3] BARTEL, A. P., S. LACH, AND N. SICHERMAN (2012): “Technological Change and the Make-or-Buy Decision,” *Journal of Law, Economics, and Organization*, 30(1), 165–192.
- [4] BARTELSMAN, E., G. V. LEEUWEN, AND M. POLDER (2016a): “CDM using a cross-country micro moments database,” *Economics of Innovation and New Technology*, <http://dx.doi.org/10.1080/10438599.2016.1202517>.
- [5] BARTELSMAN, E. J. AND M. DOMS (2000): “Understanding Productivity: Lessons from Longitudinal Microdata,” *Journal of Economic Literature*, 38, 569–594.
- [6] BARTELSMAN, E. J., P. A. GAUTIER, AND J. DE WIND (2016b): “Employment Protection, Technology Choice, and Worker Allocation,” *International Economic Review*, 57, 787–826.
- [7] BARTELSMAN, E. J., J. C. HALTIWANGER, AND S. SCARPETTA (2004): “Microeconomic evidence of creative destruction in industrial and developing countries,” *IZA Discussion Paper*, No. 1374.
- [8] BLACK, S. E. AND L. M. LYNCH (2001): “How to Compete: The Impact of Workplace Practices and Information Technology on Productivity,” *Review of Economics and Statistics*, 83, 434–445.
- [9] BLOOM, N., C. GENAKOS, R. SADUN, AND J. V. REENEN (2012a): “Management Practices Across Firms and Countries,” Working Paper 17850, National Bureau of Economic Research.
- [10] BLOOM, N., R. SADUN, AND J. V. REENEN (2012b): “Americans Do IT Better: US Multinationals and the Productivity Miracle,” *American Economic Review*, 102, 167–201.
- [11] BOONE, J. (2008): “A New Way to Measure Competition,” *The Economic Journal*, 118, 1245–1261.
- [12] BRESNAHAN, T. F., E. BRYNJOLFSSON, AND L. M. HITT (2002): “Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence,” *The Quarterly Journal of Economics*, 117, 339–376.
- [13] BRYNJOLFSSON, E. AND L. M. HITT (2000): “Beyond Computation: Information Technology, Organizational Transformation and Business Performance,” *Journal of Economic Perspectives*, 14, 23–48.
- [14] COLECCHIA, A. AND P. SCHREYER (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, 408–442.

- [15] CORRADO, C., P. LENGERMANN, E. J. BARTELSMAN, AND J. J. BEAULIEU (2007): “Sectoral Productivity in the United States: Recent Developments and the Role of IT,” *German Economic Review*, 8, 188–210.
- [16] CRÉPON, B., E. DUGUET, AND J. MAIRESSE (1998): “Research, Innovation And Productivity: An Econometric Analysis At The Firm Level,” *Economics of Innovation and New Technology*, 7, 115–158.
- [17] CRÉPON, B. AND T. HECKEL (2002): “Computerization in France: An Evaluation Based on Individual Company Data,” *Review of Income and Wealth*, 48, 77–98.
- [18] CRISCUOLO, C., P. GAL, AND C. MENON (2014): “The Dynamics of Employment Growth,” OECD Science, Technology and Industry Policy Papers 14.
- [19] EUROSTAT (2008): *Information society: ICT impact assessment by linking data from different sources*, Luxemburg: Eurostat.
- [20] FALK, M. AND F. BIAGI (2016): “Relative demand for highly skilled workers and use of different ICT technologies,” *Applied Economics*, <http://dx.doi.org/10.1080/00036846.2016.1208357>.
- [21] FALK, M. AND E. HAGSTEN (2015): “E-commerce trends and impacts across Europe,” *International Journal of Production Economics*, 170, Part A, 357–369.
- [22] GRUNFELD, Y. AND Z. GRILICHES (1960): “Is Aggregation Necessarily Bad?” *The Review of Economics and Statistics*, 42, 1–13.
- [23] HEMPELL, T. AND T. ZWICK (2008): “New Technology, Work Organisation, and Innovation,” *Economics of Innovation and New Technology*, 17, 331–354.
- [24] HSIEH, C.-T. AND P. J. KLENOW (2009): “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 124, 1403–1448.
- [25] JORGENSEN, D. W., M. S. HO, AND K. J. STIROH (2008): “A Retrospective Look at the U.S. Productivity Growth Resurgence,” *Journal of Economic Perspectives*, 22, 3–24.
- [26] JORGENSEN, D. W. AND K. J. STIROH (1999): “Information Technology and Growth,” *The American Economic Review*, 89, 109–115.
- [27] LOPEZ-GARCIA, P. AND F. DI MAURO (2015): “Assessing European competitiveness: the new CompNet micro-based database,” Tech. rep., ECB Working Paper Series 1764.
- [28] MCELHERAN, K. (2015): “Do Market Leaders Lead in Business Process Innovation? The Case(s) of E-business Adoption,” *Management Science*, 61, 1197–1216.
- [29] OECD (2015): “The Future of Productivity,” Paris: OECD.
- [30] OLLEY, G. S. AND A. PAKES (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64, 1263–1297, 1026358051.

