DELIVERABLE

# Validation workshop

## Summary

This deliverable reports on the validation workshop. The objective of the session was to engage scholars to validate the work carried out to extend the CDM model with web-scraped data. The session focused on discussing two deliverables. The first deliverable discussed focused on the updated literature review, on estimations from the econometric model, and on discussing the policy implications, while the second focused particularly on the productivity effects of digitalisation while introducing a novel distinction between process vs. product digitalisation. This deliverable summarises the experts' comments and includes the material presented.

## Deliverable Information

| | |
|---|---|
| **Deliverable number and name:** | **Validation workshop** |
| **Due date:** | 31 March 2022 |
| **Deliverable:** | D29 |
| **Work Package:** | WP5 |
| **Lead Partner for the Deliverable:** | UNU-MERIT |
| **Author:** | Arho Suominen and Matthias Deschryvere |
| **Reviewers:** | Hugo Hollanders and Scott Cunningham |
| **Approved by:** | Arho Suominen |
| **Dissemination level:** | Public |
| **Version** | 19th April 2022 |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

## Disclaimer

This document contains a description of the **BIGPROD** project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium coordinator for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of BIGPROD consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu.int/)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

# Interactive learning workshop

The BIGPROD project's validation workshop focused on WP2 outcomes. The focus was on augmenting the CDM model with web-scraped based variables. Prior to the event two deliverables D4 and D5 were finalized, and the validation workshop focused on sharing the results of these deliverables and offering key stakeholders the opportunity to comment upon the work. The D4 deliverable focuses on providing an updated literature review, on estimations from the econometric model, and on discussing the policy implications, while D5 focuses particularly on the productivity effects of digitalisation while also introducing a novel distinction between process vs. product digitalisation. The presentation, attached as slides in the annex and shared online, discusses these deliverables in detail.

The validation workshop was organized online due to the persisting COVID-19 restrictions. The session was organized on 19th April 2022 at 14.00-15.00 CET. The workshop took place over Microsoft Teams. The project team invited several top scholars that are well-positioned to evaluate the contribution of the project and particularly its relevance to the productivity literature. The invited scholars were:

1. Van Vuong, Maastricht University
2. Jacques Mairesse, CREST-ENSAE
3. Pierre Mohnen, Maastricht University
4. Hans Lööf, KTH Royal Institute of Technology
5. Rene Belderbos, KU Leuven
6. Marco Vivarelli, Università Cattolica del Sacro Cuore
7. Martin Andersson, Blekinge Tekniska Högskola
8. Michele Cincera, Université libre de Bruxelles

From the eight invited scholars five accepted the invitation, but due to sudden changes only three were able to attend, namely Rene Belderbos, Jacques Mairesse and Pierre Mohnen. In addition the session was joined by 10 project team members.

The main comments from the invited scholars focused upon the web-scraped data and its validation. The experts highlighted that if the work were to go beyond an exploratory paper, additional work should be done to get all necessary control variables in the productivity function and to carefully validate the novel web-scraped data-based variables with more traditional variables from statistical offices such as product digitalization measures already provided in selected Community Innovation Surveys. Some of the more detailed comments focus for example on the following:

- As the data has an over representation on Italian firms, further work could indeed focus on the Italian data only and get additional key variables from the Italian statistical office. This additional avenue would be supported by the fact that Italy has been the focus of several CDM research papers, all published based on extensive datasets.

- A strength of the analysis is that it also contains small firms. This can lead to interesting further analysis.

- Several additional indicators should be updated to have fewer missing values or should be additionally collected: the capital measure (available already: total assets variable for 2019 has only 2 641 missing values out of 47 237 observations), the labor measure (firm-level human capital composition measures are not available in ORBIS data), and the innovation expenditures (R&D) measure (coverage is very poor in ORBIS and typically available for big publicly listed firms only).

The discussion also focused on the web-based measures that were created during the project. The experts highlighted that:

- The work should explain the digitalization measures in detail and convince the reader that the measures are sound. In addition,

- the digitalization measures must be further validated with existing data, and

- one should add in case examples of product and process digitalization measures to demonstrate how they are derived and how they might work.

In addition there should be sectoral considerations. For some lagging sectors the digitalization of products may imply a product innovation or the presence of an existing patent while for other leading sectors a product digitalization may have little to do with actual innovation. The analysis needs to be explicit that it does control for sectors or sector averages. The analysis should not mix innovation, patent, and digitalization measures. Potentially the work could also consider including further non-digital measures.

