ISSN 1831-9424

European Commission

JRC Technical Report

Methodology to project Digital Decade trajectories towards 2030

Torrecillas, J., Papazoglou, M., Cardona, M.,
Vázquez Prada-Baillet, M., Calza, E., Righi, R.

Editors:
López Cobo, M., De Prato, G.

2023

Joint Research Centre

EUR 31608 EN

Contact information
Address: Edificio Expo, C/ Inca Garcilaso 3, Sevilla 41092, Spain
Email: JRC-T1-SECRETARIAT@ec.europa.eu

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC133748

EUR 31608 EN

PDF ISBN 978-92-68-05749-0 ISSN 1831-9424 doi:10.2760/442136 KJ-NA-31-608-EN-N

Luxembourg: Publications Office of the European Union, 2023

How to cite this report: Torrecillas, J., Papazoglou, M., Cardona, M., Vázquez Prada-Baillet, M., Calza, E., Righi, R., *Methodology to project Digital Decade trajectories towards 2030*, López Cobo, M. and De Prato, G. editors, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/442136, JRC133748.

Contents

Abstract

The Digital Decade Policy Programme sets the path for the Digital Decade, and aims at strengthening the EU's digital economy for businesses and citizens to digitally transform European societies and economies. It has set specific targets by 2030 on four areas: digital skills, digital infrastructures, digital businesses and digitalisation of public services. It also proposes a monitoring system that will measure the progress of the European Union and its Member States towards the achievement of those targets.

In this context, the Joint Research Centre (JRC) of the European Commission (EC) in collaboration with the EC's Directorate General Communications Networks, Content and Technology (DG CNECT), launched a study aimed at supporting the monitoring of the Digital Decade key performance indicators towards their 2030 targets. This report presents the methodology developed to estimate the EU trajectories towards the Digital Decade targets. It explains the general framework and the empirical specification applied for estimating these trajectories for the targets with data availability at the moment of writing this report. It proposes the estimation of a baseline trajectory by extrapolating historical series, and a benchmark Digital Decade trajectory displaying a steady progressive path towards achieving the targets. The report also details the concrete application of the framework to each target, outlining specific adjustments and presenting the data used.

Acknowledgements

The authors would like to thank Paola Annoni (EC DG Communications Networks, Content and Technology (DG CNECT)) for providing critical insights to the methodology. Additionally, the authors are grateful for the collaborative work performed throughout the project with the following colleagues from DG CNECT: Valentina Parziale, Alexandra Rotileanu, Petra Kleininger, Eléonore Wemaere and Kristine Schuetterle. We also acknowledge the panel of experts who participated in the Workshop *Digital Decade targets: Methodology to estimate trajectories and investment gaps*, held online on May 3 2022. They provided valuable comments and useful critiques to a preliminary version of the methodology, from which this one represents an evolution: (in alphabetical order): Fabio Bacchini (Italian National Institute of Statistics (Istat)), Juan Fernández de Guevara Radoselovics (University of València, Spain and Valencian Institute of Economic Research (IVIE), Spain), • Aleix Pons (Fundación Cotec para la Innovación, Spain), Andreas Reinstaller (Austrian Institute of Economic Research (WIFO)), and Francesco Venturini (University of Perugia (Italy), NIESR (United Kingdom), CIRCLE (Sweden)).

Authors

Torrecillas, Juan

Papazoglou, Michail

Cardona, Melisande

Vázquez Prada-Baillet, Miguel

Calza, Elisa

Righi, Riccardo,

López Cobo, Montserrat (Editor)

De Prato, Giuditta (Editor)

1 Introduction

Digital technology has become a significant driver of social and economic transformation. Over the last years, technological change and innovation have driven productivity gains in Western countries (Byrne and Corrado, 2013; Gordon and Sayed, 2020). The digitalization of jobs has also profoundly affected labour markets by automatizing certain occupations (Acemoglu et al., 2020; Autor, 2022; Gónzalez-Vázquez et al., 2019). Furthermore, the diffusion of digital technology has improved the way governments and citizens interact, increasing public services efficiency (Nograsek and Vintar, 2014), healthcare access, education, and government accountability (Andersen 2009; Cegarra-Navarro, Córdoba and Moreno, 2012; Kim and Lee, 2019).

The Digital Compass Communication([1]) and the recently adopted Digital Decade Policy Programme 2030([2]) aim to support the development and adoption of digital technology in order to drive economic growth and competitiveness, as well as to improve citizens' lives and government effectiveness. More specifically, the Digital Compass Communication defines the European way for the Digital Decade along four cardinal points — digital skills, digital infrastructures, digital businesses and digitalisation of public services—. The Policy Programme sets the 2030 targets and proposes a monitoring system that will measure the progress of the European Union and its Member States towards the achievement of those targets. At the same time, it implements the ambitious vision and specific objectives presented in the Digital Compass Communication with the ultimate goal of digitally transforming European societies and economies.

In this context, the Joint Research Centre (JRC) of the European Commission (EC) in collaboration with the EC's Directorate General Communications Networks, Content and Technology (DG CNECT), launched a study aimed at supporting the monitoring of the Digital Decade (DD) key performance indicators (KPIs) towards their 2030 targets. In its first phase, the study provides estimations of the Digital Decade trajectories using simple econometric techniques. For each indicator, two different scenarios are reported: first, a baseline trajectory, which extrapolates the historical trend of the indicator towards 2030; and second, the Digital Decade trajectory, which illustrates a steady growth path towards reaching the target in 2030. The trajectories built with this methodology will provide insights into which targets are more likely to be achieved, and which ones may require more focused policy interventions in order to be met. The second phase of the study offers a mapping of the main investment funds that are instrumental in achieving the DD targets. Finally, the third phase will perform an international benchmarking of investment in order to assess amounts invested in each target with respect to other countries. The findings of this study will provide policy makers with valuable information on the state of digitalisation in the European Union and its progress towards meeting the Digital Decade targets.

