TAMÁS BERKI

ICT USAGE INTENSITY IN THE HUNGARIAN CORPORATE SECTOR: STYLISED FACTS ON MICRODATA

MNB OCCASIONAL PAPERS | 149

2023

JUNE

ICT USAGE INTENSITY IN THE HUNGARIAN CORPORATE SECTOR: STYLISED FACTS ON MICRODATA

MNB OCCASIONAL PAPERS | 149

2023

JUNE

MNB Occasional Papers 149

ICT Usage Intensity in the Hungarian Corporate Sector: Stylised Facts on Microdata

(A magyarországi vállalati szektor IKT használati intenzitása: stilizált tények mikroadatokon)

Written by Tamás Berki

Budapest, June 2023

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

H-1013 Budapest, Krisztina körút 55.

www.mnb.hu

ISSN 1585-5678 (online)

Tartalom

Abstract

In this paper, we investigate the information and communication technology (ICT) usage patterns of Hungarian companies, relying on a company-level questionnaire survey conducted in 2020, and combining it with data from corporate annual reports. Our main objective is to construct an index of ICT usage patterns in the Hungarian business sector that reflects the digitisation level (ICT usage intensity) of companies and amalgamates into a single indicator the questionnaire survey information relevant to ICT usage. After aggregating (scoring) the qualitative questionnaire data, we used factor analysis to identify four distinct areas that are associated with the digitisation levels of the different business functions. These are the following: area responsible for integrating business functions, e-administration, IT infrastructure, and marketing and communications. The digitisation headline index is derived from the subindices of these four areas. Our paper also aims to map the patterns of ICT use by company size, sector, productivity and export activity. The intensity of ICT use is closely correlated with company size, export activity and productivity. Companies with more employees or higher productivity tend to be more intensive users of technologies. Firms that also export tend to use ICT more extensively than non-exporting ones. Technology use varies markedly from sector to sector, both in terms of the range of technologies used and their sophistication.

Keywords: use of information and communication technologies, enterprise microdata, questionnaire survey, composite index, factor analysis

JEL-codes: C43, C81, L25, O30

Kivonat

A tanulmányban a magyarországi vállalatok információs és kommunikációs technológia (IKT) használati szokásait vizsgáltuk meg, amire egy 2020-ban felvett vállalati szintű kérdőíves felmérést használunk, összekapcsolva vállalati beszámoló adatokkal. Fő célunk, hogy a magyarországi vállalati szektor IKT használati szokásairól a vállalati digitalizáltság (IKT használati intenzitás) mértékét tükröző indexet konstruáljunk, amely az IKT használatra vonatkozó releváns kérdőíves információkat tömöríti egy indikátorba. A kvalitatív jellegű kérdőíves adatokon aggregálás (pontozás) után négy jól elkülönülő területet azonosítottunk faktor elemzés segítségével, amelyek különböző üzleti funkciók digitalizáltságához köthetőek. Ezek az üzleti funkciók integrálásáért felelős terület, elektronikus ügyintézés, IT infrastruktúra, valamint marketing és kommunikáció. A digitalizációs főindexet e négy terület alindexe képezi. A tanulmány további célja, hogy feltérképezze az IKT használat mintázatait vállalatméret, ágazat, termelékenység, illetve exporttevékenység tekintetében. Az IKT használat intenzitás mértéke szoros összefüggésben van a vállalatmérettel, az exporttevékenységgel, valamint a termelékenységgel is. Minél több főt foglalkoztat, illetve minél termelékenyebb egy vállalat, jellemzően annál intenzívebb a technológiák használata. Egy exporttevékenységet is folytató vállalkozás jellemzően nagyobb mértékben alkalmaz IKT-t, mint egy nem exportáló cég. A technológiahasználat ágazatonként markánsan különbözik mind az alkalmazott technológiák körét, mind azok szofisztikáltságát tekintve.

JEL-kódok: C43, C81, L25, O30

Kulcsszavak: információs és kommunikációs technológiák használata, vállalati mikroadatok, kérdőíves felmérés, kompozit index, faktor analízis

1 Introduction

Over the past decade, the use of information and communication technologies (ICT) in the corporate sector has expanded rapidly, gradually transforming the internal processes of companies and the structure of the economy as a whole. In particular, the internet has brought about radical changes, with the emergence and growth of e-commerce and online networking opportunities such as data analytics, big data, and cloud-based solutions. Initially, these radical changes were limited to a few cutting-edge industries, such as the ICT sector and manufacturing (electronics). Over time, as ICT became more accessible, affordable and widespread, previously lagging sectors found themselves under increasing, partly external, pressure to adopt these technologies, and digital transformation reached and permeated the entire economic sphere. All areas of business processes are becoming increasingly digitalised. The competitiveness of firms is now fundamentally determined by their ability to innovate and their efforts to make the widest and most intensive use of ICT.

In addition to cutting costs, the drivers of the ICT-based transformation of the economy also include other benefits enabled by these technologies. Most important is the demonstrable impact they have on business productivity. The early adoption of emerging ICT solutions can provide a competitive advantage and adopting older technologies now requires less investment, which makes convergence less costly for businesses. By contrast, the rapid spread of new technologies makes their implementation increasingly expensive, which larger and/or more highly digitalised firms are in a better position to finance than smaller firms, which may see the gap widen as a result. Businesses that are at the forefront of digitisation can reap a disproportionately larger share of the rapidly emerging benefits. The implementation of complex and sophisticated technologies (e.g. enterprise resource planning systems) can be cumbersome and costly for smaller businesses simply because of their size.

The progress of digitisation has varied considerably across companies of different sizes. Small and medium-size enterprises (SMEs) tended to delay their adoption of new ICT solutions until the outbreak of the COVID-19 pandemic. Preliminary evidence suggests that the pandemic and the associated lockdowns are most likely to have triggered a shift in SMEs' attitudes to digitisation; they appear to be making greater efforts to catch up and increasingly move their general operations online in order to survive lockdowns and offset the disruptions in supply chains (OECD 2021). Relatively little is known at this point about the ICT usage habits of microenterprises (businesses employing fewer than 10 employees), which presumably lag behind even in comparison with small enterprises. As far as we know, no in-depth survey of this kind has been conducted among them; analyses have tended to focus on SMEs.

In this paper, we attempt to develop a digital maturity index that accurately reflects the differences among Hungarian enterprises in terms of the extent of their adoption of older and newer digital solutions. We provide an extensive overview of the differences in digital maturity and a variety of its aspects by sector, size, geographical location and productivity. The latter is done by the construction of a headline index (indicator) and four subindices, which are generated by aggregating the responses to a questionnaire survey. The questionnaire survey by the MNB and the Budapest University of Technology and Economics (BME)[1] was conducted in two rounds, in July and October 2020, both after the outbreak of the COVID-19 pandemic and contains self-reported responses from a sample of 2,500 companies that are registered in Hungary and employ at least 5 people. Set of firms in the sample is representative of the entire Hungarian corporate sector with at least five employee in terms of geographical location, size (sales revenue) and business sectors.

The design of the indicator(s) will be discussed in detail. We describe the steps in the process of aggregating the quantitative and qualitative datasets of the survey, and the methodology of information compression, including the robustness of the results. First, the data, which are overwhelmingly qualitative in nature, are transformed into numerical data through a scoring process. In many respects, this step may seem arbitrary, as within each subject area (e.g. enterprise resource planning, ERP) decisions must be made concerning the available options such as ranking and preference intensity (how much higher digitisation level does the use of a particular ICT tool represent compared to another tool). The scoring is

[1] The research was carried out as part of the MNB–BME Digital Maturity Project – Survey of the Digital Maturity of SMEs. In the following, we refer to this questionnaire survey as the MNB–BME survey.

fine-tuned according to the degree to which different alternatives are preferred in different constellations – e.g. how much micro and small enterprises are ‘penalised’ because, unlike large corporates, they have simple rather than complex ERP software; by their sheer size, large firms can more easily implement such systems and, in fact, their complex and diversified business processes render this necessary. The indices generated in the scoring process for each subject area (the subject area indices or subindices) are further aggregated by factor analysis. Factor analysis is a data-driven linear aggregation procedure designed to condense the information from the thirteen subject area indicators into a sufficiently small number of factors. The more important and, at the same time, higher-weight factors explain a larger proportion of the variance of the population; these are the areas where there is a relatively large difference in ICT use within the corporate sector as a whole. This step makes visible the previously latent groups that are correlated and may belong together. These groups can be interpreted as subindices. The headline index, also referred to as the ICT usage intensity indicator[2], is derived from the weights assigned to the subject area indices in the factor analysis.

Even though the questionnaire is very detailed and the main index is based on a large amount of data, the resulting composite index is by no means a perfect representation of the digital maturity of the Hungarian corporate sector. We can only look at the use of certain technologies and their intensity, not differences in accessibility (e.g. internet connection speed). Government regulation (e.g. GDPR compliance, mandatory data reporting requirements for e-invoicing) and sector-specific (e.g. e-commerce proliferation) as well as broader trends (e.g. ERP adoption, rise of online payment solutions) may affect different groups of companies differently, and this may have an impact on the intensity of ICT adoption, even at sector level. The questionnaire did not include explicit questions on the use of several major technologies, such as customer relationship management (CRM) and supply chain management (SCM), nor do we have direct information on the specific digital transformation strategies or the product and process innovation efforts of individual companies.[3] We have no information on the management (digital) skills or the digital knowledge of employees. In other words, the composite index relies solely on whether the different technologies are used and with what intensity; an absence of information means that we are unable to examine the impact of ICT use on business processes in greater depth.

The data were collected immediately after the start of the COVID-19 pandemic. As the pandemic became severe, businesses increasingly moved their operations online; it may be assumed that they then introduced efforts to raise their digitisation levels. For a comprehensive understanding of the shift in ICT usage patterns, further data collection will be necessary to analyse the impact of the pandemic. This survey is thus more of a snapshot, capturing the situation right before the outbreak of pandemic; after all, businesses were unable to respond to that shock in a meaningful way in just a few months.

Our paper is divided into three parts. In our summary of the literature, we briefly review the benefits of using information and communication technologies; besides reviewing the literature on the measurement of ICT use (and digital maturity), we discuss the methodology for constructing the composite index.

In the section on data and methodology, we offer a detailed description of our methodology for generating indicators to capture the intensity of ICT use in the Hungarian corporate sector. This includes a description of the questionnaire survey on which the indices are based, the selection of the variables, data cleaning, and the steps of aggregating the transformed data within each subject areas. From the subject area indicators we construct four subindices and the digitisation headline index. We then conduct sensitivity analysis to compare our headline index with an indicator produced using an alternative methodology, and also test the validity of our underlying data.

In the third chapter, we first show the main statistical characteristics of the headline index and the subindices, and the subject area summary indices. Using the resulting indices, we then present stylised facts on ICT usage habits in the Hungarian corporate sector and identify patterns of ICT usage by size, geographical location and sectoral breakdown. To do this, we use aggregation variables (headline index, subindices, and subject area indices) that capture ICT use at different levels of aggregation. The appendix contains the detailed steps of the scoring process, including a description of the questionnaire questions in the 13 subject areas selected.

[2] We also use the term digital maturity (DIG) index.

[3] Under the question on whether the respondents use a particular technology, one option refers to the use of either supply chain management (SCM) or customer relationship management (CRM) or supplier relationship management (SRM).

2 A summary of the literature

2.1 BENEFITS OF INTENSIVE ICT USAGE

Corporate digitisation is a multifaceted phenomenon. In particular, it involves the use of a range of multi-purpose technologies: from making business processes as fully integrated as possible, through improving IT capabilities and user experience, to reaching markets and users as widely and quickly as possible, and communicating more effectively with the latter and with business partners. The integration of technology into business operations is becoming more widespread across the entire corporate sector. Competitive pressures are increasingly strong for lagging firms, providing an additional incentive to exploit the benefits and opportunities inherent in the use of ICT.

The intensity of ICT use by companies can be influenced by their geographical location, size or scope of activity. Their geographical location may translate into less affordable internet access and/or lower bandwidths depending on the supply side, which, in this case, is mainly determined by the number of providers and the competition among them. ICT use tends to vary significantly across the different sectors, and sectoral trends can also differ widely, as is clearly demonstrated by comparing the information and communication sector with agriculture, where the former is at the forefront of pioneering technologies, while the latter is slow to adopt them. The size of a company might affect its ability to implement the latest and more complex technologies, but the introduction of newer ICTs may provide relatively large benefits for small and medium-size enterprises as well.

Broadband internet serves as the foundation for digital infrastructures and is a prerequisite for the efficient exploitation of the potential inherent in ICT tools and for increasing corporate productivity. Equally, basic infrastructure also includes factors such as the availability of computers or portable devices, or the use of mobile internet. Other determinants of digital infrastructure include organisational capacities, human ICT skills within the organisation, and managerial techniques (e.g. Andrews et al. 2018). The lack of a digital infrastructure in this broader sense hinders business innovation, knowledge accumulation (online learning), teleworking, e-commerce, cost-effective business R&D spending and IT development projects.

The main function of ERP (enterprise resource planning) systems is to increase the efficiency of management/governance processes and to promote a strategic planning approach. CRM and SCM aim at embedding customer service/sales and the entire supply chain as fully as possible into business processes. Cloud computing solutions serve the purpose of improving IT systems and expanding their capacities in a flexible manner. Another way to increase efficiency is the use of data analytics (the latest forms of which are big data combined with data and text mining techniques), the results of which may be used in strategic planning and decision-making, administration, logistics, marketing/communications/advertising, etc. Web presence (presence and activity on social media platforms, a dedicated website with sophisticated features, online customer service) helps to get feedback from, and improve access to, consumers, support online recruitment and communication with business partners.

E-commerce can bring tangible benefits especially to the SME sector, as it allows expansion beyond traditional, often geographically limited markets, both for suppliers and consumers. Interaction with consumers is more direct, and business-to-consumer (B2C) transactions are strengthened through online orders and webshop operations, so that B2C relationships, which are less common than business-to-business (B2B) links, are intensified.

The productivity benefits of newer technologies such as data analytics (big data) and cloud computing are widely recognised. By using cloud-based services, businesses may achieve large savings on the costs of building and running an IT infrastructure. In particular, small and medium-size enterprises may benefit because they can dynamically scale up and down computing/storage capacity without the use of human resources (self-service system). In a self-service system, teamwork is easier, multiple projects may be run simultaneously, and there are no capacity bottlenecks that would otherwise be slow, cumbersome and expensive to resolve. Cloud services can be effectively combined with ERP or CRM, which have an efficiency-enhancing role, optimising supplier-customer relationships at both ends of the supply chain (Andrews et al. 2018).

The proliferation of various forms of e-administration (e.g. the "Cégkapu", e-government) contribute to reducing the administrative burden, easing bureaucracy and the complexity of regulation, as well as the associated transaction costs, leaving more resources for focusing on core business processes. The introduction of e-invoicing has strengthened integration between accounting and taxation and generally points towards a greater integration of business processes, reducing administrative costs and ensuring a more stable flow of information among business partners. The wider adoption of online financial solutions is helping to reduce administrative burdens, while integrating corporate financial processes better into the overall administrative process.

The wider and more intensive use of ICT has a positive impact on business productivity. ICT use has the potential to increase productivity for businesses through the digitisation of business processes, which can manifest itself in efficiency gains and cost savings. It also provides access to new markets and new products for better customer relationships and helps automate the supply chain and increase business innovation capabilities. The role of ICT has been examined in many studies, usually in relation to business efficiency and productivity. ICT use alone can have a positive impact on the ability of companies to grow. Clayton and Criscuolo (2002), and Goodridge and Clayton (2004) have identified a positive correlation between ICT use and productivity. Firms that actively use ICT are more innovative than those that do not, which translates into higher innovation activity, productivity and growth. This also requires higher digital literacy and ICT skills among workers (cf. Morandini et al. 2020).