Concerning the econometric analysis, the experts feedback addressed the specification and identification of the econometric model. The feedback particularly addressed the digitalization measure, potential overlaps between what the variables measure and the need to consider lags between innovative activity and when something can be identified in data. The experts also addressed the sufficiency of data and the impact of missing data in some of the sample countries.

Looking at the interpretations, the model finds the same relationship between innovation and productivity as in the traditional CDM cases of process and product innovations. However the process innovation effect is absent or at best unstable; we do not know why this is. Ideally one would have (scraped) information on (process) prices. However the experts realized that this technical approach may not be feasible.

Overall, the experts were satisfied with the work and the critique was constructive in the spirit of advancing the work towards a publication in a high-quality journal. There was a notion that the novel data already gathered made a significant methodological contribution to the literature.

# BIG PROD

## Expert session: productivity, digitalization and innovation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No870822

## Agenda

- Update on the work: scope of D4 (final report) and D5 (scientific paper)
- Presentation of D5: productivity, digitalization and innovation
- Q&A

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

## Update on the work

- Goal of WP2:
  - Understanding the effects of digitalization, open innovation, and servizitation on productivity
  - Using web-scraped novel indicators in a cross-European firm-level dataset
  - Proposed modeling framework: CDM
- Exploratory nature of the work: difficult to foresee before what problems we would encounter
- Using D4 (final report) to
  - fulfill the envisaged project goals and to document the findings
  - single out the most promising venues for D5 (scientific paper)

BIG PROD

VTT PPMi Fraunhofer ISI Maastricht University TU Delft University of Strathclyde Glasgow

## Update on the work

- State of the work in D4
- We implemented a CDM -model consisting of an R&D -stage, an innovation stage (patenting), a productivity stage
- Key-variables:
  - R&D (ORBIS), but very poor quality due to high number of missings
  - Patents (PATSTAT)
  - Labour productivity (ORBIS)
  - Good progress on the web-scraped digitalization, servitization and OI-measures

BIG PROD

VTT PPMi Fraunhofer ISI Maastricht University TU Delft University of Strathclyde Glasgow



BIG PROD

## Update on the work

- Major findings of D4
- Reasonable key-relationships hold (R&D affects patenting positively and patenting is mostly positive for productivity)
- Strong positive direct productivity effects of servitization and product digitalization
- Positive indirect pass-through effects, yet quantitatively small as compared to direct effects

BIG PROD

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow

## Update on the work

- Singling out major venues forward
- Heavy quality issues of the R&D data casting doubts on the strategy of forcing a full CDM-type of model (dropping R&D equation)
- Cooperation variables widely available and used in the CIS -surveys (plus a further in-depth paper in the BIGPROD team)
- Servization measure is interesting but would need more detail to be effective
- Nice differentiation of the digitalization measure into a product and process dimension

BIG PROD

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow



BIG PROD

## Productivity, digitalization and innovation

- The analysis of the effects of digitalization on productivity in many respects mirrors the well-established literature on the productivity effects of innovation
  - Digitalization is usually also an innovation
  - Digitalization like innovation can refer to processes or products
  - Similar econometric methodologies can be implied

## Productivity, digitalization and innovation

- Digitalization as product and process innovations
  - organisation-wide digitalisation of routines and processes (Battistella et al. 2017, Eller et al. 2020, Parida et al. 2019, Annarelli et al. 2021)
  - changes in the product portfolio away from analogous products towards digital versions of them (van Ark 2016, Hatzius et al. 2016)
- Gaps in the literature
  - Mostly conceptual discussions and thus little empirical work on the productivity effects
  - No explicit differentation between the product and the process component

## Productivity, digitalization and innovation

- Results from the innovation literature (see the reviews of Hall 2009, Hall and Mohnen 2013)
  - Overall, positive effects of innovation on productivity
  - Effects appear to be somewhat more stable for product innovations than for process innovations

## Productivity, digitalization and innovation

- A triangular model

$$\log(patint_i) = x_i\beta + \theta \cdot digiproc_i + \psi \cdot digiprod_i + u_i$$
$$\log(prod_i) = x_i\delta + \vartheta \cdot \log(patint_i) + \lambda \cdot digiproc_i + \varphi \cdot digiprod_i + v_i$$

- Possibility to define direct and indirect effects

$$TE(digicap) = \frac{\partial \log(patint)}{\partial digiproc} = \underbrace{\lambda}_{direct\ effect} + \underbrace{\vartheta \cdot \theta}_{indirect\ effect}$$