This report focuses on the methodology developed to estimate the EU trajectories towards the Digital Decade targets. It explains the general framework and the empirical specification applied for estimating these trajectories. Two different models, depending on the KPIs nature, are employed. For indicators that exhibit a diffusion behaviour, the estimation is based on a logistic curve – also called S-curve, and non-linear least squares (NLS), which allows to capture the non-linear nature of the adoption process (Beal et al., 1957). Targets that show a linear behaviour in the past and that are expected to exhibit a steady evolution over time are estimated applying ordinary least squares (OLS) linear models with a time trend. The report also describes in detail how this framework is applied to each KPI, outlining particular adjustments and presenting the data used. Finally, it illustrates the procedure to calculate the Digital Decade trajectory to achieve the target. While results of the trajectories will be presented in the Communication establishing Union-level projected trajectories for the digital decade targets, this technical report should be considered a companion publication to explain in detail the methodology adopted in the former (for the KPIs to which this methodology is applicable), which will be published at a later stage.

The remainder of the report is organised as follows. Section 2 presents the data definition and sources for each KPI. Section 3 presents the general framework and the empirical approach to estimate the baseline trajectories and the Digital Decade scenario. Section 4 details the estimation procedure of each KPI. Finally, section 5 provides some concluding remarks.

[1] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and The Committee of the Regions 2030 Digital Compass: the European way for the Digital Decade (COM/2021/118 final) https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0118

[2] Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030 (OJ L 323, 19.12.2022, pp. 4–26). https://eur-lex.europa.eu/eli/dec/2022/2481/oj

2 Data: sources and definitions

This section presents the targets, and briefly describes the KPIs([3]), data sources and particularities of each KPI, for those targets with data availability at the time of writing this report, and for which a trajectory can be estimated following the methodology presented here.[4]

Table 1 summarises the target, its associated KPI, time coverage and the selected functional form for the estimation of the trajectory. For the sake of completeness, the table includes all the Digital Decade targets. Figure 1 displays the historical data for these KPIs, where breaks in series are marked with a vertical dashed line.

Table 1. Summary of Digital Decade targets and associated key performance indicators

Target	KPI (data source)	Time Coverage	Expected Behaviour
Basic Digital Skills: At least 80 % of those aged 16-74 have at least basic digital skills	Percentage of individuals aged between 16 and 74 years old disaggregated by sex with 'basic' or 'above basic' digital skills in each of the following five dimensions: information, communication, problem solving, digital content creation and safety skills (Eurostat, EU Survey on ICT Usage in Households and by Individuals)	2015-2021	Linear growth
ICT specialists: At least 20 million ICT specialists are employed within the Union, while promoting the access of women to this field and increasing the number of ICT graduates	Number of individuals aged 15-74 who are employed as ICT specialists; and gender convergence, measured as the percentage of women and men among those individuals employed as ICT specialists (Eurostat, Labour Force Survey)	2011-2022	Linear growth
Gigabit coverage: All end users at a fixed location are covered by a gigabit network up to the network termination point	Percentage of households covered by fixed Very High Capacity Networks (VHCN) (Broadband coverage in Europe studies). The evolution of the Fibre to the Premises (FTTP) coverage will also be monitored separately and taken into consideration when interpreting VHCN coverage data.	2013-2022	S-curve
5G coverage: All populated areas are covered by next-generation wireless high-speed networks with performance at least equivalent to that of 5G, in accordance with the principle of technology neutrality	Percentage of populated areas covered by at least one 5G network regardless of the spectrum band used (Broadband coverage in Europe studies)	2020-2022	S-curve
Semiconductors: Secure, resilient, performant and sustainable digital infrastructures where the production, in accordance with Union law on environmental sustainability, of cutting-edge semiconductors in the Union is at least 20 % of world production in value.	Value generated, in terms of revenues, by semiconductor activities in the Union, in all stages of the value chain, with respect to the global market value. For the first year, reporting will be done on the basis of those activities in Europe. (Study in progress, no data available)	-	-
Edge-nodes: Secure, resilient, performant and sustainable digital infrastructures where at least 10 000 climate-neutral highly secure edge nodes are deployed in the Union, distributed in a way that guarantees access to data services with low latency (i.e. a few milliseconds) wherever businesses are located.	Edge nodes, measured as the number of compute nodes providing latencies below 20 milliseconds; such as an individual server or other set of connected computing resources, operated as part of an edge computing infrastructure, typically residing within an edge data centre operating at the infrastructure edge, physically closer to its intended users than a cloud node in a centralised data centre. (Study in progress, no data available)	-	-
Quantum computing: Secure, resilient, performant and sustainable digital infrastructures where the Union has, by 2025, its first computer with quantum acceleration, paving the way for the Union to be at the cutting edge of quantum capabilities by 2030.	Quantum computing measured as the number of operational quantum computers or quantum simulators, including accelerators of High Performance Computing supercomputers, deployed and accessible to the user communities. (Detailed definitions can be found in the study *Key Performance Indicators for Quantum Technologies in Europe*, no data available)	-	-
Take-up of technologies: At least 75 % of Union enterprises have taken one of more of the following, in line with their business operations: (i) cloud computing services, (ii) big data, (iii) artificial intelligence	Cloud computing services: Percentage of enterprises that purchased at least one of a list of cloud computing services (Eurostat, EU survey on ICT usage and e-commerce in enterprises)	2014-2020 biennially and 2021	S-curve
	Big data: Percentage of enterprises analysing big data from any data source (internal or external). (Eurostat, EU survey on ICT usage and e-commerce in enterprises)	2016, 2018, 2020	S-curve
	Artificial intelligence: Percentage of enterprises using at least one artificial intelligence technology (Eurostat, EU survey on ICT usage and e-commerce in enterprises)	2021	S-curve
Digital late adopters: More than 90% of the Union SMEs reaching at least a basic level of digital intensity	SMEs with at least a basic level of digital intensity, measured as the percentage of SMEs using at least 4 of 12 selected digital technologies. (Eurostat, EU survey on ICT usage and e-commerce in enterprises, code: ISOC_E_DII)	2015-2022	-
Innovative businesses/scale-ups (Unicorns): The Union facilitates the growth of its innovative scale-	Sum of unicorns, defined as either: (a) an undertaking founded after 31 December 1990, which had an initial public offering or trade sale above USD 1 billion;	2008-2022	Exponential growth (log-

[3] The KPIs collected in this report are defined in the Commission Implementing Decision of 30.6.2023 setting out key performance indicators to measure the progress towards the digital targets established by Article 4(1) of Decision (EU) 2022/2481 of the European Parliament and of the Council. Further details can be consulted in the source of each KPI.