The link between ICT investment and productivity is strong. The use of 'early' technologies such as computers, mobile internet and portable devices had reached maturity by the mid-2010s. Using microdata to investigate ICT use intensity (capital investment per employee, ratio of persons to computers or proportion of people with access to the internet), Farooqui (2005) found that the greatest benefits from ICT investment are achievable in services, and that newly launched firms benefit more from investment projects than companies established earlier in the past. Bertschek et al. (2016) investigate the role of broadband internet access on productivity and innovation dynamics using corporate data from Germany. While no significant relationship was found in terms of the former criterion, a causal relationship was identified between innovation and internet access. It was also Bertschek and fellow authors (2015) who examined the impact of mobile internet on productivity before the technology reached maturity (pre-2014 data) at German companies and demonstrated a strong positive relationship between mobile internet access and productivity. Firms that use e-commerce (mainly B2C) can also benefit from increasing the share of e-commerce sales, especially smaller sized ones (e.g. Falk and Hagsten 2015 or Bertschek et al. 2006), whether they are active in retail trade or other sectors.

The positive impact on productivity of newer and not yet mature technologies, such as cloud, big data, data analytics, and their combination with ERP, CRM, SCM, may already be demonstrated empirically using econometric tools (Andrews et al. 2018). Using company-level panel data across a broad set of countries, Gal et al. (2019) show that the use of digital technologies such as high bandwidth internet access, simple and complex cloud-based services, ERP, and CRM is associated with higher productivity. The positive impact is stronger in manufacturing and in sectors with a higher share of routine/repetitive tasks; digitisation opens up greater scope for streamlining the production/service process and serves in part as a labour-saving function. Cette et al. (2020) used a questionnaire survey and annual report data from French manufacturing firms to investigate the effects of ICT use (cloud, big data) and the use of in-house (or contracted third-party) ICT experts on productivity. Both ICT use and the use of ICT experts are associated with significantly higher productivity and have reduced labour share within total added value.

2.2 MEASURING ICT USE, AND THE CONTENT OF THE DIGITISATION INDICES

To date, relatively few indices have been produced that capture the digitisation level of companies or the use of ICT in the corporate sector. Most of them were published without a specific scientific purpose or ambition. The better-known digitisation indices measure differences in digital maturity between national economies, or perhaps between industries, because this is the only level at which data are consistently available for comparison.[4] Digitisation indices typically do not focus on the corporate sector but try to capture digital development with indicators representing broader perspectives. Even where there is a corporate module, it tends to have a marginal role and, typically, data from questionnaires are not used. An exception is the Digimeter index, which is designed to measure the digitisation of SMEs in Hungary. This is a qualitative questionnaire, representative of the Hungarian SME sector, which measures the digital capabilities of enterprises in six subject areas (Falusi et al., 2021).[5]

The most relevant index for us is the **Digital Economy and Society Index (DESI)**, which was created and is regularly published by the European Commission and measures the digital maturity of EU Member States across five dimensions. The digital technology integration dimension measures the digitisation of businesses, and e-commerce.[6] It captures the level of business digitisation with four indicators covering the more recent technologies described above. (1) electronic information flow, (2) social media presence, (3) big data use, (4) cloud use (at least one out of four).[7] Another dimension that captures ICT use by companies is e-commerce.[8] The weights reflect the priorities set by the EU's digitisation policy. In the digital technology integration dimension, the digitisation level of firms is given a weight of 60 per cent, compared to 40 per cent for e-commerce. DESI also includes a composite indicator measuring the digitisation level of government administration processes relating to businesses (European Commission 2020).[9]

Statistical data collections on ICT use by international institutions are typically based on questionnaires. The DESI index presented collects relatively detailed data for the corporate sector and includes indicators on comparatively new technologies (big data, cloud). It is also on a questionnaire basis that the OECD collects extensive data capturing the measurement of ICT use (OECD 2015). It takes a modular approach, with the basic module focusing on e-commerce (web sales, payment methods accepted, geographical distribution of web sales, e-commerce as a percentage of sales),

[4] The Digital Platform Economy (DPE) Index is a composite indicator for countries that aims to capture the quality/maturity of digitisation in the business ecosystem. It is composed of four subindices (Digital Technology Infrastructure and Business Environment, Digital Skills (Literacy) and Digital Environment), which in turn have 12 pillars; as a result, the Index is calculated from a total of 61 standardised numerical variables. Its main purpose is comparability across countries, and it implements a sophisticated rebalancing procedure in the index production steps, but data aggregation is not a fundamentally data-driven process (Szerb et al. 2022).
The Mckinsey Global Institute (MGI) Industry Digitisation Index compares the digitisation levels of 22 US industries, which are composed of 27 input variables grouped into three categories: digital asset inventory, digital asset usage, and workforce digitisation. The variables are all at the industry (sector) level and tend to be numeric. Assets include indicators such as IT, hardware or software spending, and asset inventory data. Uses include digital payment solutions, customer service, the extent to which back- and front-office solutions are digital, and product development. The digitisation level of the workforce is captured by the indicators of software/hardware expenditure per employee, and time spent working on digital devices. Principal component analysis is used to determine the weights of both the subindices and the digitisation headline index (the principal component with the largest eigenvalue). MGI also created a digitisation index for comparison across countries; of the components of this index, the subindices relevant for our purposes include ICT supply, which captures the level of digitisation in the corporate sector, the innovation subindex, and the business use subindex. The weights are in line with the methodology of the Industry Digitisation Index (Mckinsey Global Institute 2015).
The BBVA DiGiX index, similar to the MGI's digitisation index designed for countries, is based on a large number of countries, using factor analysis mainly of macroeconomic variables (internet access bandwidth, coverage and affordability, quality of e-government, quality of regulation, e-commerce penetration in the business sector and attitude to innovation, internet usage and digital skills across the population). The difference is not only in the indicators, but also in the methodology. A two-step principal component analysis (PCA) is used, where first the subindices are determined using a PCA, and the headline index as a latent variable is constructed as a linear combination of the six subindices using a PCA (Cámara and Tuesta 2017).
The Networked Readiness Index (NRI) commissioned by the World Economic Forum is produced regularly and assesses the readiness of countries to use ICT (Portulans Institute 2020). It ranks nearly 140 countries on how they put the promotion of ICT use at the service of competitiveness and economic growth. The indicators used for constructing the index mostly concern the supply side. They reflect the market, policy, regulatory and infrastructural environment in which ICT is used, and the readiness of key actors to use it. The indicators capturing the welfare-enhancing effects of ICT use in many aspects constitute a separate subindex. Relatively few indicators are used to measure the digitisation of the business sector and there are no detailed indicators on device usage. The subindices are simply averaged to produce the headline index.

[5] The six subject areas are: digital presence (webshop, online customer service), digital everyday life (working from home, online communication within the company), business management (corporate governance system, use of data), sales and marketing (online advertising, customer acquisition tools, digital finance (e-invoicing, online banking), IT security (use of VPN, antivirus software, firewall). The exact methodology for producing the Digimeter index has not been published.

[6] See here for the methodology and database for generating DESI: https://digital-agenda-data.eu/.

[7] It is quantified using the ratio of firms using a given technology to the total number of firms with 10+ employees. The four options of cloud use are database hosting, invoicing, CRM, and the use of computing capacity.

[8] It measures e-commerce by capturing the proportion of SMEs selling online, SMEs with an e-commerce/total turnover ratio of at least 1% and SMEs that also sell cross-border.

[9] This indicator measures the availability of online options for setting up a company and/or conducting regular business activities online, and the availability of related information and help.

infrastructure (computer or internet use, broadband and mobile internet use, remote access), web presence (company website, possibility to order online), information management system (ERP, CRM, EDI[10], RFID[11], SCM, use of e-invoicing), security and privacy (security incidents, ICT security risk policy, collection of personal information for analytical purposes, data protection and data storage). The supplementary module includes e-administration, use of other ICT tools (big data, cloud), ICT skills of employees, ICT spending, social media use, open-source software, corporate feedback module on the impact of ICT adaptation.

ZEW's periodic data collection is considered the most comprehensive in terms of ICT use in enterprises; in addition to questionnaire data on general ICT use, it also tracks the use of specific technologies at a representative sample of firms with at least 5 employees in Germany, supplemented by company-level background information (Bertschek et al. 2018).[12] Overall ICT usage intensity is usually captured in numerical data related to general ICT endowment: percentage of employees working with computers or having internet access, IT capital projects and spending, number of IT professionals, etc. Under specific technology use, the questions concern the use of common technologies (ERP, e-commerce, social media, cloud, big data, mobile application development) and the implementation of Industry 4.0 projects.

[10] Electronic data exchange between computers without human intervention.
[11] Radio frequency identification, a technology used for automatic identification and data communication.
[12] This information is mostly related to the operations, the management, or the innovation or R&D activities of the firms.

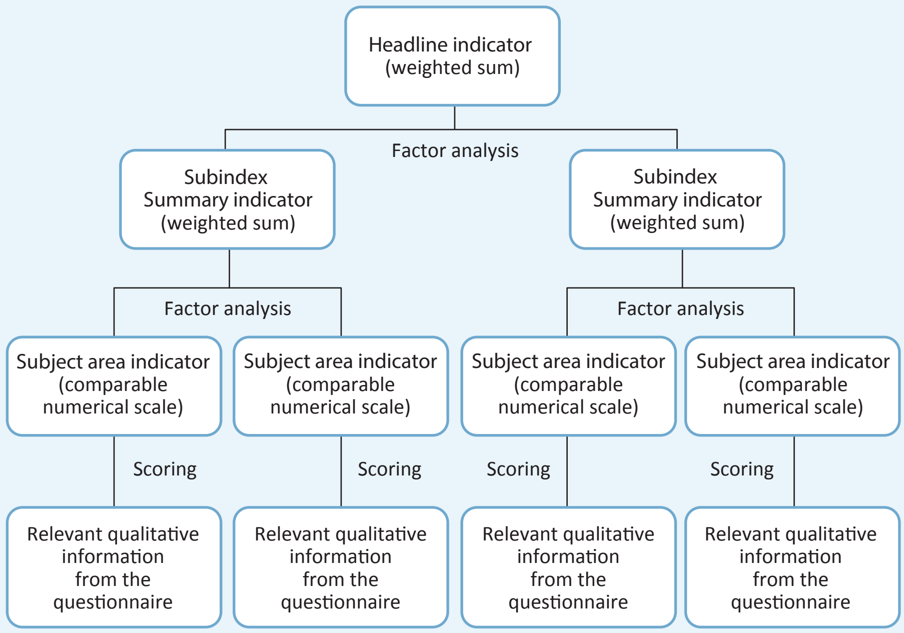
3 Data and methodology

3.1 OVERVIEW OF THE METHODOLOGY OF THE ICT USAGE INTENSITY INDEX

One of the main motivations for this paper has been to produce a composite index of the digitisation levels (ICT use) in the corporate sector based on individual company data collected through a questionnaire survey, using mainly qualitative data. The composite index and its components at different levels can be used to map the differences in the digital sophistication of the corporate sector, supporting policy makers in designing economic policies to promote ICT use.

The ICT usage indicator is based on a common scale that is easy to interpret, allows the ranking of firms and captures differences in ICT usage intensity, ensuring comparability across companies based on the same information base. However, comparing ICT usage intensity in the corporate sector is not a trivial task, because we do not have the quantitative information (R&D spending, IT capital investment and expenditure volumes, internet bandwidth information) that would, in certain aspects, capture the difference in scale better than qualitative data.[13] Nevertheless, qualitative information may also be suitable for this task due to the multifaceted nature of digitisation; this requires the provision of detailed and relevant information on the use of as many relevant technologies and their functions as possible.

Chart 1
The construction of the headline digitalisation index and its subindices



[13] For a significant number of small and microenterprises, detailed balance sheet or profit and loss account data are still missing, as this group is required to submit only simplified reports containing a limited set of data.

The main indicators are generated in four main steps from the questions in the questionnaire (Chart 1):

- first, the raw data (the responses to the individual questions in the questionnaire) are selected, assigned to subject area, cleaned and imputed to fill in the missing data points;
- in a second step, the processed raw data are aggregated (scoring) to produce subject area indices for the 13 subject areas deemed relevant;
- the 13 subject area indices are then grouped according to their correlations using factor analysis;
- finally, principal component analysis is used to estimate the digitisation headline indicator from the four subject area summary indicators (subindices) produced in the previous step.

The first two steps inevitably involve subjective value judgements. Naturally, the selection of the questions/questionnaire subject areas considered relevant is guided by the findings described in the relevant literature, as presented in our literature review. In the second step, aggregation is conducted by scoring, whereby the qualitative questionnaire responses are grouped by subject area and then summarised in a series of steps into a numerical value. This step also partly reflects a subjective assessment. It is a kind of weighting procedure, which embodies our preference intensity by ranking interrelated variables by importance. In the scoring process, we aim to apply consistent principles when designing the scoring criteria for each subject area. The summary indicators (subindices) and the main indicator are estimated in a single step using a multivariate data analysis method, which is essentially a data-driven process.

3.2 DATA SOURCE AND DATA PREPARATION

The questionnaire survey was conducted in two rounds of computer-assisted telephone interviews (CATI) in July and in October 2020. It includes responses from a total of 2,500 companies registered in Hungary with at least 5 employees; the interviews were taken with the CEOs or other senior managers of the companies.[14] Identical questions were asked of all the selected businesses. Besides the questions on ICT usage, additional information is also available, such as data from the companies' annual reports for the period 2017–2019, including the most important information in their balance sheets and income statements, the number of employees, the postcode of their head office , their four-digit sector classification by main activity (NACE Rev. 2), their year of establishment and export sales figures. We do not use external data sources that capture ICT usage specifically, because these are not yet evenly available for individual firms (e.g. IT capital investment volumes, software acquisition budgets, etc.).

In the quota sampling, the quotas were defined by size (number of employees), geographical region and sector, while in weighting, the individual weights were adjusted to the main headings of the Hungarian company structure, as described above (representative by sales revenue, sector, headcount and region).[15] In the sampling procedure, larger companies with 50–249 employees and corporations with at least 250 employees are oversampled to ensure that, given the pyramid shape of the Hungarian corporate structure in terms of size, as many large employers as possible are included in the sample, thus providing a more accurate picture of ICT usage patterns. The descriptive statistics used in this paper for the continuous and category variables are enclosed in the Appendix (Table F1).

We have found the 13 subject areas and most of the questions within them to be suitable for constructing an ICT usage intensity indicator.[16] In all areas, the questions are very closely related, mostly building on each other, so reclassification of very many elementary questions into subject areas added little value. Some questions in the subject area on the

---

[14] The data was collected on behalf of the MNB by internet research and consulting company eNET Internetkutató és Tanácsadó Kft.

[15] Geographical region is determined by the location of the company's head office, and sector is determined by the main activity declared by the company. The sectoral breakdown is representative in terms of two-digit NACE codes.

[16] The questionnaire also included a section on online information gathering and communication, which asked about the extent and method of using company databases/company information websites. These were considered irrelevant to our focus on ICT use. We also excluded participation in government programs related to digital development and the coronavirus block (changes in digital practices and teleworking), as they are not relevant to our examination of the extent of ICT use. In the latter subject area, all the companies were asked again after the outbreak of the pandemic what measures they had taken to improve the level of digitisation and what they expected as a result of COVID-19.

use of online financial services were removed because they either had too high non-response rates or did not explain substantively the differences between companies in relation to the variables retained within the subject area.[17] We also excluded from our analysis the some questions on e-commerce, partly due to the high non-response rate and partly due to the particularly specific nature of the subject area.[18]

Data cleaning is the process of checking the consistency of the data, which in practice only corrects for the erroneous values of a small number of numeric variables. Further, we classified the manual responses, if any, in the 'other' option of multiple-choice questions into one of the response options wherever possible; elementary variables (none, don't use/don't know) were dropped if they became redundant after the appropriate variable transformation. Accordingly, the number of transformed and selected variables was significantly lower than the number of elementary variables originally included in the survey.