- [31] O'MAHONY, M. AND M. P. TIMMER (2009): "Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database*," *The Economic Journal*, 119, F374–F403.
- [32] RÖLLER, L.-H., R. SIEBERT, AND M. M. TOMBAK (2007): "Why Firms Form (or do not Form) RJVS," *The Economic Journal*, 117, 1122–1144.
- [33] SCHREYER, P. (2002): "Computer Price Indices and International Growth and Productivity Comparisons," *Review of Income and Wealth*, 48, 15–31.
- [34] VAN WELSUM, D., W. OVERMEER, AND B. VAN ARK (2013): "Avoiding the Digital Desert," , New York: The Conference Board.

A Data Appendix: MMD Documentation

Table A1: MMD variables (-STAT tables)

Table	data type	Variable name	Description
PS-	numeric	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 national currency)
		K	capital services measure
		NX	Value of exports (in national currency)
		LPV*	Labor productivity (value added)
		LPQ*	Labor productivity (gross output)
		TFP*	TFP (gross output)
		MFP*	MFP (value added)
		HKPCT*	pct workers with post upper secondary education
		HKITPCT*	pct workers with post upper secondary IT education
		HKNITPCT*	pct workers with post upper secondary non-IT education
		DE*	Change in employment
VAY*	Value added to gross output ratio		
EC-	Boolean	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
		DIALUP	Firm uses a dial-up connection to access the internet
		ITERP	Enterprise Resource Planning
		ECOM*	E-commerce
		SISC*	Sharing information on supply chain management

		CRM*	Use of CRM
		INTENS*	ICT intensity
		INTENS2*	Alternative to INTENS (see table A3)
		ITSP	firm employs IT specialists
		ITSPT	firm provides IT training
		XFSP	firm outsources IT functions
		ADEGOV	ADE for sending or receiving data to/from public authorities
		ADENGOV	ADE for sending or receiving data (not public authorities)
		INVOICE*	Firm sends or receives electronic invoices
		ADEINV*	Firms uses ADE or e-invoices
		SISAPU*	Sharing information on sales/purchases with customers/suppliers
	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
	numeric	ICTi*	ICT intensity based on average propensity of adoption
		ICTi2*	Alternative to ICTi (see table A2)
IS-	Boolean	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
		FUNNAT	Did enterprise receive external funding for R&D, local or national government
		FUNEU	Did enterprise receive external funding for R&D from EU
		FUNRTD	Did enterprise receive external funding for R&D from framework program
		CO	Cooperation arrangements on innovation activities
		ORGIN*	Innovative business practices

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

Notes: '*' composite indicators are derived from underlying source variables, see table A2.

All monetary variables are expressed in euro.

Table A2: MMD Composite indicators

MMD name	Input variables	Operator/mapping
LPV*	Real value added, employment	Ratio
LPQ*	Real gross output, employment	Ratio
MFP*	Value added, employment, capital measure	Residual
TFP*	Gross output, employment, capital measure, intermediates	Residual
DE*	Change in employment	Difference
VAY*	Value added, gross output	Ratio
SISC	SISU, SICU	Maximum
CRM	CRMSTR, CRMAN	Maximum
BROADPCT*	BROAD, EMPIUSEPCT	Product
ECOM*	AEBUY, AESELL	Maximum
ECPCT*	AEBVALPCT, AESVALPCT	Sum
INPD	INPDGD, INPDSV	Maximum
INPS	INPSLG, INSPD, INPSSU	Maximum
MRKIN	MKTDGP, MKTPDP, MKTPDL, MKTPRI, MKTMET	Maximum
EXPMKT	MAREUR, MAROTH	Maximum
FUND*	FUNLOC, FUNGMT, FUNEU, FUNRTD	Maximum
ORGIN	ORGBUP, ORGWKP, ORGEXR	Maximum
ICTi*	CRM, ERP, SCM, ADENGOV	average propensity
ICTi2*	MOB, ECOM, SISAPU, ADENGOV	average propensity
INTENS*	CRM, ERP, SCM, ADE	Boolean (ICTi > .5)
INVOICE*	INVREC, INVSND	Maximum
ADEINV*	ADENGOV, INVOICE	Maximum
SISAPU*	SISAINV, SISAACC, SISAPROD, SISADIST, SIPUINV, SIPUACC	Maximum