[4] Note that there is no data available for Semiconductors, Quantum, Edge. Additionally, data for Digital late adopters is not comparable over time due to yearly changes in the indicator definition.

Target	KPI (data source)	Time Coverage	Expected Behaviour
ups and improves their access to finance, leading to at least doubling the number of unicorns	or (b) an undertaking that has been valued at over USD 1 billion in its last private venture funding round, including where the valuation has not been confirmed in a secondary transaction (Dealroom)		linear model)
Digital public services: There is 100 % online accessible provision of key public services and, where relevant, it is possible for citizens and businesses in the Union to interact online with public administrations	Citizens: Online provision of key public services for citizens, measured as the share of administrative steps that can be done fully online for major life events for citizens. The following life events are considered: moving; transport; starting a small claims procedure; family; career; studying; and health. Businesses: The online provision of key public services for businesses measured as the share of administrative steps needed to start a business and conduct regular business operations, which can be done fully online. (eGovernment Benchmark)	2013-2022	S-curve
Electronic health record: 100 % of Union citizens have access to their electronic health records	(i) the existence of a nationwide mechanism, for citizen online access to health data, such as a patient portal, or a patient mobile app with additional measures in place that enable certain categories of people (guardians for children, people with disabilities, elderly) to also access their data, and (ii) the percentage of individuals that have ability to obtain or make use of their own minimum set of health-related data currently stored in public and private electronic health-record (EHR) systems (Empirica GmbH and PredictBy)	-	-
Electronic identification (eID): 100 % of Union citizens have access to secure electronic identification (eID) means that are recognised throughout the Union, enabling them to have full control over identity transactions and shared personal data	Measured by two KPIs: (1) number of Member States that have notified at least one national eID scheme in accordance with Regulation (EU) No 910/2014; and (2) number of Member States that have provided access to secure privacy-enhancing eID via the European Digital Identity Wallet in accordance with the Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity (The European Digital Identity Framework)	-	-

Basic digital Skills

The Digital Decade Policy Programme sets the target that, by 2030, 80% of citizens aged 16-74 should have at least basic digital skills. The KPI used to monitor this target is the “At least basic digital skills”, defined as the percentage of individuals aged between 16 and 74 years old with “basic” or “above basic” digital skills in each of the following five dimensions: information, communication, problem solving, digital content creation, and safety skills, considering activities carried out during the previous three months. The source of this KPI is the Digital Skills Indicator (DSI 2.0) from the European Union Survey on ICT Usage in Households and by Individuals. The latest edition of the survey in 2021 included some methodological changes that lead to a break in series. These changes include: variation in the mode of data collection from face-to-face to telephone or online interviews due to the Covid pandemic; changes in the sampling of some countries; and the modification of the DSI by the inclusion of a fifth dimension of digital skills([5]). Consequently, data from 2015 to 2019 are not comparable with results from the 2021 survey.