We decided to impute missing data to avoid losing observations/variables. This was possible because the proportion of missing data around the selected questions did not exceed 10 per cent of observations for any single subject area (Table 1), and the data gap was partly random in subsamples of 35–40 variables based on the Little MCAR test.[19] Our data imputation method was multiple imputation, which is a common and sophisticated model-based statistical technique for data imputation.[20]

Table 1
Number of variables within subject areas, and imputation

subject areas	number of elementary variables	variables after selection and transformation	total imputed data points	average imputation rate (%)
Basic ICT endowment	11	7	263	1.5
E-government, e-administration	8	5	1,040	8.3
Online financial services	31	13	1,810	5.6
Web presence, online marketing	11	6	224	1.5
Enterprise resource planning	8	6	0	0
Digital HR practices and systems	2	2	0	0
Cloud usage	7	4	800	8.0
E-commerce	35	3	85	1.1
E-invoicing	9	3	226	3.0
Data analysis/Big data	14	10	246	1.0
ICT security software	7	4	0	0
ICT expertise	7	5	146	1.2
UX maturity	3	3	0	0
Total	**153**	**71**	**4,840**	**2.7**

[17] These questions were the following: the number of business bank accounts, future cooperation with fintech companies, the use of credit calculators.

[18] The main profile of around 15% of the companies in the sample is retail or wholesale trade, and companies in other sectors showed little such activity.

[19] Missing Completely at Random (MCAR). The Little MCAR test (Nainar package of R statistical software) does not give plausible results when run on all variables at the same time. The significance of the test statistic averaged around 5% across the subsamples tested, i.e. there is weak evidence that the data gap is entirely random rather than random and systematic.

[20] The MICE software package of the R statistical software was used to replace the missing data. When imputing, quantitative variables for all ICT usage were used in addition to the additional information already mentioned. In the stochastic imputation of each variable, the goodness of fit of the predictions obtained by the different regression procedures and the convergence of the algorithm were prioritised. In addition to the survey variables, we used a large number of variables from the annual reports in our prediction, which mitigates the potential distortion due to non-random data gaps. For more details on the procedure in general, see van Buuren and Groothuis-Oudshoorn (2011).

The subject areas selected for the interviews are:[21] **General (basic) ICT endowment** (1); **E-government, administration** (2); **Use of online financial services** (3); **Web presence** (4); **Enterprise resource planning** (5); **Digital HR practices and systems** (6); **E-invoicing** (7); **Data analytics/Big data** (8); **Cloud usage** (9); **E-commerce** (10); **Use of ICT security software** (11); **Use of ICT expertise** (12); **UX maturity**[22] (13).

Most of the subject areas cover the digitisation level of the business functions of companies, covering a relatively broad range of domains. The thirteen subject areas covered in the questionnaire survey can be grouped into four categories according to the business function to which they are related. The classification is based on the OECD's groupings of statistical indicators on ICT use (OECD 2015, 2021). The questions on ICT use in the MNB–BME questionnaire survey typically show a high degree of similarity with data collected by international organisations such as the OECD or the European Commission (DESI).

The classification of subject areas into business functions also determines the meanings of the summary indicators. The **area primarily responsible for the integration of various business functions** includes the technologies used for information management purposes, such as enterprise resource planning systems and strategic planning. The specific tools related to this are ERP, SCM and, in part, CRM, as well as other strategic areas such as HR. The use of data analytics (+big data) and cloud-based systems[23] is also included here, as they support decision-making and strategic planning directly. Of the 13 subject areas, included here are **enterprise resource planning systems, digital HR systems and practices, data analytics and cloud-based services**.

The second group includes the following subject areas: **e-government administration, e-invoicing and online financial services**; collectively, we call these **e-administration**. E-administration may be seen as a specialised business function that helps companies operate cost-effectively by integrating accounting, invoicing and tax administration, complemented by the digitisation of financial services.

The **marketing/advertising/communication** group includes the **web presence**, **e-commerce** and **UX maturity** subject areas of the questionnaire.[24] The digitisation of this area is a low-hanging fruit for smaller firms because it may improve the quality and reduce the time requirement of engagement with suppliers and customers, result in a substantively wider range of business relationships that are less limited geographically, and allow for more cost-effective advertising campaigns.

The last group includes **basic ICT endowment**, the **use of ICT expertise** and the **use of ICT security software**. These sub-areas are mainly intended to capture the quality of the corporate **IT infrastructure**, which is also part of the general governance and IT business function area.

3.3 AGGREGATION WITHIN SUBJECT AREA: SCORING

We define data points as the information content of the standardised answers to the different elementary questions. The questions in the survey on ICT use can be either single-select (yes/no) or multiple-choice questions.[25] The questions build on each other, i.e. for the single-choice (usually yes/no) questions, a positive answer is followed by further questions to find out more about the given area of technology. Most data points provide qualitative information. We therefore decided to convert into category variables only a negligible number of numerical data points. Factor analysis of mixed data may

[21] The questions in the 13 subject areas are described in more detail in the Appendix (Tables F2–F14).

[22] UX: User eXperience

[23] The use of cloud services (for day-to-day operations, computing capacity, hosting, etc.) also supports the integrated and cost-effective operation of the previous three areas and serves general governance/administrative processes and ensures the flexibility of IT capacities. Despite it being a general-purpose ICT tool, we find that the cloud makes a greater contribution to the integration of business functions and is even able to partly replace them where no enterprise resource planning software is available. General business administration and IT systems are related to business functions in the OECD's categorisation (OECD 2020), but in this paper we subdivide this function to achieve a larger number of subject area groups.

[24] Even though CRM also falls under this category, we do not include it here because the questionnaire puts it in the supply chain (SCM) / customer relationship management (CRM) / supplier relationship management (SRM) answer choice. We allocated it to the area responsible for integrating business functions.

[25] For questions where the other response option was offered and a unique response was received, we assigned the latter to one of the predefined possible options.

be a possible option, but it is difficult to implement and interpret if the weighting is also considered. Since we have more than 70 elementary variables, dimensionality reduction does not work effectively enough to explore the relationships between categorical variables using multiple correspondence analysis.[26]

Scoring involves transforming the raw qualitative information from the survey by assigning numerical values to each response option (usually a positive or negative response) and then aggregating these numerical values to the level of subject areas closely following the structure of the questionnaire survey. A total of five questions included in the analysis are non-qualitative; these were normalised into interval scaled numerical variables and categorised. For multiple-choice questions, the response options were generally treated together and assigned a numerical value based on the number of options for which the given company answered in the affirmative. Most subject area indices are normalised on a common 0–6 scale, each reflecting the intensity of use in a particular ICT/digitalisation area.

The scoring procedure was carried out according to three criteria, following the widely cited study by Nicoletti et al. (2000). A company using no ICT at all is scored zero, while a company using all its functions is scored 6. A discrete scale of 0–6 points was considered sufficient to capture the variance of the raw data as much as possible *(variance-preserving scoring)*. The binary categories of yes/no questions are converted into cardinal values by treating related questions together *(categorical scoring)*. *Hierarchical scoring* assumes the hierarchy between questions on a given use of ICT, i.e. it fine-tunes the scoring of ICT usage intensity based on the responses to related but subordinate questions. In this fine-tuning, we also consider the constraints imposed by company size. For example, micro and small enterprises are not 'penalised' with substantively lower scores for not using an ERP solution, only a simpler enterprise management system. Similarly, for multiple-choice questions, we often do not consider proportionally if a smaller company uses fewer features than a larger company.

Variance-preserving scoring is used only for a marginal part of the data set, a total of five questions for which numerical answers were expected. While the allocation of numerical values to categories may seem arbitrary, we have tried to establish category boundaries that reflect the variance of the original data set, following a similar principle for all the variables involved.

Following Table F1 in the Appendix, the multiple-choice questions on data analysis/big data were converted to single-select ones because the dispersion of the data allowed it: there are very few firms using multiple types of data processing, just as there are very few that do so on multiple types of data. Those that do not use any data processing technology were only asked if they buy data from other entities. The maximum score of six points is given to companies with positive answers to all the questions, and half the score is awarded if a company uses some kind of data analysis/data processing technology.

In the case of categorical scoring, questions that are grouped together by subject area are aggregated for scoring. In Table F8 in the Appendix, when scoring the answer choices for the multiple-choice question on the use of cloud computing services, only the number of 'yes' responses in any combination is considered. Here, we deliberately do not distinguish between the response options according to how sophisticated the technology is, to compensate the less digitised and smaller firms at the expense of companies that, typically due to their larger size, find it easier to adopt technologies.

For many subject areas, we use a combination of hierarchical and categorical scoring. We score each group of questions separately, either hierarchically or categorically, and then aggregate the sub-scores using a new weighting system based on the importance of the questions, in which we have drawn on the literature. This is the case, for example, for the subject area of general ICT endowment (see Table F11 in the Appendix), where, after converting the numerical variables into categorical ones, we assigned different weights to the questions, reflecting the relative importance of the areas and the maturity of the technology[27], using a mix of categorical and hierarchical scoring.

We have been consistent in collecting, scoring, normalising and aggregating the raw data across all subject areas and groups of questions, following the principles set out above. Scoring will inevitably include subjective elements such

[26] The number of relevant dimensions is too high due to too many categorical variables, making it difficult to interpret the information or assess the associations between variables, and to extract weights.

[27] Computer / internet / mobile device and mobile internet usage among employees: 40-20-10%.

as expert judgements about weighting questions within the same subject area, or certain simplifications in variance-control scoring, which do not substantially reduce post-scoring variance compared to the raw data, and there may also be measurement errors in the raw data. This is most likely to be due to the misinterpretation or misunderstanding of the question by the respondent when completing the questionnaire, especially if the interviewer was not competent enough on certain topics to provide meaningful help on the response options. The impact of potential distortion caused by measurement errors can be reduced by the fact that the questionnaire survey was conducted in the same way for each company. Also, the database was compiled on a large number of companies. By contrast, distortion impacts may be intensified by the fact that the respondents were different officers/experts of the companies, so the information on ICT use within the company may not be of a consistent quality.

3.4 AGGREGATION OF SUBJECT AREA INDICATORS

In the scoring process, the indicators generated per subject area are further aggregated into summary indicators (subindices) (of which there are four in total), and then into a headline indicator measuring ICT usage. The methodology used in this chapter also closely follows the procedure used by Nicoletti et al. (2000) to produce a product market and a labour market indicator for OECD countries, based solely on the degree of market-friendliness of public regulation in these two areas.

Factor analysis, a multivariate statistical procedure is used to aggregate the subject area indicators. Factor analysis/principal component analysis are popular methods for constructing composite indices. Their main purpose is dimensionality reduction, to compress as much information as possible into as few components as possible, while at the same time ensuring objectivity and transparency in the application of the method.[28] The essence of this data-driven method is that it assigns to each component of the summary and headline indicators a weight corresponding to its contribution to the variance of the underlying data set. Factor analysis can be used to reveal the structure of the data set and the factors that are not directly observable. Within each factor, the individual indicators are weighted in proportion to the variance of the subset of the corporate database explained by the factor. These factors are mostly sub-areas of ICT use that are well interpretable in terms of the business functions of companies. Each enterprise is ranked on each of the identified factors using estimated weights. Factor-specific scores are used as input for the summary indicators. Each factor contributes differently to explaining the variance of the company data, but it is sufficient to consider only the few most explanatory components, given that they already cover most of the variance. The relative contribution of the few factors considered to the total variance forms the basis for weighting each summary indicator.

Factor analysis is a data-driven procedure, as it assigns weights based solely on the contribution to the total variance of the raw data when decomposing into components. It assigns the greatest weight to the components that show the greatest variation across companies for a given set of ICT use, so that it is independent of any *a priori* assumptions about the relative importance of each component. The only *a priori* weighting consideration is the weighting system intended to ensure company representativeness, which is used as an input in factor analyses. In our case, the factor analysis reveals dimensions of ICT use that are highly heterogeneous across the business sector as a whole, whereas elements of ICT use that are qualitatively very similar across the business sector are of little interest because they are unlikely to explain differences in, for example, company (labour) productivity.

Factor analysis is sensitive to input data, especially the presence of outliers and extreme values that can cause spurious correlations. However, it does not require normality of the data, nor are there any small sample problems in this analysis. The robustness of the results should be examined in any case, therefore we also use alternative aggregation procedures.

The procedure for principal component analysis[29] consists of the following steps:

- Our main goal is to reduce the number of input variables (dimensionality reduction) in such a way that the information contained in the raw data is not significantly reduced and the retained factors can be interpreted. The observed variables

[28] Typical methods of composite index construction in terms of weighting, aggregation and robustness are reviewed in detail for instance in Greco et al. (2019).
[29] A special case of factor analysis.

must be correlated, otherwise the dimensional reduction will not work, they will not have a common factor. Bartlett's test is used to test the correlation of indicators constructed using scoring.

- The second step is to extract the factors and determine the number of factors extracted. Each factor is described by a set of coefficients (factor weight) that measures the correlation between the observed variables and the latent factor. Principal component analysis is used to extract the factors.[30] Principal components are linear combinations of the variables included that are uncorrelated, with the first principal component describing the largest fraction of the variance squared in the database, the second principal component describing the second largest fraction, and so on.

- The third step is the rotation (orthogonal rotation) of the factors (principal components). The goal of varimax rotation is to produce as many factor weights close to zero as possible, or in other words, to minimise the number of observed variables that have a high factor loading on the same factor. This procedure simplifies the factor structure and improves the interpretability of the factors.

- Fourthly, the weighs needed to produce the summary indicators and the headline index is spelled out. Each subject area indicator is assigned a weight per factor that is proportional to the square of the normalised factor loading associated with it, while each factor is weighted according to the proportion of variance it explains in the database.

In the following, the aggregation/loading procedure using factor analysis/component analysis is presented. First, the subject area indicators are aggregated into summary indicators based on the four groupings presented above.

Calculation of the Kaiser-Meyer-Olkin (KMO) criteria for the 13 subject area indicators is used to assess the adequacy of the sample. Conformity refers to the meaningfulness of the dimensionality reduction. In our case, KMO is 0.85, which makes the database suitable for factor analysis/principal component analysis because of the high correlation of the variables. The same is confirmed by the value of Bartlett's chi-square test measuring the correlation of the variables, as the hypothesis of the independence of the variables is rejected.

Consistent with best practice, we focus on factors that preferably have an eigenvalue greater than 1 (the so-called Kaiser criterion), explain at least 10 per cent individually and at least 60 per cent cumulatively of the total variance in the data set (see e.g. OECD and JRC-EC 2008). These conditions are applied flexibly because the cross-section of the database (n = 2,500) is much larger than the number of subject area indicators (13), and the high heterogeneity of the corporate sector means that the degree of dimensional reduction is far from being significant (above 60 per cent cumulated) for the first few factors/main components.

Table 2
Factor analysis: eigenvalues and variance

Factor	Eigenvalue	Explained variance (%)	Cumulative explained variance (%)
1	3.29	25.3	25.3
2	1.17	9.0	34.3
3	1.03	7.9	42.2
4	0.92	7.1	49.3
5	0.90	6.9	56.2
6	0.85	6.6	62.7
7	0.80	6.2	68.9
8	0.76	5.8	74.7
9	0.74	5.7	80.4
10	0.70	5.4	85.8

Factors were extracted using principal component analysis.

[30] Factor analysis and principal component analysis give roughly the same results for factor extraction.

In Table 2, the thirteen subject area indicators correlate best with the first four factors. The variance explained by the second factor is already below 10 per cent, although its eigenvalue is still above one. The same is true for the third factor. Overall, the explained variance is low from the second factor onwards. The decrease in the slope of the curve as the number of factors increases stops almost completely after the fourth factor. This is a sign that certain subsets of the variables are weakly correlated. An argument in favour of retaining the first four factors is the fact that they condense nearly 50 per cent of the information in the original database into just four factors, of which the fourth has an eigenvalue close to one (OECD and JRC-EC 2008).

The varimax rotation ensures that the information compressed by the new factors is not redundant, the total variance explained does not change, but the eigenvalues do. The squares of the high factor loadings will be close to one and the low ones close to zero after rotation, resulting in better interpretable factors.

The extracted four factors after varimax rotation represent the summary indicators (subindices), where the weights of the subject areas are proportional to the normalised loading squares. For each factor, the highest weighted subject area indicators that determine the interpretation have a combined loading of at least 70 per cent.