Notes: '*' are unique to MMD. See table A4 for input variable definitions.

Table A3: Coding of Disaggregation Variables

Variable	Description	MMD name	Coding
Size	Size classes	SZ_CLS	1: < 20; 2: 20–49; 3: 50–249; 4: ≥ 250 employees
HGE	High growth enterprise	HGE	Boolean, more than 10 employees in t and more than 10% annual growth for 3 years
Gazelle	Gazelle	GZL	Boolean: high growth enterprise younger than 5 years old in year t = 3
MNC	Multinational company	MNC	Boolean
Foreign owned	Foreignly owned firm	FRGN_OWN	Boolean
Exporter	Exporting firm	EXPORT	Boolean
ICT intensive	ICT intensive	INTENS	Boolean
Broadband	Categories of broadband usage	BROADCAT	0: BROADPCT = 0; 1: > 0 to 40%; 2: 40% to < 90%; 3: ≥ 90% ^a
ICT outsourcing	Firm has ICT specialists, provides ICT training, has outsourced ICT functions	ITOUT	ITOUT = 8 - 4*ITSP - 2*ITSPT - XFSP. Eight possible combinations for ICT specialists, training, and/or outsourced ICT ^b
Sharing data	Sharing data on inventories, accounting, production, distribution	LINK	0: no sharing; 1: on inventories; 2: sharing on inventories and at least one other category; 3: sharing data on all categories ^c
e-business	Combinations of types of ICT	BUSORG	BUSORG = 8 - 4*ITERP - 2*CRM - SISC. Eight possible combinations for ERP, CRM, and/or SCM ^d
Product	Product innovation	INPD	
Process	Process innovation	INPS	
Organizational	Organizational innovation	ORGIN	
Innovation combos	Combinations of innovations	INNOV	INNOV = 8 - 4*INPD - 2*INPS - ORGIN. Eight possible combinations product, process, and/or organizational innovation
innov. collaboration	Collaboration on innovation	CO	Boolean

Notes for Table A3:

- a. BROADPCT is the percentage of workers using the internet in a broadband-enabled firm (0 if the firm does not have broadband).
- b. ITSP, ITSPT, XFSP are EC input Boolean variables reflecting ICT specialists, training, outsourcing, see table A4.
- c. LINK is based on EC input Boolean variables SISAINV, -ACC, -PROD, -DIST; and SIPUINV, -ACC.
- d. ITERP, CRM, SISC are EC input Boolean variables reflecting ICT specialists, training, outsourcing ERP, CRM and SCM, see table A4.

Table A4: Input variables in firm-level data source files.

Datasource	Variable	Description
BR	Emp_BR	Number of employees given in Business Register
	Ind_BR	National Industry Code on the Business Register (NACE)
	Ind2_BR	NACE2 code on the Business Register
	Frgn_own*	Dummy for Foreign Ownership
	MNC*	Dummy for Multinational Corporation
	Birth*	Birth year of firm
PS	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
	HKPCT*	pct workers with post upper secondary education
	HKITPCT*	pct workers with post upper secondary IT education
	HKNITPCT*	pct workers with post upper secondary non-IT education
	EXPORT*	Dummy for firm exporter
	NX*	Value of exports (in national currency)
EC	BROAD	Firm has broadband
	AEBUY	Firm orders through internet
	AEBVALPCT	% of orders through internet
	AESELL	Firm sells through internet (or EDI)
	AESVALPCT	% of sales through internet (or EDI)
	IACC	Firm has internet
	EMPIUSEPCT	% of workers with access to internet
	INTRA	Firm has intranet
	EMPINTRAPCT	% of workers with access to intranet
	CUSE	Firm uses computers
EMPCUSEPCT	% of workers using computers	