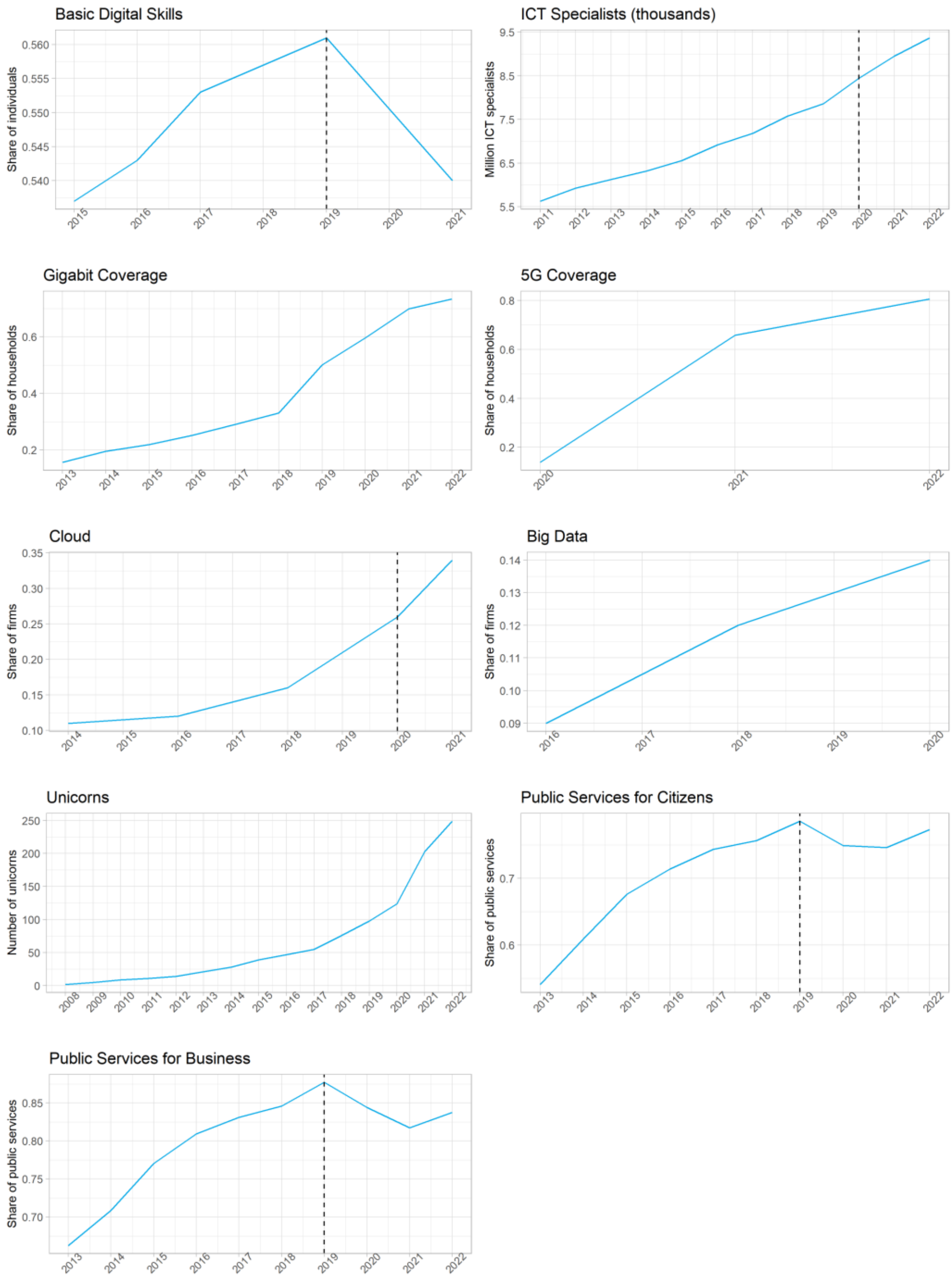
ICT specialists

The Digital Decade Policy Programme has set the target to have 20 million information and communication technology (ICT) specialists in employment by 2030, promoting the access of women. The KPI to monitor the evolution of this target is the number of individuals aged 15-74 who are employed as ICT specialists , derived from the Eurostat’s Labour Force Survey[6]. More specifically, the indicator is operationalised in terms of ISCO codes, with a list of ISCO-88 codes for the period 2004-2010, and using the Eurostat-OECD statistical definition based on a list of ICT specialists’ ISCO-08 occupations as of 2011 onwards (OECD, 2015). This includes jobs such as ICT service managers, ICT professionals, ICT technicians, electronics and telecommunications installers and repairers. Data for this indicator is available from 2004 to 2022, but due to the break in the series in 2011 caused by the revision of the ISCO (from ISCO-88 to ISCO-08), series are only comparable for the period 2011-2022.

[5] The previous version of the DSI indicator (DSI 1.0) is computed based on four dimensions of digital skills instead of five (the same ones than DSI 2.0, except “safety skills”).

[6] Following OECD (2004, p. 216), Eurostat defines ICT specialists as "workers who have the ability to develop, operate and maintain ICT systems, and for whom ICT constitute the main part of their job". https://ec.europa.eu/eurostat/cache/metadata/en/isoc_skslf_esms.htm

Figure 1. KPIs historical data



Note: The vertical dotted line indicates breaks in historical series of the indicator in that year. Source: Authors' elaboration based on data from official sources.

Gigabit network coverage

The target specifies that 100% of end users at a fixed location should be covered by a Gigabit network up to the network termination point. To monitor this target, the KPI employed is the percentage of households covered by fixed Very High Capacity Networks (VHCN). The technologies considered as VHCN are fibre to the premises (FTTP) and DOCSIS([7]) 3.1. The evolution of FTTP will be monitored as a proxy for the growth of VHCN. The source is the Broadband coverage in Europe studies for the EC, by Omdia and Point Topic. Data are available for the period 2013-2022.

[7] Data Over Cable Service Interface Specification.

5G coverage

The target for 5G coverage is that all populated areas are covered by next generation wireless high-speed networks with at least 5G equivalent performance, in accordance with the principle of technology neutrality. To provide monitoring, the KPI is defined as the percentage of populated areas covered by at least one 5G network regardless of the spectrum band used. The source is the Broadband coverage in Europe studies for the EC, by Omdia and Point Topic. Data availability for 5G data exists for 2020-2022.

Take-up of digital technologies: Cloud computing services, Big data and Artificial intelligence

The target specifies that at least 75% of EU enterprises should have taken up one or more of the following technologies: cloud computing services, big data, artificial intelligence.

Cloud computing services: In order to monitor the evolution of cloud computing adoption, the KPI is computed based on the European Union survey on ICT usage and eCommerce in Enterprises, and defined as the share of enterprises that purchased, at least, one of a list of advanced services. Since 2021, the indicator changed the services included and is defined as the share of enterprises that purchased, at least, one of the following services: finance or accounting software applications, enterprise resource planning (ERP) software applications, customer relationship management (CRM) software applications, security software applications, hosting the enterprise's database(s), and computing platform providing a hosted environment for application development, testing or deployment. Data available for 2014, 2016, 2018, 2020 and 2021.

Big data: To monitor the evolution of this indicator, the KPI is based on the European Union Community survey on ICT usage and eCommerce in Enterprises. More specifically, the variable used is defined as "the share of enterprises analysing big data from any data source", which provides information for 2016 and 2018; and "the share of enterprises analysing big data internally from any data source or externally" for 2020.

Artificial intelligence: The KPI is calculated based on the European Union Community survey on ICT usage and eCommerce in Enterprises and defined as the share of enterprises that use at least one AI technology. Starting in 2021, and every two years, the Eurostat survey includes a dedicated AI module to provide the official data on AI adoption in the EU and its Member States. The sample is, hence, limited to one data point in 2021.