Table 3
Detailed result of factor analysis

	Factor 1		Factor 2		Factor 3		Factor 4	
Interpretation	**Integration of business functions**		**IT infrastructure**		**E-administration**		**Marketing and communications**	
	factor loading	**indicator weight in the factor**	**factor loading**	**indicator weight in the factor**	**factor loading**	**indicator weight in the factor**	**factor loading**	**indicator weight in the factor**
ERP	0.49	0.14	0.45	0.12	0.13	0.01	0.19	0.02
data analytics	0.74	0.32	0.00	0.00	0.11	0.01	0.03	0.00
digital HR	0.62	0.22	0.28	0.04	0.04	0.00	0.02	0.00
cloud	0.43	0.11	0.05	0.00	0.37	0.09	0.19	0.02
ICT security	0.12	0.01	0.75	0.31	0.10	0.01	0.01	0.00
ICT expertise	0.41	0.10	0.54	0.16	-0.03	0.00	0.08	0.00
gen. ICT endowment	0.01	0.00	0.60	0.20	0.28	0.05	0.22	0.03
e-administration	0.02	0.00	0.23	0.03	0.64	0.28	-0.07	0.00
e-invoicing	0.09	0.00	-0.09	0.00	0.69	0.32	0.07	0.00
online fin. serv.	0.15	0.01	0.27	0.04	0.55	0.21	0.10	0.01
web presence	0.01	0.00	0.38	0.08	0.12	0.01	0.61	0.25
e-commerce	0.02	0.00	0.03	0.00	-0.01	0.00	0.83	0.46
UX maturity	0.38	0.08	0.02	0.00	0.06	0.00	0.53	0.19
Factor weights		27.8%		23.0%		22.7%		26.5%
Eigenvalue before rotation		3.30		1.17		1.03		0.93
Variance explained by factors: 49.35%								
KMO value		0.86						
Bartlett test	Chi-2	4,796						
	df	78						

In the principal component analysis, the factors are found to be well interpretable after rotation (Table 3). ERP, data analytics, digital HR solutions and the cloud take the highest loading value on the first factor with a weight of 28%, meaning that these subject areas are the most correlated. Initially under the area responsible for **integrating business functions**, these technologies were allocated to a single factor; these four subject areas account for almost 80 per cent of the factor. ICT security, ICT expertise and basic digital coverage were also allocated to a single factor after rotation, with a weight of almost 70 per cent on the factor. This component can be interpreted **as IT/ICT infrastructure** and coincides exactly with the preliminary grouping. The combined weight of the e-government, online financial services and e-invoicing subject areas is above 80 per cent on the third factor, which may be described as **e-administration**. Also belonging together are web presence, e-commerce and UX maturity on a 27-per cent factor, which is the **marketing/communications/advertising** business function; their combined weight is 90 per cent on the factor. Overall, the subject areas expected in advance were allocated to the same factor in the principal component analysis.

The second largest and still significant factor loading value for the cloud is found in the e-administration factor. This may be related to the fact that the cloud is a more general-purpose ICT tool, which can serve and supports other areas as well, as it is more cost-effective for smaller businesses to use (e.g. using Gmail for corporate e-mailing) than to make the same investment in-house. Cloud use may be seen as providing more of the basic infrastructure for enterprise resource management and also for other areas such as general administration. This is evidenced by the fact that the cloud subject area is present in both factors, with a share of at least 10 per cent. Similarly, the weight of ERP is high in the IT infrastructure factor and ICT expertise in the integration of business functions factor.

3.5 ROBUSTNESS AND VALIDITY TESTING

A composite index of digitisation that is not directly observable should preferably condense as much information as possible from each of the basic indicators and should be balanced, i.e. not biased towards one or more highly correlated variables. Principal component analysis tends to overweight indicators with a high correlation. Therefore, we also use an alternative weighting and aggregation method to compare the results with the one-step principal component analysis described above, using the same thirteen subject area indicators.

Essentially, it is assumed in this methodology that there is a latent structure for the diverse set of indicators included in the model, and the latent variables may be expressed as a linear combination of the indicators (Nagar and Basu 2002).[31] We perform principal component analysis in two steps (alternative PCA). First, principal component analysis is carried out separately for each of the four pre-identified subject areas, where factors are extracted according to the principles presented. First, as a result of the principal component analysis performed on all four indicator groups, only one factor may be identified and extracted (the eigenvalue of the factors is greater than one) in line with standard practice, and these factors condense between 45 and 53 per cent of the total variance of the subject area indicator groups. In a second step, principal component analysis is performed on the four summary indicators generated; the factor loadings of the principal components generated determine the weighting scheme applied to the four summary indicators and, at the same time, the principal digitisation index. In this step, we only considered the first principal component, which explained 50 per cent of the total variance.

Table 4
Final weights of the headline index in the different procedures (%)

subject areas	baseline PCA (1)	alternative PCA (2)	difference (2)-(1)
enterprise resource planning (ERP)	7.4	7.7	0.3
data analytics	9.1	7.1	-2.0
digital HR	7.2	6.6	-0.7
cloud	5.9	6.2	0.3
integration of business functions	**27.8**	**27.5**	**-0.2**
ICT security	7.6	9.2	1.6
ICT expertise	6.6	9.2	2.6
basic ICT endowment	6.8	9.2	2.4
IT infrastructure	**23.0**	**27.6**	**4.6**
e-administration	7.2	7.7	0.6
e-invoicing	7.6	6.5	-1.2
online financial services	6.2	8.1	1.9
e-administration	**22.7**	**22.3**	**-0.4**
web presence	8.7	7.8	-0.9
e-commerce	12.2	8.1	-4.2
UX maturity	7.4	6.8	-0.6
marketing/communications	**26.5**	**22.6**	**-3.9**

[31] Latent variable models assume that the latent variable can be written as a linear combination of variables included in a factor analysis and an (orthogonal) error term, where the estimated parameters of the explanatory variables represent the variable weights.

There is a difference between the results of the two procedures, with the weight of IT infrastructure increasing, within which all three subject areas increase substantively, while marketing/communications decreases, including a significant decrease in e-commerce (Table 4). Nevertheless, when we arrange in deciles the company headline index values for both methods and look at which deciles the companies fall into using the two methods, we find that firms change their positions by at most one decile, with a slight, 7 per cent changing their decile position on average (see Chart F1 in the Appendix). In the middle deciles, the index values situated within a very narrow range, which means a higher probability of a shift in decile.

The validity of the questionnaire data used can be tested by comparing them with relevant aggregate data from the DESI measuring the digitisation level of the corporate sector. Comparing the latter with the indicators of the MNB–BME questionnaire survey with similar information content, it can be concluded that they are close to each other (Table 5). Overall, compared to similar indicators in the DESI, the variables in the MNB–BME survey indicate a wider use of ICT, especially among small businesses.[32] While the differences between the results of the surveys are not negligible, the differences in company size within the surveys are almost identical.

Table 5
Comparison of the answers to the common questions in MNB–BME and DESI

	MNB–BME	DESI	DESI			MNB–BME		
Percentage of users (%)	**at least 10 people**		**10–49 persons**	**50–249 persons**	**250+ persons**	**10–49 persons**	**50–249 persons**	**250+ persons**
ERP use*	27	14 (21)	10 (15)	32 (46)	62 (80)	23	45	60
big data use**	13	7	6	11	19	12	20	31
cloud usage (any)	32	25	22	37	59	31	33	57
ICT specialist employed	24	29	24	50	83	19	43	79
own website/webpage	83	63	60	78	86	81	90	92
social media presence***	55	38	34	46	67	55	52	73
web sales	16	14	13	16	18	16	13	11

** data for 2019 (2021); ** MNB–BME: data analysis use; *** DESI: data for 2019.*

[32] The Hungarian data in DESI were collected by the Hungarian Central Statistical Office (HCSO) on a sample of around 6,500 (see European Commission 2021). It is important to note that the DESI indicators measuring ICT usage rates in the corporate sector may vary significantly from one year to the next. The MNB–BME survey was carried out in July and October 2020 and the HCSO survey in early 2020; the impacts of the COVID-19 pandemic may have been reflected in the data to a greater extent. Besides their different dates, the two surveys also differed in their sampling procedures and the exact wording of the questions. Company-level data from the HCSO survey is not available.

4 Stylized facts: ICT use in the corporate sector

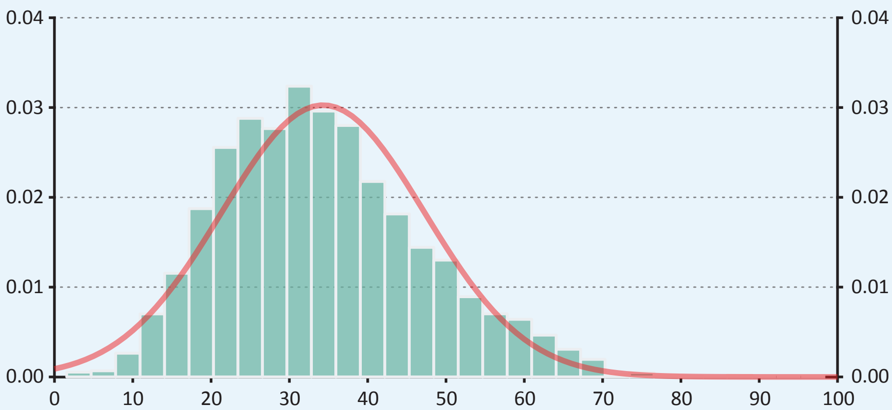
In this chapter, we explore the main patterns of ICT usage intensity in the Hungarian corporate sector in terms of size, productivity, geography and industry. The study of these factors offers significant new information, as there have been no systematic surveys of ICT use conducted in Hungary using company-level data. The intensity of ICT use is of primary importance for the future performance of firms, as the use of newer technologies can help them increase their productivity and boost their competitiveness. Further, we explore where ICT use differs significantly by company size and whether the ICT usage patterns of exporting firms diverge from those of non-exporting firms. We also look at whether younger businesses use ICT technologies more extensively than the ones founded earlier in time.

4.1 CORPORATE DIGITISATION HEADLINE INDEX, SUB- AND SUBJECT AREA INDICES

In this section, we briefly present the main digitisation indicator, as well as the underlying summary and subject area indicators. As our sample size is relatively large, we use histograms, boxplots and radar charts to visualise the key results for the whole population. The data used in the following as a basis for the diagrams are weighted, so they are representative of the Hungarian corporate sector and its subsets. The headline indicators and subindices can take values on a scale of 0 to 100.[33]

Chart 2 shows the distribution of the headline indicators with a fitted normal distribution. The final digitisation indicator follows a near-normal distribution, but standard tests (e.g. Shapiro-Wilk) do not confirm this,[34] which is caused by the boundedness of the indicator and partly related to this, by a slight right-skewedness.

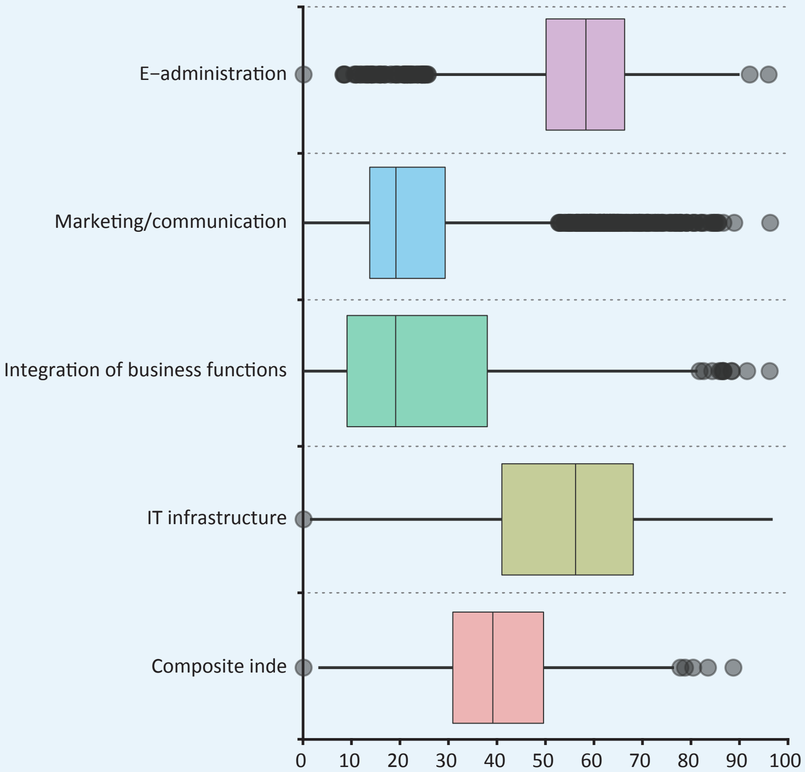
Chart 2
Distribution of digitisation headline indicator in the weighted sample



[33] The headline index is zero for a company that does not use any of the technologies or any of their functions at all, and 100 for a firm that fully uses all the technologies.

[34] Taking the natural logarithm of the headline index, its distribution is still not a normal distribution, but the test statistic suggests that it is closer to that than the raw data.

Chart 3
Distribution of subindices and the headline index

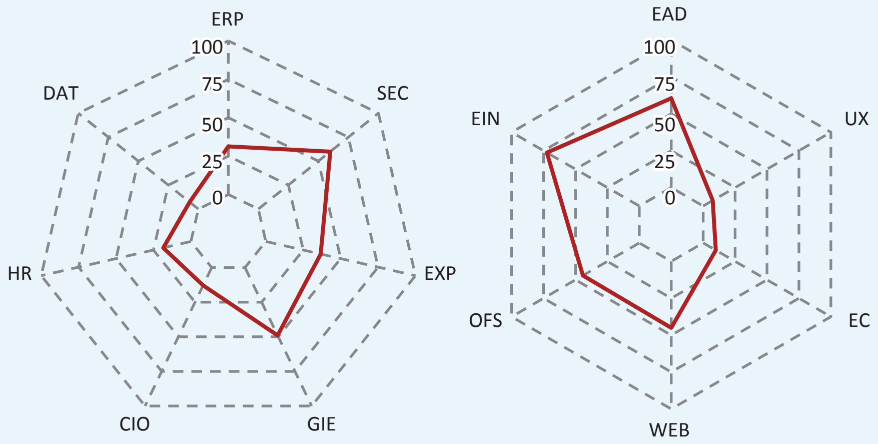


On the boxplot, the two edges of the rectangles mark the boundaries of the bottom and top quarters, and the middle line marks the median value. The dots represent outliers and are not included in the quartile bounds, they are outside the median by about ±3 standard deviations.[35] The median and average values of the headline indicator are very close: 34.2 vs. 35.5 points. Based on the boxplot of the headline indicator, half of the companies score between 30 and 50 points, within a relatively narrow range, with the top and bottom quartile of firms scoring roughly 25 to 25 points.

The integration of business functions and marketing/communication subindices had the lowest average scores of only around 20 points, i.e. only one fifth of the maximum score available on average (Chart 3). By contrast, the average score is close to 55 for e-administration and is 48 for IT infrastructure, but this is largely based on the use of mature technologies, unlike in the previous two areas. The high number of extreme values for marketing and communication at the top end of the distribution suggests that the question group results in a strong segmentation of the companies. The main reason for this is that e-commerce and online customer relationships are a top priority for trading and services companies, which is not the case in other sectors. The middle 50 per cent of firms in the integration of business functions and IT infrastructure subindices are scattered over the widest intervals, it is here that the rectangle is the widest. By contrast, in e-administration and marketing/communications it is approximately half as high, at roughly 15 points. The headline indicator is the weighted average of the above four subindices, with a relatively narrow range of the middle 50 per cent.

[35] Outliers are values that are more than one and a half interquartile ranges from the limits of the interquartile range.

Chart 4
Median scores for the subject areas[36]



The radar charts (Chart 4), which show the weighted average scores for the 13 subject areas, reveal even more significant differences between the subject area indices. The use of enterprise resource planning/administration systems and related modules (SCM, CRM) is still the most widespread, with about half of all cloud-based services being used on average in the 4 different groups (day-to-day business operations, hosting, storage, computing capacity). The penetration of the use of data analytics (big data) is very low. Median levels of basic digital coverage are relatively high, with a widespread use of mobile internet, and relatively high levels in computer coverage, the proportion of employees using the internet and the use of portable devices, as is the use of ICT security solutions (especially firewalls). A relative lag is evident in ICT expertise: few companies have a dedicated IT team/professional, IT training for employees is not widespread, and firms typically do not use external IT support.