	WEB	Firm has website
	MOB	Firm has mobile access to internet
	ITERP	Enterprise Resource Planning
	ADEGOV	ADE for sending or receiving data to/from public authorities
	ADENGOV*	ADE for sending or receiving data (not public authorities)
	ITSP	firm employs IT specialists
	ITSPT	firm provides IT training
	XFSP	firm outsources IT functions
	INVREC	receiving e-invoices
	INVSND	sending e-invoices
	SISU	sharing Supply Chain Management (SCM) data with suppliers
	SICU	sharing SCM data with customers
	CRMSTR	share of information with other business functions
	CRMAN	analyse information for marketing purposes
	SISAINV	management of inventory levels
	SISAACC	accounting
	SISAPROD	production or services management
	SISADIST	distribution management
	SIPUINV	management of inventory levels
	SIPUACC	accounting
IS	ENTGP	Enterprise part of a group
	HO	Country of head office
	MAREUR	Other EU/EFTA/CC market
	MAROTH	All other countries
	INPDGD	Introduced onto the market a new or significantly improved good
	INPDSV	Introduced onto the market a new or significantly improved service
	INPDTW	Who mainly developed these products?
	NEWMKT	Did the enterprise introduce a product new to the market?
	TURNMAR	% of turnover in new or improved products that were new to the market
	INPSPD	Introduced onto the market a new or significantly improved method of production
	INPSLG	Introduced onto the market a new or significantly improved logistics

INPSSU	Introduced onto the market new or significantly improved supporting activities
INPCSW	Who mainly developed these processes?
RRDIN	Engagement in intramural R&D
RDENG	Type of engagement in R&D
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)
FUNLOC	Public funding from local or regional authorities
FUNGMT	Public funding from central government
FUNEU	Public funding from the EU
FUNRTD	Funding from EU's 6th or 7th Framework Programme for RTD
CO	Cooperation arrangements on innovation activities
ORGBUP	New business practices for organising work or procedures
ORGWKP	New methods of workplace organisation
ORGEXR	New methods of organising external relations
MKTDGP	Significant changes to the aesthetic design or packaging
MKTPDP	New media or techniques for product promotion
MKTPDL	New methods for product placement or sales channels
MKTPRI	New methods of pricing goods or services
MKTMET	New or significantly changed sales or distribution methods

Notes for table A4: All source files included a unique firm identifier and the reference year, which allows linking the data. Firm-level datasets are used to construct the MMD; they are not at the disposal of external researchers. Not all variables may be available for all countries x industries x years (see Eurostat(2013)). Not all variables in the firm-level input dataset were actually used for the MMD database. '*' denotes variables that may not be directly available from the source, but are derived as a composite indicator and/or taken from a linked source. Different alternatives have been used across countries (see Eurostat(2013)). In general, variable names correspond to Eurostat transmission format names. All monetary variables are expressed in national currency.

Table A5: Distribution Tables CR and ST

Means of:	By Quartile of:	Sample
de dq dv dw e	'prod' ₋₁	PS
de dkl d'prod' dq dv dw	Broadpct ₋₁ ICTi ₋₁	PSEC PSECIS
de dkl d'prod' dq dv dw	Turnmar ₋₁	PSIS PSECIS
HKITpct HKpct kl 'prod' w	Turnmar RDqi	PSIS PSECIS

Notes: The entries show contents of the 'CR' files, with the first column showing 'vname', or the variable for which means are computed, by quartile of 'qname', which is given in the second column. 'prod' stands for LPV, LPQ, TFP, MFP. 'd' denotes growth rate for employment (de); turnover (dq); value added (dv); wages (w, dw). KL is the capital-labor ratio. X₋₁ denotes lagged valued of the variable X. The 'ST' files contain means of all the variables in the table, by their own quartile for the given samples.