Innovative businesses/scale-ups (Unicorns)

By 2030, the EU intends to double the number of unicorns. For this trajectory, the KPI is based on Dealroom, a global provider of data and information on Startups and Technology Ecosystems. Dealroom defines a unicorn as a firm that fulfils the following conditions: (i) have been founded after 1990 (excluding subsidiaries or joint ventures of established companies), (ii) are currently valued at over $1 bn or have successfully completed a $1+ bn exit, and (iii) are regarded as "tech". Each unicorn is counted from the first time it was valued at $1 bn. After considering the aforementioned conditions the final sample consists of 15 observations over the 2008-2022 period([8]). Regarding the "tech" definition by Dealroom, it does not consider companies that have more traditional business models (e.g., telecommunication or airplane leasing company), but includes other sectors arguably related to "tech", like online betting companies.

Digital Public Services

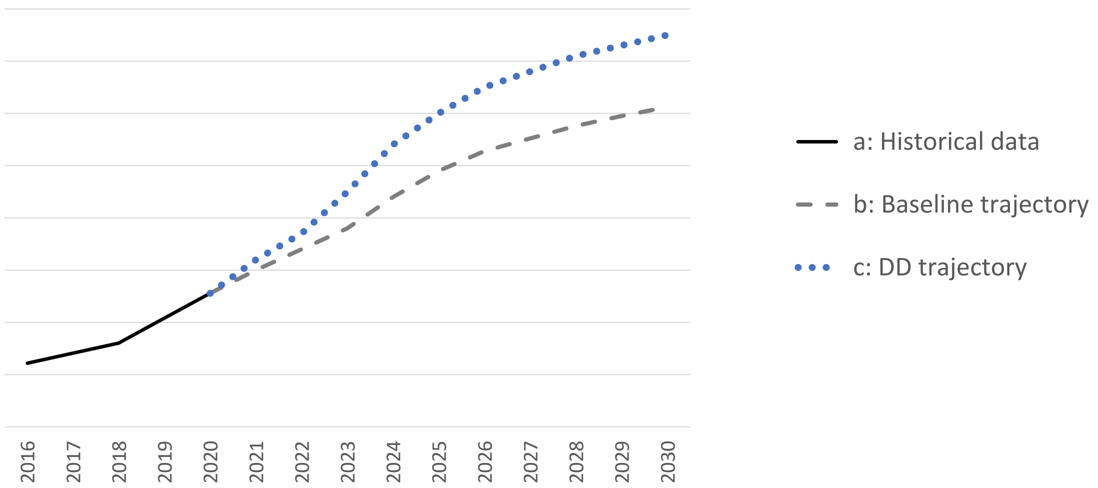
The target aims to have 100% online accessible provision of key public services. To estimate the online provision of key public services, the KPI comes from the eGovernment Benchmark's Online Availability indicator, defined under the User Centricity top-level benchmark. The indicator for online provision of digital public services to citizens is computed as the average of online availability of services for citizens and for citizen mobility (services directed to foreign citizens), to also account for cross-border services. Until 2019, the key public services for citizens considered by the eGovernment Benchmark were services related to the following life events: career (including losing and finding a job), studying, family, moving, transport (owning and driving a car), starting a small claims procedure. After 2019, these life events changed and included: moving; transport; starting a small claims procedure; family; career; studying; and health. The key services for businesses considered by the benchmark referred to: business start-up and regular business operation. As the eGovernment Benchmark measures a different set of services in odd and even-numbered years, the indicator uses biennial averages of yearly data.

[8] A new year is included only when the December data is available on Crunchbase.

3 General framework for the estimation of trajectories

For each indicator, we project two trajectories: a baseline trajectory, which extrapolates the historical trend of the indicator towards 2030, and the Digital Decade trajectory, which illustrates the ideal path towards the achievement of the targets in 2030. Figure 2 illustrates a theoretical trajectory, showing the baseline trajectory in the grey dashed line, the DD trajectory in the blue dotted line and the historical series in the black solid line.

Figure 2. Baseline and Digital Decade trajectories (example with an S-curve)



For estimating the trajectories of the Digital Decade KPIs, we employ two different models, depending on the nature of the indicator:

— On the one hand, indicators that exhibit a technology diffusion nature are estimated using an S-curve model. This model is appropriate for describing user's behaviour regarding technology adoption. It starts with an initial slow growth, followed by faster uptake to finally plateau near the saturation point (Beal et al., 1957). Growth may decline as late adopters are usually more reluctant to adopt the new technology, or because infrastructure rollout cost increases for less populated areas.

— On the other hand, we estimate an OLS linear model for those indicators in which a linear growth is observed in the past and expected to continue in the future.

A key aspect in forecasting trajectories is data availability and comparability over time. Ideally, a long time series should be available for each KPI to monitor the evolution of the Digital Decade. In some cases, Digital Decade KPIs are already monitored within the European Commission's Digital Economy and Society Index (DESI), with sources from Eurostat, the eGovernment Benchmark, or other ad hoc studies launched by the EC, and can therefore be readily monitored and projected. Thus, the main difficulties to implement the methodology relate to data scarcity (e.g. artificial intelligence, 5G) and breaks in series due to changes in the indicator's definition (e.g., basic digital skills, digital public services). While some KPIs present changes in the indicator definition (e.g., basic digital skills or digital public services) that limit comparability over time, in other cases there is no existing data to measure it, and the analysis must rely on the use of proxies. We will address these concerns in section 4. Finally, the absence of data for some KPIs; such as semiconductors, edge, quantum, electronic health records and electronic identity; makes not possible the estimation of trajectories, hence they are not covered in sections 3 and 4. In the case of digital late adopters there is data availability, but yearly changes in the definition of the KPI makes it not possible to estimate a trajectory. These limitations also are behind the choice of a simple econometric approach, leaving aside other more sophisticated methods like time series analysis. We also discarded the introduction of control variables to account for country specific variation (e.g. GDP per capita) due to the reduced number of data observation in many of the KPIs.

3.1 Empirical specification: Baseline trajectories

We define baseline trajectories as projections in a business-as-usual scenario, namely, assuming no external shocks or future policy interventions. To calculate baseline trajectories, the following procedure is applied and further adapted to each KPI([9]):

1. We collect data for the corresponding KPI.
2. We estimate the model selected.
3. Using the coefficients obtained in step 2, we project the evolution of the KPI up to 2030.