Within the marketing/communication subject areas, web presence has the highest average score, followed by e-commerce and then UX maturity. The latter two subject areas are more sector-specific, which explains the low average score. By contrast, e-administration (i.e. e-government) is widespread and firms are likely to conduct their business online (whereas the use of non-banking digital financial solutions is less widespread), as is e-invoicing, which is naturally highly developed due to the requirement for mandatory data supply. The average score for online financial services is close to but remains below fifty.

The (Pearson) correlations between the four subindices are all significantly different from zero and relatively strong, i.e. the extent of ICT use related to business functions is correlated (see Chart F2 in the Appendix). The greatest degree of co-movement is between IT infrastructure and the integration of business functions. The higher the IT infrastructure score, the higher the integration of business functions score; the relationship is non-linear overall, with a higher correlation observed from a score of approx. 55. This suggests that a well-developed IT infrastructure is a prerequisite for supporting the integration of business processes effectively.

There is also a significant correlation between the IT infrastructure and e-administration domains (0.52). The ICT usage intensity associated with e-administration also requires an IT infrastructure of high standard, although to a lesser extent, given that accessing these services is easier and less complex than implementing administration systems/ERP, SCM, CRM or data analytics or complex cloud-based services.

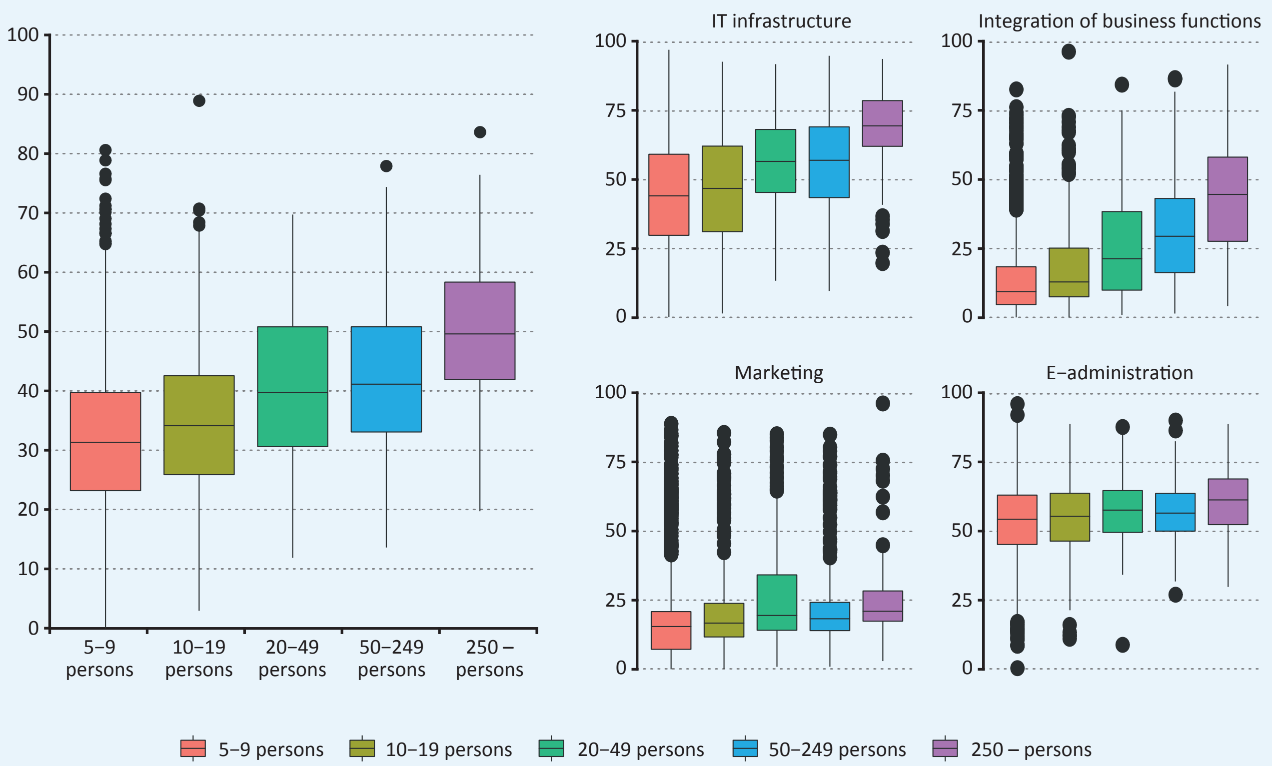
[36] The following abbreviations are used for the subject areas in the diagrams below: GIE – General ICT endowment; EAD – e-government, e-administration; OFS – Online financial services; WEB – Web presence; ERP – Enterprise resource planning; HR – Digital HR practices and systems; EIN – E-invoicing; DAT – Data analytics/big data; CLO – Cloud; EC – E-commerce; SEC – Use of ICT security software; EXP – Use of ICT expertise; UX – UX maturity.

When examining the correlations of the 13 subject area indices, it is obvious that correlation is highest between the subject areas of the individual subindices. The IT infrastructure and integration of business functions domains show the highest correlation coefficients, especially between corporate and ICT expertise, and the use of ICT security software (Chart F3 in the Appendix). ICT expertise also shows a strong co-movement with the subject areas UX maturity and web presence. The low penetration of e-commerce is shown by its weak correlation coefficient with the other subject areas. Similarly, the correlation with other variables is weaker than average for e-invoicing and e-administration. The highest correlation with the other variables is found for enterprise resource planning.

4.2 COMPANY SIZE AND EXPORT ACTIVITY

By company size, the digitisation headline indicator shows significant differences. Company size is approximated using the number of employees, applying five employee categories in the analysis and to visualise the results. The median digitisation headline indicator increases in value as the size of the company rises (Chart 5). The dispersion of the DIG index does not differ by category for firms in the middle 50 per cent, although the sample element number decreases substantially as size increases, i.e. there is no significant dispersion. There are larger jumps in the case of companies with 20–250 employees compared to microbusinesses, followed by firms with more than 250 employees, with a 6-to-7-point jump in the median firm size for the former and a nearly 10-point jump for the latter.

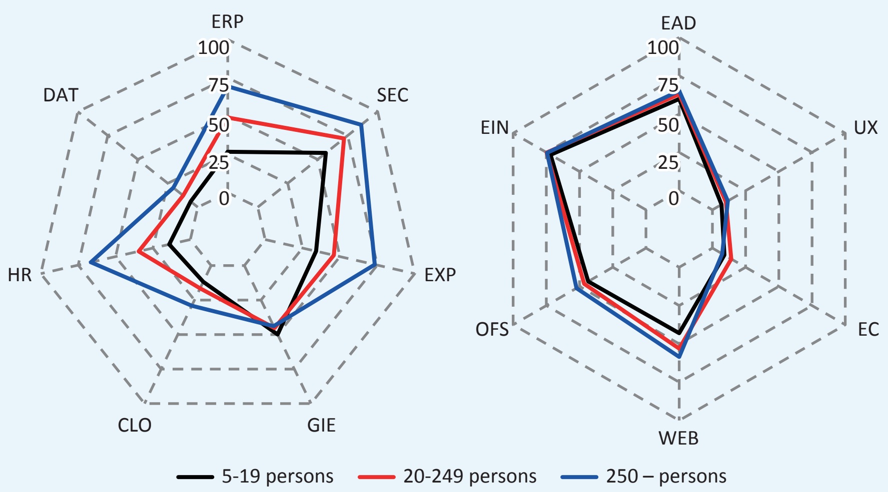
Chart 5
Digitisation headline index and subindices by company size



By subindex, the largest difference is seen at the integration of business functions (Chart 5), with the median score rising sharply with increasing company size, especially for the subject areas of ERP/administration systems and cloud usage. Large enterprises score higher than SMEs mainly in the areas of IT infrastructure and integration of business functions. With increasing size, the use of ERP systems becomes more common, as is the use of technologies that support it and decision-making (e.g. cloud use, data analytics); however, the maturity of IT infrastructure is not significantly higher than at smaller companies. The largest corporates (250 or more employees) stand out for their better IT infrastructure, mainly in terms of ICT expertise, but there is no difference in terms of basic digital coverage. Rather, the latter confirms that access may be even across the corporate sector and geographically, and that these technologies (mobile internet or portable devices, computerisation) are already in their mature phase.

Microbusinesses lag behind enterprises with 10 to 19 employees only in the integration of business functions area. Firms with less than 20 employees score markedly differently in the areas of integration of business functions and IT infrastructure, while there is no significant difference in e-administration and marketing/communications. Among these firms there is a relatively large number of outliers, companies with less than 20 employees that are significantly above the median in terms of integration. The small size differential in e-administration and marketing communications may be related to the cost-effectiveness of using these technologies, their mandatory use (e.g. certain online financial services, administration via the company e-portal "Cégkapu", e-invoicing) and the fact that web presence may be key to growth for smaller firms. On the flip side, the importance of customer relationships is not necessarily key to all larger firms, especially in the manufacturing industries and in construction.

Chart 6
Subject area indicators by company size



Based on the radar charts of the 13 subject areas, digital sophistication increases significantly with increasing company size in the subdomains of integration of business functions (VV, BB, DD, FF), in the use of ICT security software (IK) and in the use of ICT expertise (ICT) (see Chart 6). These areas are relatively closely interlinked, with the use of sophisticated ERP systems presupposing a high level of IT infrastructure, and the latter tending to result in the intensive use of data analytics or cloud-based services.

ICT use varies significantly not only by company size but also by other characteristics. We investigate the differences in the level of digitisation measured by the headline index and subindices by export activity[37] and headcount categories, using simple OLS regression. We run regressions in which dummy variables represent the ten company categories examined in terms of headcount and export activity. This method offers the possibility to show the average variation across the ten company categories by eliminating sectoral effects.

[37] We define exporters as companies that derive at least 1 per cent of their net sales revenue from exports. The results are similar if we define an exporter as a company with exports representing at least 10 per cent rather than 1 per cent of total sales revenue.

The first column of Table 6 shows the average headline index score for company size and export activity category. Within an industry, firms that also export tend to have a headline index score that is 4–5 points higher than non-exporting firms with a similar headcount category. In the subindex capturing IT infrastructure, up to headcounts of maximum 50 the average score of exporters differs by 10–15 points compared to firms of the same size. The same is true, but with a smaller average difference per pair, in the case of the subindex of areas responsible for integrating business functions. As the size increases, the average score for the headline index and the IT infrastructure and business function integration subindices rise steadily. There is an average difference of 15 points between firms with 5 to 9 employees and those with over 250 employees across the three indices for both exporting and non-exporting firms.[38] When only the export dummy variable is included in the regression alongside the sectoral dummy variables, there is a statistically significant difference between exporters and non-exporters (see Table F14 in the Appendix).[39]

Table 6
Differences in digitisation indices by company size and exports

	Dependent variables				
	Headline index	**IT infrastructure subindex**	**Integration of business functions subindex**	**E-administration subindex**	**Marketing/ communications subindex**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Exporter, 5–9 staff	32,730***	45,657***	14,281***	58,815***	13,335**
	(4,130)	(6,142)	(5,186)	(4,528)	(5,926)
Exporter, 10–19 staff	41,006***	54,991***	24,532***	63,932***	21,133***
	(4,322)	(6,428)	(5,427)	(4,739)	(6,203)
Exporter, 20–49 staff	44,012***	54,685***	30,924***	62,571***	28,411***
	(4,069)	(6,052)	(5,110)	(4,462)	(5,840)
Exporter, 50–249 staff	40,990***	52,383***	31,343***	61,749***	18,402***
	(3,911)	(5,816)	(4,911)	(4,288)	(5,612)
Exporter, 249+ staff	47,429***	61,370***	42,381***	64,358***	20,206***
	(4,412)	(6,561)	(5,540)	(4,837)	(6,331)
Not exp., 5–9 staff	27,922***	35,106***	9,851**	56,667***	12,337**
	(3,675)	(5,466)	(4,615)	(4,029)	(5,274)
Not exp., 10–19 staff	30,369***	37,709***	13,391***	58,478***	13,949***
	(3,685)	(5,480)	(4,627)	(4,040)	(5,288)
Not exp., 20–49 staff	35,867***	45,856***	19,183***	60,517***	19,320***
	(3,754)	(5,583)	(4,714)	(4,116)	(5,387)
Not exp., 50–249 staff	36,896***	46,336***	24,412***	60,534***	17,119***
	(3,835)	(5,704)	(4,816)	(4,205)	(5,504)
Not exp., 249+ staff	42,659***	56,095***	35,403***	60,917***	17,305**
	(5,094)	(7,576)	(6,397)	(5,585)	(7,310)
N	2,500	2,500	2,500	2,500	2,500
R^2	0.902	0.884	0.642	0.948	0.629
Adjusted R^2	0.901	0.882	0.638	0.947	0.625
F stat (df = 28; 2472)	815,090***	671,219***	158,035***	1,606,995***	149,820***

*Separate regressions per column, estimated without constant and with sectoral dummy variables. *p<0.1; **p<0.05; ***p<0.01.*

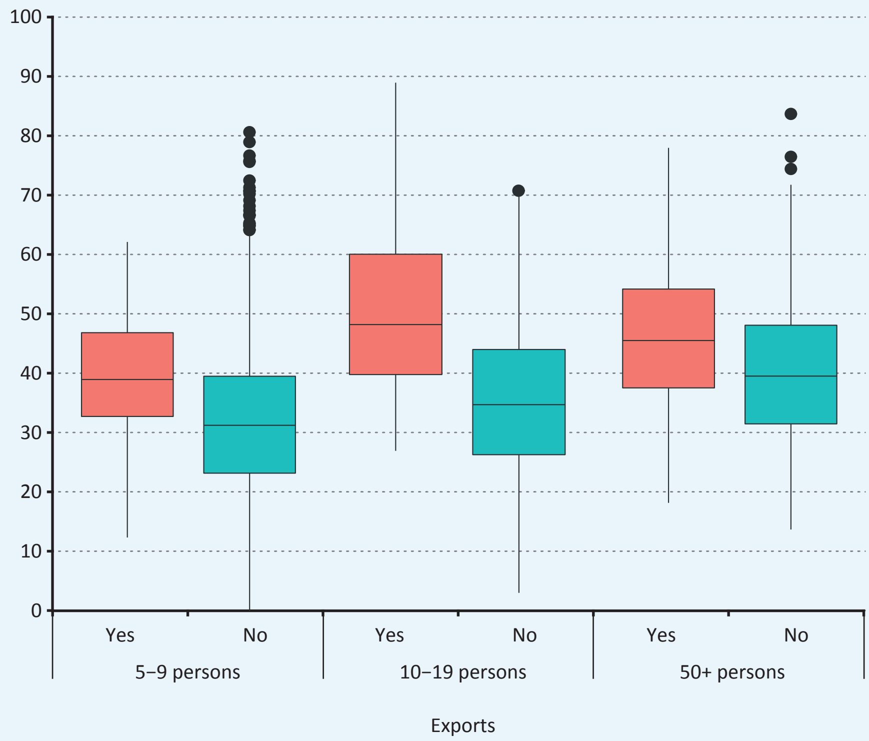
[38] We chose the 1 per cent export threshold because just over 8 per cent of the weighted sample is engaged in export activity. There is no significant change in the coefficients of the regression if a company is considered an exporter at an export ratio of above 10%, but the standard errors increase in the categories as the number of items is small.

[39] The dummy coefficient has a 95% confidence interval of 48.3 ± 2.9 points for exporters and 34.5 ± 2.3 points for non-exporters.

Neither the e-administration nor the marketing/communication subindices show a significant difference in the export dimension. The average score increases slightly with increasing company size and there is also a difference in the export dimension. The exception to the former is the marketing/communications sector, where typically an above-average number of firms with 20 to 49 employees have a marketing specialisation.

Among enterprises with 5 to 9 employees, 10 to 49 employees or 51 or more employees, the median value of exporters is significantly higher than that of non-exporters (Chart 7). The upper bound of the interquartile range for non-exporters is just below the median for exporters in all three categories. For smaller firms, the difference is more significant in terms of exports.