Table A6: Industry Dynamics Table

Variable	Description
CHURN	Sum of absolute value of market share changes of firms (turnover Q, or value added V)
OP(prod)	Olley-Pakes cross term: difference between aggregate and average productivity
I(prod)	Aggregate inputs
P(prod)	Average productivity
S(prod)	Standard deviation of productivity distribution
W(prod)	Aggregate, ie weighted average, productivity
Nobs_op	Number of firms used in OP calculation
P(prod)(status)	Average productivity
I(prod)(status)	Aggregate inputs
C(prod)(status)	Productivity contribution
S(prod)(status)	Standard deviation of productivity

Industry classification: EU-KLEMS (SRC = 1) or ALT (SRC = 2)

Prod = LPV, LPQ, TFP, MFP

Status = CO(ntinuing); EN(try); EX(it); O(ne-)Y(ear)

Table A7: Long-Panel Tables

Variable	Description
lpvar	Standard deviations of employment, output, and productivity (only ALT classification)
lpdemogr	Long-panel version of demogr table
Lpit0demogr, lpit1demogr	Long-panel version of demogr table, broken down by ICT-use
lpecst	Long-panel version of ecst table
lpinddyn	Long-panel version of inddyn table
Lpst	Long panel version psst
lpitst	Long panel version psst for selection of ICT intensive firms
lpstat	Long panel version of psstat, breakdown only ICT (BROADPCT and INTENS)

For 38 EUKLEMS industries, and higher aggregates

Figures for total sample and sample splits by ICT from 2001 on.

Countries (ISO-3166-2 codes): DK, FI, FR, NL, NO, SW, SI. 1995 to 2010

Table A8: Observations in the MMD, by table

table	countries	industries		samples	breakdowns	years		
		bottom	/total			average	(min/	max)
PSSTAT	14	29	/54	4	18	10	(2/	16)
ECSTAT	14	29	/54	4	15	9	(1/	12)
ISSTAT	13	29	/54	4	13	8	(1/	9)
Demogr	14	29	/53			11	(2/	16)
Demograge	12	29	/53			11	(2/	16)
Demogragesz	12	29	/53			11	(2/	16)
ECJoint	13	7	/9			8	(3/	9)
ISJoint	13	7	/9			8	(5/	9)
ECISJoint	13	7	/9			8	(3/	9)
PSst	14	2	/16			11	(2/	16)
PSECst	13	29	/38			10	(3/	12)
PSISst	12	7	/9			8	(5/	9)
PSECISst	13	7	/9			8	(3/	9)
PScr	14	29	/36			10	(1/	15)
PSECcr	13	29	/38			10	(3/	10)
PSIScr	12	7	/9			8	(5/	9)
PSECIScr	13	7	/9			8	(3/	9)
Inddyn	14	29	/54			10	(2/	14)
LPStat	7	29	/54	3	2	20	(1/	31)
LPst	7	7	/9			18	(12/	31)
LPecst	7	7	/9			18	(12/	31)
LPdemogr	7	29	/53			19	(1/	31)
LPinddyn	7	29	/54			19	(1/	31)
LPitst	7	7	/9			18	(12/	31)
LPit0demogr	7	29	/53			19	(3/	31)
LPit1demogr	7	29	/53			19	(1/	31)
LPvar	7	29	/54	3	2	20	(12/	31)

Note: For Germany, the innovation data are unavailable. For Finland, information by quartile is unavailable. Number of industries is shown for the lowest level of detail available ('bottom'), and total number of industries (mixed hierarchies). Samples refer to surveys or combinations of surveys (e.g. PS, or PSEC). Breakdowns are splits of the data beyond country, industry, year, see table 2. Number of years refer to averages across country, industry, sample and breakdown.

B Accessing the MMD at Eurostat

Table B1: Relevant Eurostat Website Pages

Information on the MMD: http://ec.europa.eu/eurostat/web/microdata/micro-moments-dataset
An overview of microdata access at Eurostat: http://ec.europa.eu/eurostat/web/microdata/overview
Downloadable public use version of the MMD: http://ec.europa.eu/eurostat/documents/341889/725524/2006-2008-micro-aggregated-data-esslait-indicators-v42-final.xls/026ef74b-e996-4af7-ad58-0fcc57264456
Information on applying for access to the full MMD: http://ec.europa.eu/eurostat/documents/203647/203698/How_to_apply_for_microdata_access.pdf/82d98876-75e5-49f3-950a-d56cec15b896
To check eligibility or to find out how to become recognized: http://ec.europa.eu/eurostat/documents/203647/771732/Recognised-research-entities.pdf/7b59b2fb-d53d-4049-8dc7-42f44a565e70
Information on the ESSLait and ESSLimit Projects: http://https://ec.europa.eu/eurostat/cros/content/esslait_en https://ec.europa.eu/eurostat/cros/content/esslimit-finished_en
Background papers from the Eurostat projects: http://ec.europa.eu/eurostat/web/information-society/methodology http://www.cros-portal.eu/content/esslait