Analytically, this can be expressed as:

$$Y_{t,i} = F_i(t) + \varepsilon_{t,i} \qquad [1]$$

Where $Y_{t,i}$ is the value of indicator *i* in time t, and $F_i(t)$ is the functional form of indicator *i* and depends on time, and $\varepsilon_{t,i}$ is an i.i.d (independent and identically distributed) normally distributed error term with zero mean and constant variance. Thus, the baseline trajectory can be expressed as follows:

$$\forall\,\{t \leq T\}, \hat{Y}_{t+k|t,i} = F_i(t+k)\,;\forall i \in I = \{L, J\}$$

Where $\hat{Y}_{t+k|t,i}$ is the indicator *i* forecast for time *t+k* conditional on time t information; and F(t) is a function of *t*. *I* is a set that contains all KPIs, L is a subset of *I* that contains KPIs estimated using OLS, and J is the subset of KPIs estimated using an adoption model.

3.2 Estimation of baseline trajectories

3.2.1 Estimation of S-curve: Technology diffusion

An S-curve diffusion model is estimated for the following indicators: gigabit network coverage, 5G coverage, take-up of digital technologies (cloud computing services, big data, and artificial intelligence), and digital public services (for business and citizens). Following Gruber and Verboven (2001), we estimate a logistic model of diffusion using NLS. Concretely, the model in equation [1] takes the following form:

$$\forall\,\{t \leq T\}, Y_{t,j} = \frac{M_j}{1+e^{-\left(a_j+b_jt+\varepsilon_{t,j}\right)}}\,;\forall j \in J \qquad [2]$$

Where M_j represents the saturation point of the diffusion curve of indicator *j*, a_j is a parameter that represents the timing of diffusion, i.e., that shifts the curve backwards or forwards without affecting its shape([10]); b_j represents the speed of diffusion; and $Y_{t,j}$ is the value of KPI j at time t. Accordingly, diffusion rates are increasing in a, with higher values of a associated with higher initial levels of coverage or adoption; and in b, where larger values of b represent a higher speed of diffusion. Finally, $\varepsilon_{t,j}$ is the error term. The saturation level (parameter M) is set to 100% for all cases, except for take-up of digital technologies, where the saturation point is set to 80%([11]).

We first estimate Equation [2] using NLS to obtain estimates of parameters $\hat{a}$ and $\hat{b}$. After that, parameter *a* is set such that the value of the projection in the last year of the historical series coincides with the observed value in that year, given $\hat{b}$; this allows continuity between the historical series and the projection. Concretely, we assume that the timing parameter *a* is set such that the cumulative distribution function fulfils the following identity:

$$F_j\left(Y_{t=T,j}\right) = \frac{M_j}{1+e^{-(a_j+\hat{b}_jT)}} = z_{t,j=T} \qquad [3]$$

Where T equals 0 in the last year of available data and increases sequentially by 1 every year up to 2030. Z is the value of the indicator in the last year of the historical series. Solving for *a*, we have that the timing of diffusion is:

[9] The modifications applied in each case are detailed in section 4.

[10] The rationale behind this adjustment is that adoption series data is not available from adoption equal to 0, so we adjust the parameter to ensure continuity between the historical series and the trajectory.

[11] We assume that adoption could still increase beyond the 75% target.

$$\bar{a}_j = -(ln\left(\frac{M}{z_{t=T,j}} - 1\right) + \hat{b}_j T) \qquad [4]$$

Finally, we employ these parameters' estimates to calculate the projected trajectories of indicator *j* up to *t+k* conditional on time *t* information:

$$\hat{Y}_{t+k|t,j} = \frac{M_j}{1 + e^{-(\bar{a}_j + \hat{b}_j (t+k))}}$$

Where $\hat{b}$ represents the estimated speed of diffusion from equation [2] and $\bar{a}$ is the adjusted timing.

3.2.2 Estimation of linear trajectories using OLS regression

We estimate a linear model using OLS to forecast the following indicators: basic digital skills, ICT specialists, and unicorns (innovative businesses / scale-ups)[12]. The model is specified as follows:

$$Y_{t,l} = \alpha_l + \beta_l t + \varepsilon_{t,l} \quad \forall\, l \in L \qquad [5]$$

Where $Y_{t,l}$ is the value of indicator *l* in time t, α is a constant term and t is a time trend. Finally, $\varepsilon_{t,l}$ is the error term.

For linear trajectories, we use OLS to estimate equation [5]. Finally, we use these parameters' estimates to forecast the evolution of indicator *l* up to t=2030 using the following expression:

$$\hat{Y}_{t+k|t,l} = \hat{\alpha}_l + \hat{\beta}_l\,(t+k)$$

Where hatted parameters represent OLS estimates from equation [5].

3.3 Approach to deal with breaks in series

One of the main drawbacks of our empirical exercise is the presence of breaks in series due to changes in definitions in many of the analysed indicators. To solve this issue, we propose a slight modification of the main model in order to estimate the size of the break. Concretely, we add a dummy variable, *d*, that takes value 1 after the year of break in series, T^b, and 0 otherwise, as follows:

$$Y_{t,j} = \frac{M_j}{1 + e^{-(a_j + b_j t + c d_t + \varepsilon_{t,j})}}$$

Or, in the case of a linear model:

$$Y_{t,l} = \alpha_l + \beta_l t + \delta \mathrm{d_t} + \varepsilon_{t,l}$$

If we assume that no other structural breaks happened in year T^b, parameters *c* and δ can be interpreted as a measurement of the magnitude of the break in series. Finally, trajectories are calculated using parameter *c* and δ to adjust the whole series to the most recent definition. This method is a refined version of the error prediction approach proposed to deal with indicator's breaks in series and used to backcast series with breaks, as done with DESI indicators([13]).