Chart 7
Headline index value by export activity

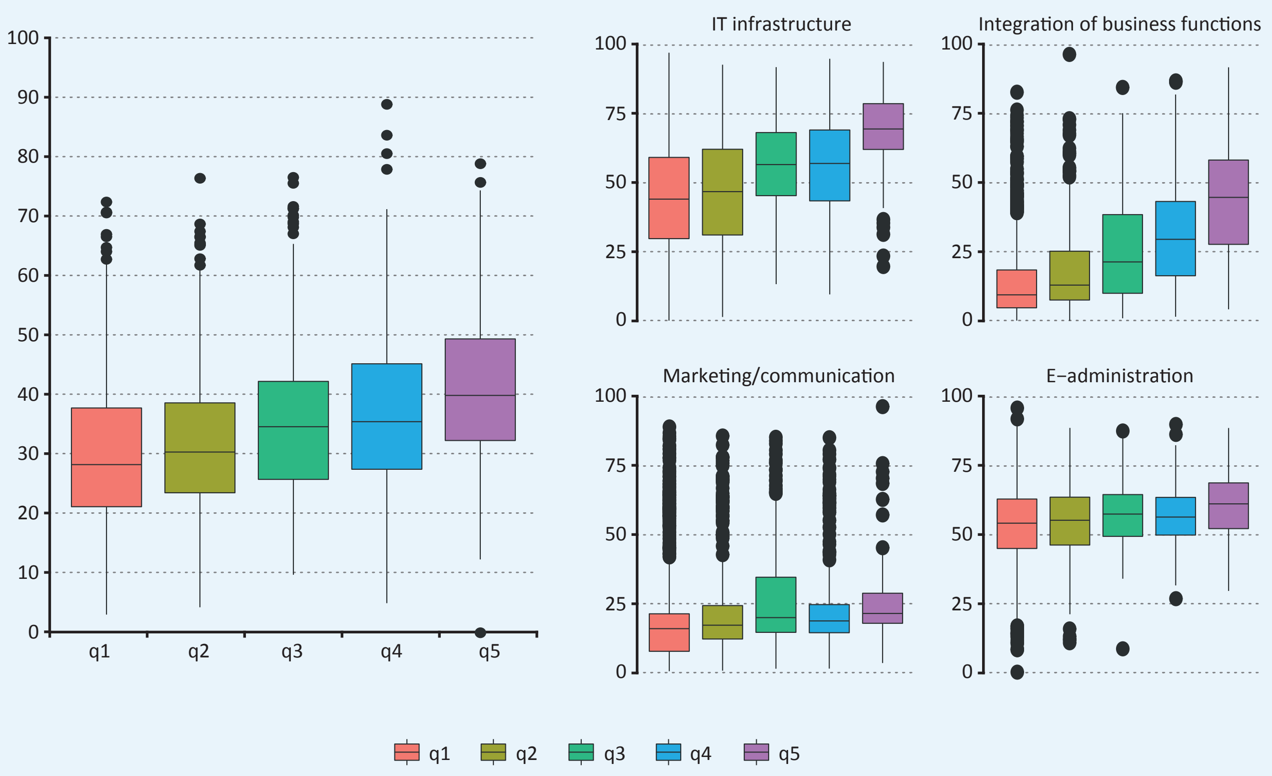


4.3 LABOUR PRODUCTIVITY

We define company productivity as labour productivity and approximate it with the variable of value added divided by the number of employees.[40] Where the variable of value added per headcount is available, the relationship between the productivity and ICT use intensity of the companies is not nearly as strong as the relationship with exports or company size.[41]

Classifying companies into quintiles according to their labour productivity, the median values of the index capturing ICT usage intensity rise gradually as productivity increases (Chart 8). Jumps may be observed between the second and third quintiles and the fourth and fifth quintiles. There is a large overlap between quintiles in the middle 50 per cent of firms. Of the four subindices, only the areas of IT infrastructure and integration of business functions differ in terms of business productivity. Going further into details, these differences can be attributed to general ICT endowment, the use of data analytics (big data) and ICT expertise, and to enterprise resource planning systems.

Chart 8
Headline index and subindices by productivity quintile



By examining the relationship between productivity and ICT use in regression analysis, we find that, after controlling for location, sector and firm size, the digitisation index is 4 points higher for the third, 5 points higher for the fourth and 7 points higher for the fifth quintile compared to the lowest productivity quintile (Table 7), while firm size remains significant.

[40] The nominal amount of value added is obtained using the following rows in the 2019 income statements: net sales revenue plus variation in stocks of finished goods and work in progress less material expenditures. Where the year 2019 turns out to be an outlier based on the 2017–2018 income statements, we take an average of the available years by bringing the 2017 and 2018 values to the 2019 level using the relevant sectoral deflators. Around 40 firms have a negative value added (in the earlier available years too); we replace these with the sectoral bottom decile median value to avoid losing observations.

[41] For 35 companies there were no annual reports available.

Table 7
Differences in digitisation indices by company size and exports

	Dependent variable	
	DIG headline index	log(productivity)
	(1)	(2)
log(DIG headline index)		0.215***
		(0.037)
does not export	-4.805***	-0.465***
	(0.980)	(0.072)
employee_group: 5–9 persons	-2.130***	-0.072*
	(0.554)	(0.041)
employee_group: 20–49 persons	4.753***	0.172***
	(0.835)	(0.062)
employee_group: 50–249 persons	5.200***	-0.005
	(1.077)	(0.080)
employee_group: 250 or above	11.295***	-0.203
	(2.143)	(0.159)
established: 2008 to 2014	0.040	0.008
	(0.549)	(0.041)
established: after 2014	0.450	-0.020
	(0.930)	(0.069)
productivity quintile2	1.090	
	(0.732)	
productivity quintile3	3.959***	
	(0.746)	
productivity quintile4	4.714***	
	(0.762)	
productivity quintile5	6.794***	
	(0.785)	
Number of observations	2,459	2,459
R2	0.257	0.182
Adjusted R2	0.246	0.171
Residual standard error	11.448 (df = 2422)	0.849 (df = 2425)
F stat	23.297*** (df = 36; 2422)	16.341*** (df = 33; 2425)

**p<0.1; **p<0.05; ***p<0.01. Standard errors in brackets. Estimated with constants, sectoral and regional dummy variables. Reference categories: lowest productivity quintile, companies established before 2008, 10–19 employees, exporters.*

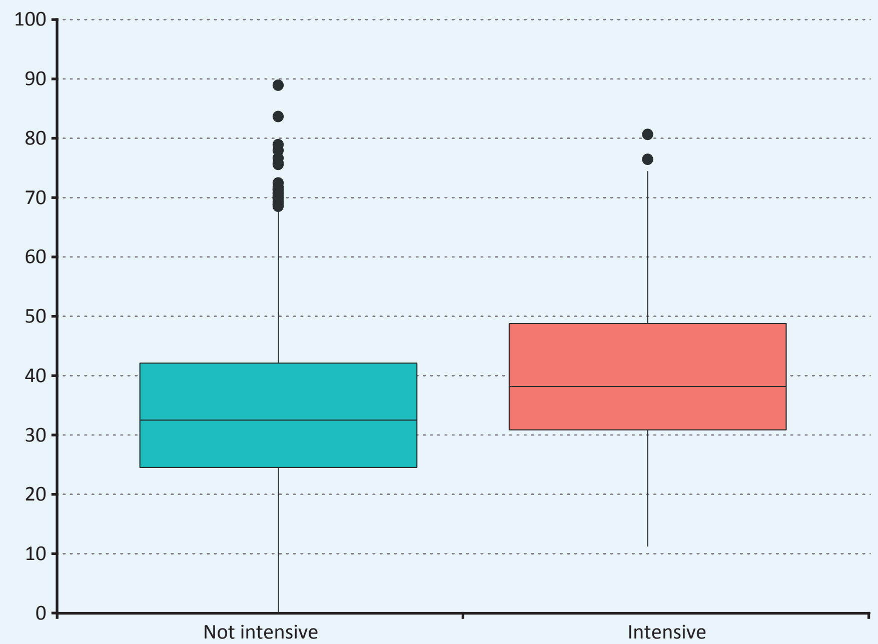
Using regression to examine the relationship between labour productivity and ICT use, with the logarithm of labour productivity as the dependent variable and the logarithm of the headline index of digitisation as the independent variable, a 10% higher value of the headline index is associated with approximately 2.2% higher productivity.[42] In this specification company size no longer explains variation in productivity to the same extent as when we regressed productivity quintiles on the headline index. There is a strong correlation between productivity and the headline index of digitisation, but we do not investigate the causal relationship as this has not been an objective in this study.

[42] The value of the coefficient may vary marginally depending on whether the outliers (top and bottom one per cent) are removed from the estimation or not.

4.4 SECTOR, GEOGRAPHICAL LOCATION, AGE OF COMPANY

At a detailed sectoral level, there is a large difference in ICT usage intensity between sectors (see Chart F4 in the Appendix). The median scores for the sectors range from 25 points to as high as 50 points. The information communication sector leads the way, followed by financial intermediation. These are followed by sectors that typically employ workers with above-average educational qualifications (professional and scientific activities, civil service). Trade and manufacturing, two sectors with a large number of companies, are slightly above the mid-range, typically with a large standard deviation. The least digitised sectors are agriculture, accommodation services, mining, transport, storage and construction.

Chart 9
Digitisation headline index by knowledge-intensive classification



In the category pair of knowledge-intensive versus non-knowledge-intensive industries generated on the basis of sectoral classification[43], the latter have a median digitisation index score that is five times higher; knowledge-intensive firms account for 17 per cent of the sample (Chart 9). This appears to suggest that there is a gap between the qualification levels of the workforce as determined by the sector and the level of digitisation.

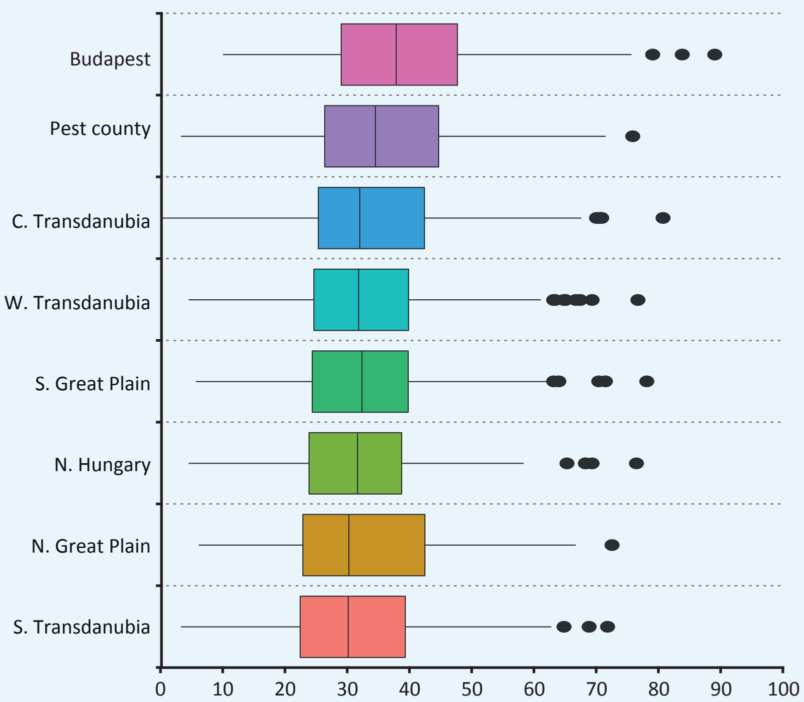
In terms of the geographical location of the head office, median values are close to each other, and the interquartile range is also uniform across the eight regions (Chart 10). Access inequalities due to geographical location[44] are not detectable. Typically, less developed regions have lower median scores, but the difference is not significant. When compared to the other regions, the median score of Budapest is clearly higher, by 6–7 points; the difference is even smaller compared to the median score of firms in Pest county (constituting part of Budapest metropolitan area). The difference becomes statistically significant when considering the locations of regional company headquarters. On average, companies in Budapest have an index score 3 points higher than firms in underdeveloped regions and 1.5 points higher than firms in Pest county, although the latter difference is not significant.

[43] An industry can be considered knowledge-intensive if tertiary graduates account for at least one third of all employees on average in the EU-27. For more details on the classification methodology, see: https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an8.pdf.

[44] Based on the postcode of the corporate head offices, we categorised our sample of companies into eight geographical regions.

The uptake of certain key ICTs (measured as a percentage of firms that have adopted them) shows significant differences by sector, as well as by the company size subject area (Chart F5 in the Appendix). The penetration of technologies varies widely across sectors and follows sector-specific patterns. In accommodation services and catering, social network usage is saliently high, and e-invoicing is also widespread. The information and communication sector is at the forefront of almost all the technologies examined here, but particularly in the use of cloud-based services and employing ICT experts. Social media use and webshop penetration are above the average in the retail sector, as is ERP or CRM/SCM/ERM use in manufacturing, and cloud and data analytics in the scientific professions. The relative lag of the construction sector is clearly shown in the radar charts.

Chart 10
Digitisation headline index by region



There is no significant difference in the level digitalisation by firm age (three categories: firms founded after 2014, between 2008–2014 and before 2008) when adding it to regression along with export and company size (see Table F15 in the Appendix). Incidentally, the level of digitisation is slightly higher at the youngest companies than in the other two categories.

5 Conclusion

The digitisation of business has advanced rapidly over the last few years, regardless of company size and business activity. The digitisation of business functions and the introduction of new technologies is not happening at the same speed; smaller businesses tend to react more slowly and some sectors more quickly than others. The use of ICT in the corporate sector is diversified, and the mix of technologies adopted varies from company to company, depending on organisational capacity and digitisation strategy. The aim of this study has been to construct a digitisation index and subindices of the ICT usage patterns in the Hungarian corporate sector in 2020 (pre-COVID-19), and to map the patterns of ICT usage by company size, sector, productivity and export activity, providing stylised facts based on a snapshot. We are not aware of any representative published surveys result of ICT use in the corporate sector conducted on a larger sample of Hungarian data. Compared to most international surveys of similar data, we also have data on a subset of micro-enterprises, those with 5–9 employees.

In terms of content, the questionnaire survey closely follows other, typically international data collections measuring various aspects of ICT use in enterprises and shows a significant degree of similarity with the content of the OECD's and the European Union's DESI data collection. Other than the questionnaire survey, no additional information on ICT use was incorporated due to an absence of such data. When processing qualitative questionnaire data, the information was first converted into numerical data by subject area using a scoring procedure. The composite index and the subindices were aggregated and weighted from the data aggregated at lower levels using principal component analysis, a technique commonly used in the literature.

Ex ante, four subindices were expected; this was confirmed by principal component analysis. The four areas can be distinguished according to business functions and are in line with the literature (OECD 2020). The areas responsible for the integration of business functions are the enterprise management systems, data analytics and cloud services supporting (strategic) decision-making, and the use of digital HR solutions. E-administration as a business function includes communication with government agencies, electronic invoicing and the use of online financial services. The IT infrastructure supports the functioning of the other areas, such as ICT expertise, use of ICT security software, and general ICT endowment (computerisation, personal digital devices, or mobile internet use). The fourth area identified is marketing and communications, which covers the fields of web presence, e-commerce and UX maturity. For each of these three subject areas, the quality of customer relationships is of paramount importance.

The intensity of ICT use captured by the digitisation index as well as its subindices is closely related to the size of the company. The larger a company is (the more employees it has), the more sophisticated ICT solutions it uses and the better its IT infrastructure is; this, in turn, supports the efficient operation of such solutions. Medium-size and large enterprises tend to employ complex ERP systems and the most highly digitised ICT areas responsible for integrating business functions, as they tend to use cloud-based services or data analytics/big data. IT infrastructure does not differ as much by company size as it does in the above domain. In the marketing/communication dimension of digitisation, company size is not a cardinal factor.

We also found no significant differences between regions in terms of the geographical location of firms, although the average digitisation level of companies in Budapest is slightly higher than in other regions. This shows that accessibility (e.g. broadband affordability, availability, access to ICT skills) may play little or no role in explaining differences in ICT use. Companies that also export tend to be more digitally active than those that do not trade abroad. This characteristic of the companies is closely related to company size; however, even among firms of the same size within a sector, exporters have significantly higher intensity of ICT usage than non-exporters.