## Productivity, digitalization and innovation

- Construction of the digitalization measures:

$$prod\_digi\_cont = \frac{n_{digital}}{n_{non\ -digital} + n_{digital}}.$$

$$proc\_digi\_cont = \frac{\sum_{i=0}^{n} x_{digital}^{i}}{\sum_{i=0}^{n} x_{digital}^{i} + \sum_{j=0}^{m} x_{non\ -digital}^{j}}$$

- To level out sector averages, we created dummies indicating of whether the variables are larger than the NACE-2-sector average

BIG PROD

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow

## Productivity, digitalization and innovation

- Estimation procedure inspired by Baum et al. (2017), who used general structural equation model including a latent factor accounting for cross-equation error correlations
- We use a SUR approach, where the latent factor has been estimated by factor analysis
- To-dos: implement a bootstrapping algorithm accounting for the first-stage root-n-consistent estimation error

BIG PROD

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow



BIG PROD

# Productivity, digitalization and innovation

| Variable | Full sample Obs | Mean | Std. Dev. | Estimation sample Obs | Mean | Std. Dev. |
|---|---|---|---|---|---|---|
| Ln labour productivity | 24263 | 3.932 | 0.811 | 16162 | 3.991 | 0.795 |
| Ln patent intensity 2015–2019 | 40445 | 0.003 | 0.055 | 16162 | 0.002 | 0.055 |
| Patent active 2015–2019 | 45795 | 0.014 | 0.119 | 16162 | 0.015 | 0.122 |
| Product digitalisation | 32199 | 0.337 | 0.473 | 16162 | 0.330 | 0.470 |
| Process digitalisation | 45795 | 0.379 | 0.485 | 16162 | 0.360 | 0.480 |
| Servitisation | 45735 | 0.167 | 0.373 | 16162 | 0.246 | 0.431 |
| Neighbours with patent | 45795 | 0.641 | 0.480 | 16162 | 0.664 | 0.472 |
| Ln employees | 40445 | 3.306 | 1.720 | 16162 | 3.555 | 1.676 |
| Ln firmage | 45694 | 3.058 | 0.745 | 16162 | 3.144 | 0.696 |
| Multiple establishments | 45191 | 0.735 | 0.441 | 16162 | 0.751 | 0.433 |
| NACE 21 | 45795 | 0.036 | 0.187 | 16162 | 0.036 | 0.187 |
| NACE 26 | 45795 | 0.153 | 0.360 | 16162 | 0.136 | 0.343 |
| NACE 20 | 45795 | 0.140 | 0.347 | 16162 | 0.145 | 0.353 |
| NACE 27 | 45795 | 0.137 | 0.343 | 16162 | 0.132 | 0.339 |
| NACE 28 | 45795 | 0.387 | 0.487 | 16162 | 0.426 | 0.494 |
| NACE 29 | 45795 | 0.071 | 0.256 | 16162 | 0.070 | 0.255 |
| NACE 30 | 45795 | 0.009 | 0.094 | 16162 | 0.009 | 0.092 |
| AT | 45795 | 0.015 | 0.121 | 16162 | 0.011 | 0.106 |
| BE | 45795 | 0.012 | 0.107 | 16162 | 0.015 | 0.122 |
| BG | 45795 | 0.013 | 0.112 | 16162 | 0.019 | 0.136 |
| CZ | 45795 | 0.033 | 0.180 | 16162 | 0.010 | 0.101 |
| DE | 45795 | 0.164 | 0.370 | 16162 | 0.036 | 0.186 |
| ES | 45795 | 0.100 | 0.300 | 16162 | 0.144 | 0.351 |
| FI | 45795 | 0.025 | 0.157 | 16162 | 0.000 | 0.000 |
| FR | 45795 | 0.073 | 0.259 | 16162 | 0.061 | 0.240 |
| GB | 45795 | 0.049 | 0.217 | 16162 | 0.054 | 0.226 |
| HU | 45795 | 0.018 | 0.133 | 16162 | 0.000 | 0.000 |
| IT | 45795 | 0.330 | 0.470 | 16162 | 0.484 | 0.500 |
| PL | 45795 | 0.061 | 0.240 | 16162 | 0.074 | 0.262 |
| PT | 45795 | 0.012 | 0.110 | 16162 | 0.018 | 0.133 |
| RO | 45795 | 0.015 | 0.121 | 16162 | 0.016 | 0.126 |
| SE | 45795 | 0.043 | 0.203 | 16162 | 0.040 | 0.197 |
| SK | 45795 | 0.013 | 0.114 | 16162 | 0.000 | 0.000 |