3.4 Estimation of Digital Decade benchmark trajectories

The benchmark DD trajectory represents the theoretical path towards the achievement of the target by 2030, following the same functional form as its baseline counterpart. In the linear model, we can interpret this trajectory as the minimum progressive increment that the indicator should experience in order to reach the DD target in 2030. In the S-shaped KPIs, the DD trajectory can be interpreted as the one resulting from the minimum speed of diffusion that the indicator should experience in order to reach the DD target by 2030, given the baseline value or starting point.

In practice, for KPIs with a linear specification, the DD trajectory is calculated as the linear interpolation between the target and the last observed value available. That is:

[12] In the case of unicorns, we employ a log-linear transformation of Equation 5.

[13] See Annex for a technical discussion of both approaches.

$$Y_{dd} = y_0 + \frac{(t - t_0)(target - y_0)}{(t_1 - t_0)}$$

Where Y_{dd} is the Digital Decade trajectory, y_0 is the last observed value available, t represents time, $target$ is the value of the target, t_1 is year 2030 and t_0 is the year of the last observed value available.

For KPIs with S-shape specification, the DD trajectory is calculated by solving the speed of diffusion assuming the indicator reaches the target value in 2030.

$$target = \frac{M}{1 + e^{-(\bar{a} + \tilde{b} t_1)}}$$

Therefore, solving for $\tilde{b}$:

$$\tilde{b} = \frac{-\ln\left(\frac{M}{target} - 1\right) + \bar{a}}{t_1}$$

Which gives the speed of diffusion required to achieve the Digital Decade target in 2030.

4 Specific Applications to each KPI

Basic digital skills

We estimate the trajectory of basic digital skills using an OLS linear model. We assume that the indicator will steadily increase over the next years.

Because of the break in the series in 2021, data from 2015-2019 is not comparable with the most recent release. Therefore, we assume that data following the new DSI 2.0 will follow a trend similar to that of the DSI 1.0. We proceed as follows: first, we estimate the linear model using data up to 2019. Second, we adjust the intercept of the regression so that the estimation for 2021 coincides with the 2021 observed value using the DSI 2.0. Lastly, we calculate the trajectory up to 2030 using this adjusted intercept and the slope from the OLS results.

ICT specialists

We use OLS on a linear regression model to estimate the trajectory of ICT specialists up to 2030 using data from 2011 to 2022.

Gigabit coverage

We use FTTP coverage data from the Broadband coverage studies to obtain parameter speed of diffusion, which provides us with 9 observations over the 2013-2021 period. The reason to use FTTP coverage data is that operators deploy optical fibre infrastructure in response to competition pressure in cable footprint areas, or where first-mover advantage can be realised with reasonable cost per household (EIB, 2022). Consequently, the additional rollout will have to expand the existing optical fibre coverage footprint in 2017 outside cable areas. Coaxial (DOCSIS) footprints are not expanding and are not expected to expand significantly and if cable operators do expand their footprint, they will use FTTP. Therefore, future gigabit developments will expectedly reflect the rollout characteristics of FTTP. With this in mind, we estimate the diffusion speed of the FTTP data, and apply parameter estimates to the VHCN indicator.

5G coverage

To model the evolution of 5G coverage, we estimate a logistic curve of diffusion using NLS using data from 2020 to 2022.

Take-up of technologies – Cloud computing services

The sample contains 5 observations corresponding to 2014, 2016, 2018, 2020 and 2021. We apply a logistic curve of diffusion model and the adjustment for breaks in series outlined in sub-section 3.3 to estimate this trajectory.

The literature suggests that an S-shaped adoption curve is appropriate to model the evolution of cloud computing services take-up. For instance, Bloom and Pierri (2018) show that the adoption of cloud computing services in the US follow an S-curve, starting with a fast uptake from 2010 to 2014 and reaching a plateau between 2014 and 2016. According to PublicFirst (2022), the adoption cost of cloud computing services might still be expensive for SMEs which could delay their adoption and provoke slower growth near the saturation point, a behaviour consistent with adoption theory (Beal et al., 1957).

Hence, we estimate the trajectory of cloud computing services using a logistic curve. For this adoption curve, we set the market saturation point at 80% to allow for further increase after the target is reached.

Take-up of technologies – Big data

The technology adoption rationale applies here too. The projected trajectory follows an S-shaped curve, where growth is expected to be slow in the beginning, to further experience an acceleration period and finally a slower final uptake, likely driven by small firms or firms in specific sectors that may have less incentive to adopt big data technologies. For estimation purposes, it is assumed that the distinction "internally […] or externally" introduced in 2020 does not affect the series' comparability over time, hence we can consider a time series of three data points.

For this adoption curve, we set the market saturation point at 80%, above the target of 75% to allow for possible increments after the target is reached.

Take-up of technologies – Artificial intelligence

The technology adoption literature suggests an S-shaped curve for the diffusion variable (Bughin, 2018). However, the sample contains just one observation, which makes the estimation not feasible. Since artificial intelligence could be interpreted as a related and broader technology than big data , we use big data's estimated speed of diffusion to proxy that of artificial intelligence adoption. The point of market saturation is set to 80%, to account for the possibility of further adoption beyond the Digital Decade target.

Innovative businesses/scale-ups (Unicorns)

The increase in unicorns has accelerated in recent years, and the number of unicorns has risen from 60 in 2017 to 136 in 2020 and 249 at the end of 2022, which suggests the use of a log-linear model to estimate this trajectory. Parameter $\hat{\beta}$ in the log-linear transformation, differently from other KPIs estimated using OLS, can be interpreted as the percentage increase in unicorns after one time period increase.

Digital Public Services

We estimate the trajectory of the provision of digital public services using a logistic curve of diffusion. There are two reasons to think that an S-shaped curve would fit the future behaviour of the indicator. First, services that are more costly to provide online may be left until the end of the digitalization process. Second, services in rural areas, where there are fewer digital infrastructures, might be more difficult to digitalise. Thus, we can expect a slower growth near the saturation point of 100%, which is typical of diffusion curves. We assume that all key public services are suitable to be provided online, and thus that the saturation point is 100%.