Other company characteristics, such as sectoral activity, show very marked differences in the extent of ICT use. The information and communication sector, banking and insurance, as well as jobs requiring higher qualifications (professional, scientific) are at the forefront. There can be large differences in digitisation levels within sectors (manufacturing, trade).

The age of a company does not really matter: on average, newly established firms have a higher digitisation index score than those founded earlier, but the difference is negligible and insignificant.

A relatively strong relationship can be identified between productivity and the intensity of ICT use. Companies with higher productivity tend to have more extensive ICT usage patterns, considering their export orientation, the geographical location of their headquarters, their age or the size of the company. While determining the direction of causality has not been a subject for this study, we note that causality is unclear, as it may not be through ICT use that companies gain a competitive advantage and achieve higher productivity, but the other way around: companies that are more competitive and/or productive to start with can afford to adopt the relatively costly new technologies.

The impact of the COVID-19 pandemic on the digitisation level of the SME sector is unlikely to be felt yet, as the survey was conducted during the first lockdown and in October 2020, at which point there had not been sufficient time for any meaningful adjustments. In any case, the pandemic has exerted considerable pressure on small and medium-size enterprises to bring their ICT use up to speed and adopt new technologies. It will become possible to judge their convergence only when a new survey has been conducted to identify the extent of changes since 2020. Generating the ICT usage intensity index presented in this study using more recent data from after 2021 would be a suitable means for this task.

References

Andrews, D., Nicoletti G., Timiliotis, N. (2018): Digital technology diffusion: A matter of capabilities, incentives or both? OECD Economics Department Working Papers 1476, OECD Publishing.

Bertschek, I., Fryges, H., Kaiser, U. (2006): B2B or Not to Be: Does B2B E-Commerce Increase Labour Productivity? International Journal of the Economics of Business 13(3), pp. 387–405.

Bertschek, I., Cerquera, D., Klein, G.J., (2013): More Bits – More Bucks? Measuring the Impact of Broadband Internet on Firm Performance, Information Economics and Policy 25(3), pp. 190–203.

Bertschek, I., Niebel, T. (2015): Mobile and More Productive? Firm-Level Evidence on the Productivity Effects of Mobile Internet Use. ZEW – Centre for European Economic Research Discussion Paper No. 15-090, https://ssrn.com/abstract=2766252.

Bertschek, I., Ohnemus, J. and Viete, S. (2018): The ZEW ICT Survey 2002 to 2015: Measuring the Digital Transformation in German Firms. Jahrbücher für Nationalökonomie und Statistik 238(1), pp. 87–99. https://doi.org/10.1515/jbnst-2016-1005.

van Buuren, S., Groothuis-Oudshoorn, K. (2011): mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), pp. 1–67.

Camara, N., Tuesta, D. (2017): DiGiX: The Digitization Index. BBVA Working Paper No. 17/03.

Cette, G., Nevoux, S., Py, L. (2020): The impact of ICTs and digitalization on productivity and labor share: Evidence from French firms. Banque de France Working Paper No. 785, November 2020.

Clayton, T., Criscuolo, C. (2002): Electronic Commerce and Business Change. Economic Trends No. 583, June 2002.

Clayton, T., Goodrigde, P. (2004): E-business and labour productivity in manufacturing and services. Economic Trends No. 609, August 2004.

European Commission (2020): Digital Economy and Society Index (DESI) 2020. Methodological note. http://www.juntadeandalucia.es/institutodeestadisticaycartografia/digitalizacion-sociedad-digital/metodologia/Notas-Metodologicas-DESI.pdf

European Commission (2021): Quantitative and Qualitative Data of the Inventory on Information and Communication Technologies. Hungary. https://circabc.europa.eu/sd/a/baf6e9cb-c504-4095-92b4-425e0b803037/isoc_sdds_hu.htm.

Falk, M., Hagsten, E. (2015): E-commerce trends and impacts across Europe, No 220, UNCTAD Discussion Papers, United Nations Conference on Trade and Development.

Falusi, Á., Kis, E. E., Kovaliczky, T., Mándics, T., Petrezselyem, P. (2022): Digiméter 2021. A digitális versenyképességi mutató. [Digimeter 2021. Digital Competitiveness Indicator.] https://digimeter.hu/wp-content/uploads/2022/01/Digimeter-2021-jelentes.pdf. Accessed: 01.12.2022.

Farooqui, S. (2005): Information and Communication Technology use and productivity. Economic Trends, No.625, December 2005.

Gal, P., et al. (2019): Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries. OECD Economics Department Working Papers, No. 1533, OECD Publishing, Paris.

Greco, S., Ishizaka, A., Tasiou, M. et al. (2019): On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. Soc Indic Res 141, pp. 61–94.

Mckinsey Global Institute (2015): Digital America: A tale of the haves and have-mores. McKinsey&Company, December 2015. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/digital-america-a-tale-of-the-haves-and-have-mores.

Morandini, M., Thum-Thysen, A., Vandeplas, A. (2020): Facing the digital transformation: are digital skills enough? No 054, European Economy – Economic Briefs. Directorate General. Economic and Financial Affairs, European Commission.

Nagar, A. L., Basu, S. R. (2002): Weighting socioeconomic indicators of human development: A latent variable approach. In Ullah A et al. (eds) Handbook of Applied Econometrics and Statistical Inference. Marcel Dekker, New York. pp. 631–664.

Nicoletti, G., Scarpetta, S., Boylaud, O. (2000): Summary Indicators of Product Market Regulation with an Extension to Employment Protection Legislation. OECD, ECO Working Paper No. 226.

OECD, Joint Research Centre – European Commission (2008): Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing, Paris.

OECD (2015): The OECD Model Survey on ICT Usage by Businesses. 2nd Revision, Working Party on Measurement and Analysis of the Digital Economy. https://www.oecd.org/sti/ieconomy/ICT-Model-Survey-Usage-Businesses.pdf.

OECD (2021): The Digital Transformation of SMEs, OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, https://doi.org/10.1787/bdb9256a-en.

Portulans Institute (2020): The Network Readiness Index 2020. Accelerating Digital Transformation in a post-COVID Global Economy. https://networkreadinessindex.org/2020/wp-content/uploads/2020/11/NRI-2020-V8_28-11-2020.pdf.

Szerb, L., Somogyine Komlosi, E., Acs, J. Z., Lafuente, E., Song, K. A. (2022): The Digital Platform Economy Index 2020. Springer Cham. https://doi.org/10.1007/978-3-030-89651-5.

Appendix

Table F1
Descriptive statistics for the variables used

	average	median	deviation	minimum	maximum	n
corporate governance	33.2	33.3	35.0	0	100	2,500
data analytics	7.7	0.0	21.1	0	100	2,500
digital HR	21.2	0.0	35.4	0	100	2,500
cloud	13.1	0.0	24.9	0	100	2,500
ICT security	60.7	60.0	34.4	0	100	2,500
ICT expertise	38.2	30.0	22.6	0	100	2,500
basic ICT endowment	48.8	48.3	29.9	0	100	2,500
e-administration	61.0	58.3	25.9	0	100	2,500
e-invoicing	72.3	66.7	20.3	0	100	2,500
online financial services	44.4	42.5	14.7	0	100	2,500
web presence	44.7	50.0	26.5	0	100	2,500
e-commerce	9.3	0.0	25.9	0	100	2,500
UX maturity	7.2	0.0	22.3	0	100	2,500
IT infrastructure	47.9	48.4	19.6	0	97.0	2,500
integration of business functions	19.4	13.0	17.4	0	96.4	2,500
e-administration	54.8	55.4	13.2	0	96.2	2,500
marketing/communications	20.6	16.7	17.5	0	96.5	2,500
DIG headline index	**35.5**	**34.2**	**13.3**	**0**	**88.9**	**2,500**
weight variable	1.0	0.9	0.4	0.2	2.2	2,500
log(labour productivity)	8.6	8.6	0.9	3.6	14.8	2,459
export ratio (%)	5.3	0.0	18.9	0.0	100.0	2,465

Note: unweighted values

	distribution (%)		distribution (%)
exports (at least 1% of turnover)		**region** (head office)	
yes	12.2	Budapest	28.6
no	87.8	Pest county	14.2
founded		Southern Great Plain	11.8
before 2008	68.7	Southern Transdanubia	7.7
2008–2014	23.9	Northern Great Plain	10.7
after 2014	6.6	Northern Hungary	7.4
sector		Central Transdanubia	10.1
agriculture	4.4	Western Transdanubia	9.6
manufacturing	19.4		
construction	10.3		
other industry and mining	1.4	**number of employees**	
trade, transport, storage	32.4	5–9 persons	58.5
professional, scientific, technical activities	10.9	10–19 persons	16.6
real estate, accommodation, food service	9.9	20–49 persons	6.8
administrative and support service activities	3.7	50–249 persons	14.2
other services	7.6	250 or above	3.9

Note: unweighted values

Chart F1
Headline digitisation index decile position of firms by two types of PCA procedures