# Productivity, digitalization and innovation

| | (1) Ln patent intensity | (2) Ln patent intensity | (3) Ln patent intensity |
|---|---|---|---|
| Neighbours with patent | 0.00114 | 0.00110 | 0.00067 |
| | (1.00) | (0.97) | (0.56) |
| Servitisation | −0.00339** | −0.00363*** | −0.00368*** |
| | (−3.24) | (−3.45) | (−3.49) |
| Ln employees | −0.00049 | −0.00057 | −0.00056 |
| | (-1.49) | (−1.73) | (−1.68) |
| Ln firm age | −0.00036 | −0.00029 | −0.00029 |
| | (−0.54) | (−0.44) | (−0.44) |
| Multiple establishments | −0.00020 | −0.00024 | −0.00027 |
| | (−0.19) | (−0.23) | (−0.25) |
| Product digitalisation | | 0.00274** | 0.00273** |
| | | (2.89) | (2.89) |
| Process digitalisation | | −0.00033 | −0.00028 |
| | | (−0.36) | (−0.31) |
| Constant | 0.01085* | 0.01031* | −0.00044 |
| | (2.47) | (2.34) | (-0.05) |
| Sector dummies | Yes | Yes | Yes |
| Country dummies | Yes | Yes | Yes |
| Latent factor | No | No | Yes |
| N | 16162.00000 | 16162.00000 | 16162.00000 |
| r2 | 0.00426 | 0.00477 | 0.00488 |
| P | 0.00001 | 0.00000 | 0.00000 |

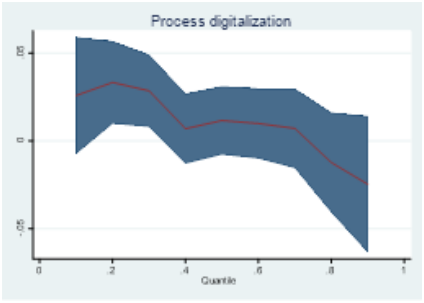
# Productivity, digitalization and innovation

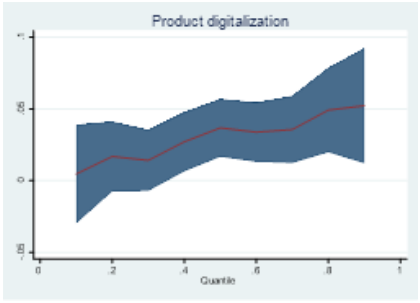
- Direct and indirect effects

| | Direct | Indirect | Total |
|---|---|---|---|
| Product digitalisation | 0.036*** | 0.001* | 0.037*** |
| Process digitalisation | 0.006 | −0.000 | 0.006 |

# Productivity, digitalization and innovation

- Differences across the productivity distribution using quantile regression



# Productivity, digitalization and innovation

- Robustness checks
  - Entropy balancing to avoid selectivity issues coming from ex-ante differences in the samples of digital and non-digital firms
  - WALS model averaging to assess the influence of the control set
  - Assessing the role of the inclusion of different country sets
  - Using larger (unequal) samples across the two equations
  - Using a patenting dummy instead of the patenting intensity

BIG PROD

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow

BIG PROD

Q&A

VTT PPMi Fraunhofer ISI Maastricht University TU Delft Delft University of Technology University of Strathclyde Glasgow

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No870822



BIG PROD

# For more information, please contact

Dr. Arho Suominen (Consortium leader)
Tel. +358 50 5050 354
arho.suominen@vtt.fi

# About BIGPROD

BIGPROD is a research project focusing on Big Data based analysis of productivity using webscraped data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

The project partners in the project are Quantitative Science and Technology Studies team, Foresight-driven Business Strategies, 1) VTT Technical Research Centre of Finland, Competence Center Innovation and Knowledge Economy (Coordinator), 2) Fraunhofer ISI, Economics of Knowledge and Innovation team, 3) UNU-MERIT, Maastricht University, 4) Public Policy and Management Institute, 5) Economics of Technology and Innovations, Faculty of Technology, Policy and Management, 6) Delft University of Technology, Economics of Technology and Innovations, 7) Faculty of Technology, Policy and Management, Delft University of Technology

www.bigprod.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822