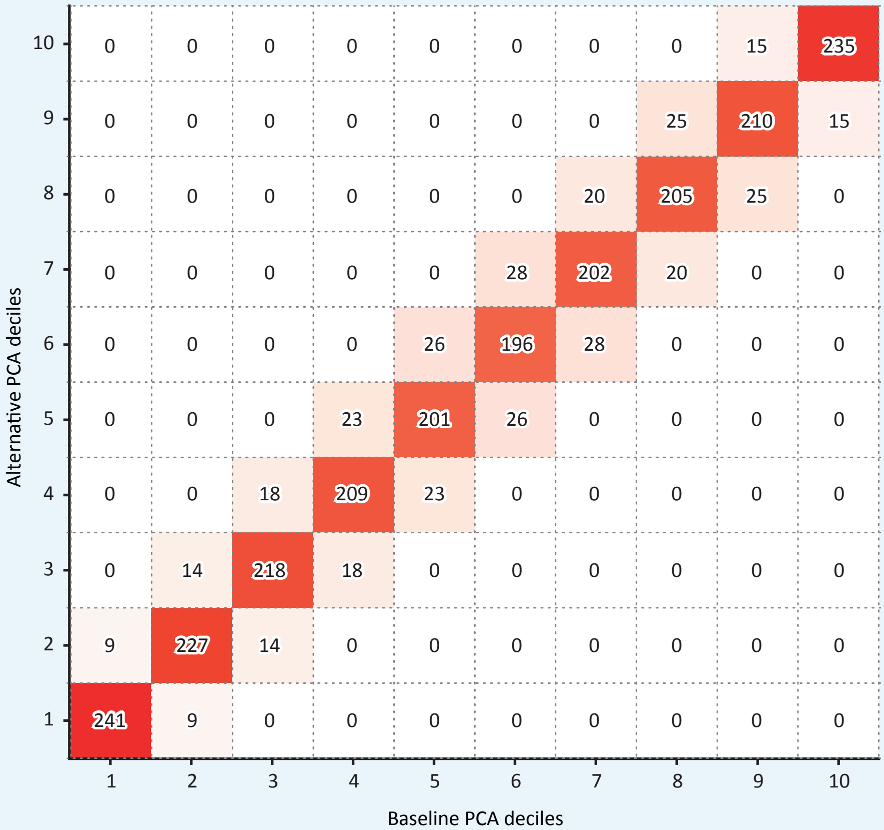


Chart F2
Correlations between subindices

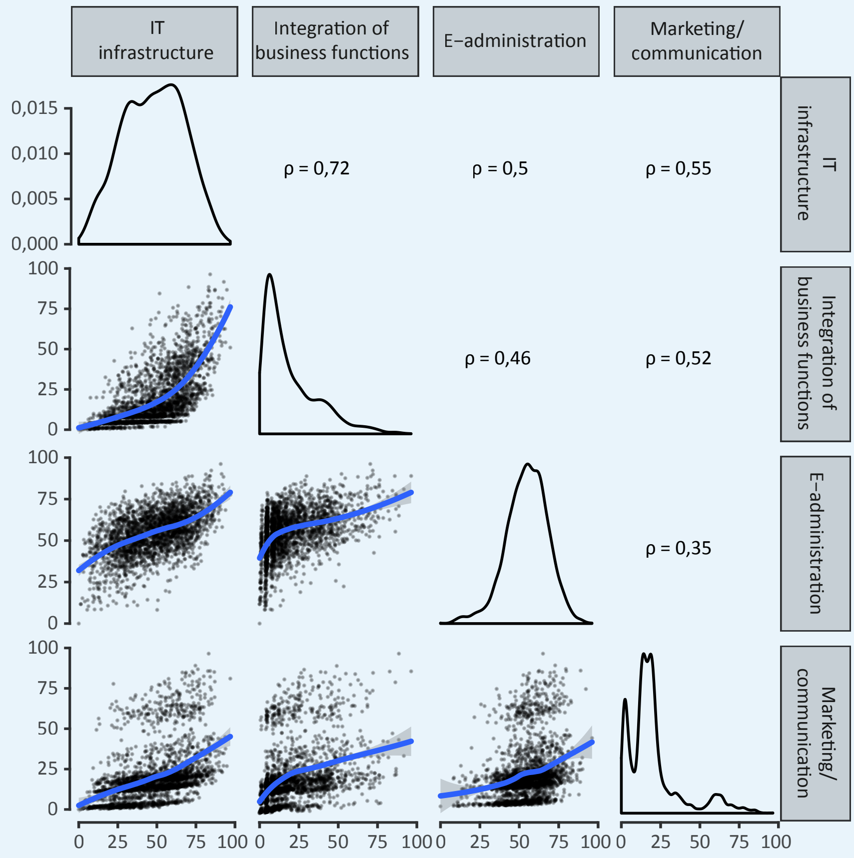


Chart F3
Correlations between subject area indicators

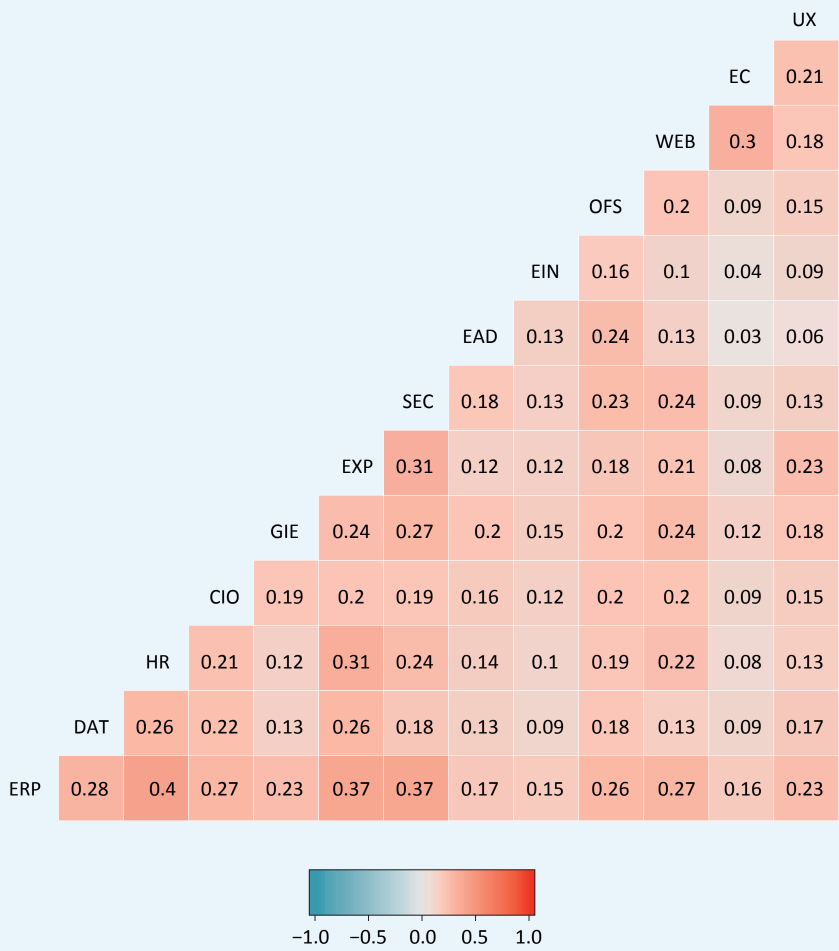
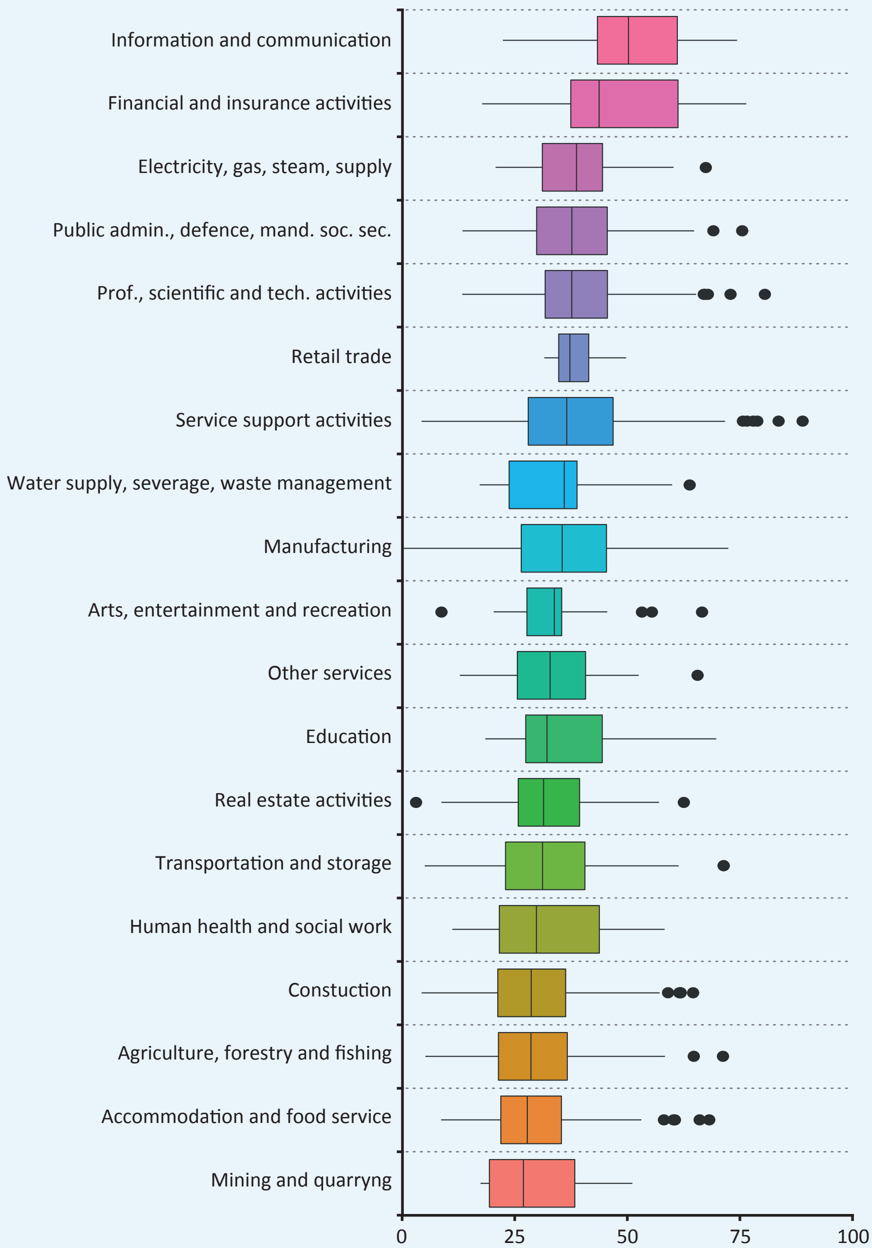


Chart F4
Digitisation headline index by industries



TABLES F2–F14: SELECTED QUESTIONS FROM THE QUESTIONNAIRE SURVEY AND THEIR SCORING IN THE THIRTEEN SUBJECT AREAS

Table F2
Data analysis/use of big data

Uses data processing technology	Type of technology used (data mining; machine learning; big data; speech recognition; text processing)	Type of data analysed (data obtained from smart devices and sensors; geolocation data; social media data)	Data purchased from other entities?	
yes/no	at least one of 5 yes/no	at least one of 3 yes/no	yes/no	Score
Yes	Yes	Yes	Yes	6
Yes	Yes	Yes	No	5
Yes	Yes	No	Yes	5
Yes	Yes	No	No	4
Yes	No	Yes	Yes	4
Yes	No	Yes	No	5
Yes	No	No	Yes	4
Yes	No	No	No	3
No	No	No	Yes	2
No	No	No	No	0

Table F3
Use of ICT security software

	weight (w_i)	yes	no
Uses			
firewall	40%	0	6
secure remote access system (VPN)	20%	0	6
encryption (computers, internal communications)	20%	0	6
data leakage protection (DLP) system	20%	0	6

Total score: $\Sigma w_i w_j score_{ij}$

Table F4
ICT expertise

	weight (w_i)	yes	no
Does the company have full-time ICT staff?	40%	6	0
further training in IT for ICT workers	10%	6	0
IT further training for staff	20%	6	0
How is IT infrastructure maintenance and office software support provided at the company?			
in-house and/or external staff	30%	6	0

Total score: $\Sigma w_i w_j score_{ij}$

Table F5
Digital HR practices and systems

	weight (w_i)	yes	no
Does the company use digital tools to perform HR tasks (either standalone software or a standalone HR system)?	85%	6	0
Does the company advertise directly on jobs portals?	15%	6	0
Total score: $\Sigma w_i w_j score_{ij}$			

Table F6
E-government, e-administration

	weight (w_i)	weight (w_j)	score	yes	no
Have there been examples of conducting online the administration of certain forms of government matters for the company?	60%				
starting business operations		25%		6	0
requesting official certificates		25%		6	0
retrieving data from registers		25%		6	0
issuing and storing accounting documents in electronic form		25%		6	0
How often does the company use the company portal Cégkapu for communication and administration purposes?	40%				
we always do it this way			6		
we often do it this way			4		
occasionally we do it this way			2		
we never do this/we do not use the company portal			0		
Total score: $\Sigma w_i w_j score_{ij}$					

Table F7
E-commerce

Does the business currently operate an online shop?	In 2019, did your company have any online, internet/webshop sales?		Score
Yes/No	in own webshop	yes, on an online marketplace	
Yes	Yes	Yes	6
Yes	Yes	No	5
Yes	No	Yes	3
Yes	No	No	2
No	No	Yes	1
No	No	No	0

Table F8
Use of Enterprise Resource Planning (ERP) systems, CRM, SCM, administration systems

Do you use an ERP system in your company's operations? Yes/No	A smaller, simpler administration system? Yes/No	Does the company use software solutions? How many out of 4?*	score
Yes	Yes	at least three	6
Yes	Yes	one or two	5
Yes	Yes	none	4
Yes	No	at least three	6
Yes	No	one or two	5
Yes	No	none	4
No	Yes	at least three	5
No	Yes	one or two	4
No	Yes	none	3
No	No	at least three	3
No	No	one or two	2
No	No	none	0

** (1) project management (e.g. Microsoft project); (2) supply chain management (SCM) / customer relationship management (CRM) / supplier relationship management (SRM); (3) manufacturing execution system (MES) / logistics / inventory / asset management; (4) finance / accounting / human resources management.*

Table F9
Cloud-based systems

	weight (w_j)	yes	no
Does the company buy cloud services on the internet for the following purposes?			
daily operational management (e.g. correspondence, finance, customer relations)	25%	6	0
corporate database hosting	25%	6	0
document storage	25%	6	0
computing capacity	25%	6	0

Total score: $\Sigma w_i w_j score_{ij}$

Table F10
Web presence

	weight (w_i)	weight (w_j)	yes	no
What online platform(s) is the business present on?	60%			
Standalone website		2/3	0	6
Facebook or other social networking site		1/3	0	6
Are the following available on the company's online platform?	40%			
live chat		25%	0	6
robot chat		25%	0	6
online ordering, reservations, purchasing		25%	0	6
automated reporting, complaint handling		25%	0	6

Total score: $\Sigma w_i w_j score_{ij}$

Table F11
E-invoicing

Which invoicing method does the company use?		Are you using electronic invoices?	What type of invoicing software do you use?		
manual invoicing	invoicing software		installed invoicing software	online, no-installation invoicing software (cloud-based)	
Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Score
Yes	Yes	Yes	Yes	Yes	6
Yes	Yes	Yes	No	Yes	6
Yes	Yes	Yes	Yes	No	5
No	Yes	Yes	Yes	Yes	6
No	Yes	Yes	No	Yes	6
No	Yes	Yes	Yes	No	5
Yes	Yes	No	Yes	Yes	5
Yes	Yes	No	No	Yes	5
Yes	Yes	No	Yes	No	4
No	Yes	No	Yes	Yes	4
No	Yes	No	No	Yes	4
No	Yes	No	Yes	No	3
No	No	No	No	No	0

Table F12
General ICT endowment

	weight (w_i)	weight (w_j)	Score	No	Yes
Proportion of employees using the internet	20%				
low			0		
average			3		
high			6		
Computer (desktop PC and laptop) coverage	40%				
low			0		
average			3		
high			6		
Mobile internet use	20%			0	6
Proportion of mobile device use (phone, tablet, laptop)	10%				
low			0		
average			3		
high			6		
Does the company provide employees with portable devices with mobile internet?	10%				
laptop		1/3		0	6
smartphone		1/3		0	6
tablet		1/3		0	6

Total score: $\Sigma w_i w_j score_{ij}$

Table F13
Use of online financial services

	weight (w_i)	weight (w_j)	Score	No	Yes
How does your business most often carry out banking transactions?	30%				
Online			6		
By phone			3		
In person			0		
Does your business use the following electronic banking services?	25%				
internet banking		1/6		0	6
banking system for corporate cash flow management		1/6		0	6
authenticated electronic statements of financial transactions		1/6		0	6
mobile application		1/6		0	6
online customs payment		1/6		0	6
POS terminal		1/6		0	6
Does your business use electronic signatures?	20%			0	6
Which non-banking digital financial solutions has your company used?	25%				
digital payment and money transfer solutions		20%		0	6
digital insurance solutions		20%		0	6
digital crowdfunding solutions		20%		0	6
digital solutions for cryptocurrency management		20%		0	6
fintech solutions for businesses		20%		0	6

Total score: $\Sigma w_i w_j score_{ij}$

Table F14
UX maturity

Is there web or software product development for sales purposes?	Are the users involved in the development process in any way?	Is there an employee or team specifically dedicated to user experience?	
Yes/No	**Yes/No**	**Yes/No**	**Score**
Yes	Yes	Yes	6
Yes	No	Yes	5
Yes	Yes	No	4
Yes	No	No	3
No	No	No	0

Chart F5
ICT usage (penetration) by sector

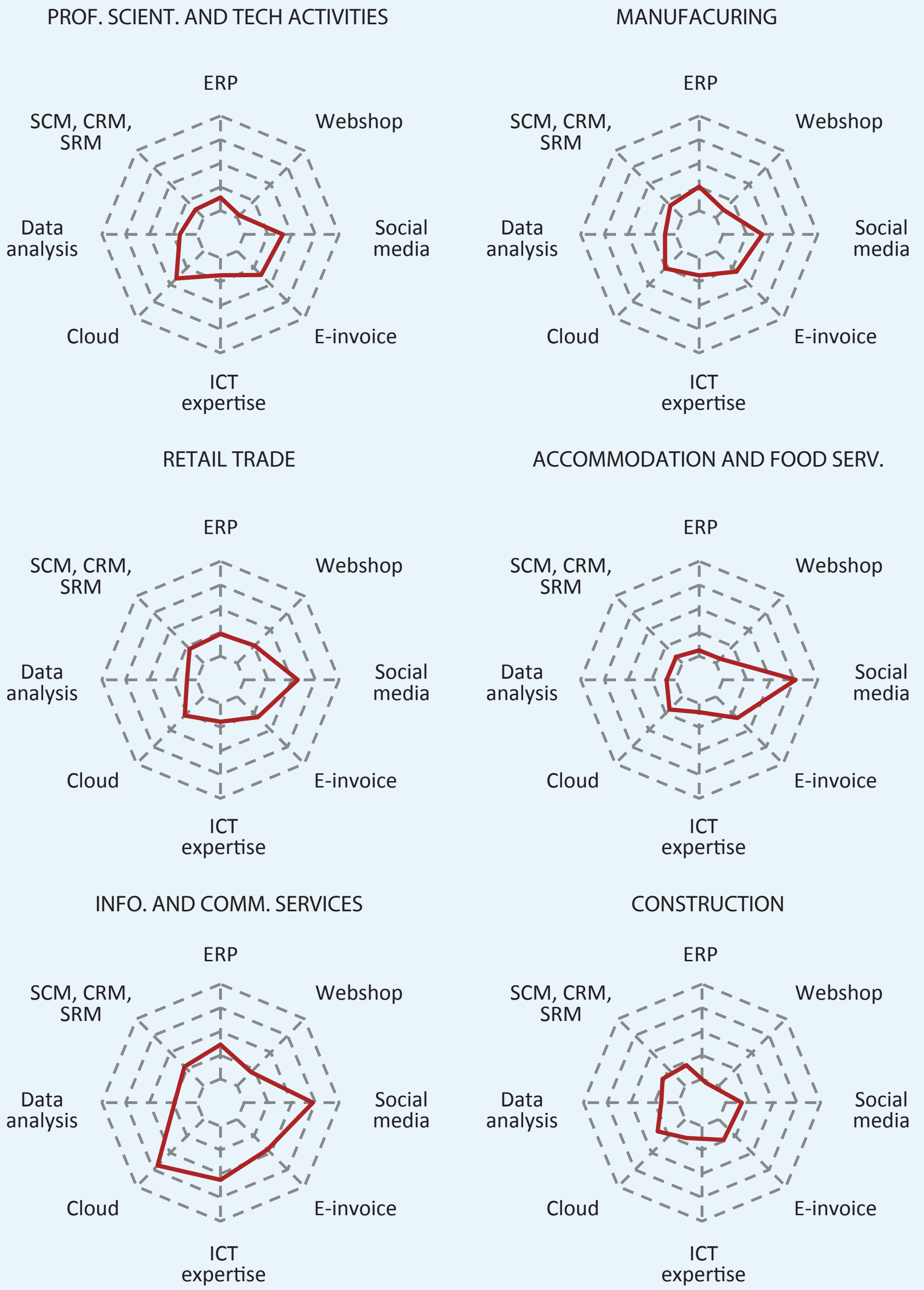


Table F15
The relationship between company characteristics and the digitisation indicator

	Dependent variable: **Digitisation headline index**
does not export	-6.041***
	(0.984)
company size: 5–9 persons	-2.429***
	(0.559)
company size: 20–49 persons	5.352***
	(0.847)
company size: 50–249 persons	5.610***
	(1.095)
company size: 250 or above	11.249***
	(2.182)
established: 2008 to 2014	0.003
	(0.558)
established: after 2014	0.359
	(0.930)
region: Southern Great Plain	-3.508***
	(0.830)
region: Southern Transdanubia	-3.650***
	(1.001)
region: Northern Great Plain	-3.781***
	(0.853)
region: Northern Hungary	-3.765***
	(0.990)
region: Central Transdanubia	-2.442***
	(0.899)
region: Western Transdanubia	-2.873***
	(0.903)
region: Pest	-1.230
	(0.763)
Number of observations	2,478
R^2	0.227
Adjusted R^2	0.217
Residual standard error	11.669 (df = 2445)
F stat	22.495*** (df = 32; 2445)

**p<0.1; **p<0.05; ***p<0.01. Standard errors in brackets. Estimated with a constant and sectoral dummy variables. Reference categories: Budapest region, companies founded before 2008, 10–19 employees, exporters*

MNB OCCASIONAL PAPERS 149

ICT USAGE INTENSITY IN THE HUNGARIAN CORPORATE SECTOR:

STYLISED FACTS ON MICRODATA

June 2023

Print: Prospektus Kft.

H-8200 Veszprém, Tartu u. 6.

mnb.hu

H-1013 BUDAPEST, KRISZTINA KÖRÚT 55.