As mentioned in section 2, the change in services measured before and after the 2019 data release, as specified in the eGovernment benchmark's method paper 2020-2023 (EC, Directorate-General for Communications Networks, Content and Technology, 2021), makes the series not comparable over time. Thus, we apply the adjustment for indicators with breaks in series outlined in sub-section 3.3.

5 Conclusions

Digitalization is of utmost importance for the future of Europe. In this direction, the Digital Decade Policy Programme is expected to critically contribute to accelerate the digitalisation of the economy and society. Hence, monitoring the progress of the Digital Decade targets is crucial to guarantee efficiency and to support evidence-based decision-making.

This report describes the methodology developed to estimate the EU-level trajectories towards the Digital Decade targets. It analyses the general framework applied and the specific applications to each KPI. It also details the empirical specification of the two different models employed: linear and S-curves. Several adaptations of this general framework to each KPI have been presented, most of them tackling breaks in series due to changes in the indicator’s definition.

The projection of trajectories is severely constrained by data availability. While many indicators present a small sample size that may yield unstable estimates —e.g. 5G coverage, big data, artificial intelligence—, others present breaks in series that limit comparability over time and, hence, the precision of the estimates —e.g., digital public services, cloud computing or basic digital skills—.

Next steps would include the application of the methodology to obtain the EU level trajectories and the gap between the baseline and Digital Decade trajectories. The analysis of the mapping of monetary investments concurring to the achievement of the Digital Decade targets and the international benchmarking of investments supporting the targets in order to assess amounts invested with respect to other countries are aspects to be covered in future work.

References

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of political economy*, *128*(6), 2188-2244.

Andersen, T. B. (2009). E-Government as an anti-corruption strategy. *Information Economics and policy*, 21(3), 201-210.

Autor, D. (2022). *The labor market impacts of technological change: From unbridled enthusiasm to qualified optimism to vast uncertainty* (No. w30074). National Bureau of Economic Research.

Beal, G.M. & Rogers, E.M. (1957). The importance of personal influence in the adoption of technological change. *Soc. F.*, *36*, 329.

Bloom, N., & Pierri, N. (2018). Cloud computing is helping smaller, newer firms compete. Harvard Business Review, 94(4).

Bughin, J. (2018). Wait-and-See Could Be a Costly AI Strategy. *MIT Sloan Management Review*, 59(4), 1-4.

Byrne, D. M., & Corrado, C. A. (2017). ICT Services and their Prices: What do they tell us about Productivity and Technology?. SSRN Working Paper.

Cegarra-Navarro, J. G., Pachón, J. R. C., & Cegarra, J. L. M. (2012). E-government and citizen's engagement with local affairs through e-websites: The case of Spanish municipalities. *International Journal of Information Management*, *32*(5), 469-478.

European Commission, Directorate-General for Communications Networks, Content and Technology, (2021). eGovernment benchmark: method paper 2020-2023, Publications Office. https://data.europa.eu/doi/10.2759/640293

Gonzalez Vazquez, I., Milasi, S., Carretero Gomez, S., Napierala, J., Robledo Bottcher, N., Jonkers, K., Goenaga, X. (eds.), Arregui Pabollet, E., Bacigalupo, M., Biagi, F., Cabrera Giraldez, M., Caena, F., Castano Munoz, J., Centeno Mediavilla, C., Edwards, J., Fernandez Macias, E., Gomez Gutierrez, E., Gomez Herrera, E., Inamorato Dos Santos, A., Kampylis, P., Klenert, D., López Cobo, M., Marschinski, R., Pesole, A., Punie, Y., Tolan, S., Torrejon Perez, S., Urzi Brancati, C., Vuorikari, R. (2019) *The changing nature of work and skills in the digital age*, EUR 29823 EN, Publications Office of the European Union, Luxembourg.

Gordon, R. J., & Sayed, H. (2020). *Transatlantic technologies: The role of ICT in the evolution of US And European productivity growth* (No. w27425). National Bureau of Economic Research.

Gruber, H., & Verboven, F. (2001). The diffusion of mobile telecommunications services in the European Union. *European Economic Review*, 45(3), 577-588.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., & Simoes, A. (2014). The atlas of economic complexity: Mapping paths to prosperity. Mit Press.

ILO (2012) International Standard Classification of Occupations: ISCO-08. Structure, group definitions and correspondence tables, Volume I, ILO, Geneva

Kim, S., & Lee, J. (2019). Does E-Government Raise Effectiveness and Efficiency?: Examining the Cross-National Effect. Public Performance & Management Review, 42(4), 1000-1023.

Nograšek, J., & Vintar, M. (2014). E-government and organisational transformation of government: Black box revisited?. *Government Information Quarterly*, 31(1), 108-118.

OECD (2004), Information Technology Outlook 2004, OECD Publishing, Paris. Available at: https://www.oecd.org/sti/ieconomy/37620123.pdf

OECD (2015), "Proposal for an Eurostat-OECD definition of ICT specialists", available at: http://ec.europa.eu/eurostat/cache/metadata/Annexes/isoc_skslf_esms_an1.pdf

PublicFirst (2022), Unlocking Europe's Digital Potential. Available at: https://awsdigitaldecade.publicfirst.co/

List of figures

List of tables

Annexes

Annex 1. Comparison of two approaches for trajectories with breaks in series

A1.1. Approaches to account for breaks in series

As explained in sub-section 3.3 *Approach to deal with breaks in series*, we propose to add a dummy variable in the main model to estimate the size of the break. The dummy variable, d, takes value 1 after the year of the break, Tb, and 0 otherwise, as follows in the S-curve model:

$$Y_{t,j} = \frac{M_j}{1 + e^{-(a_j + b_j t + c d_t + \varepsilon_{t,j})}}$$

And, in the case of a linear model:

$$Y_{t,j} = \frac{M_j}{1 + e^{-(a_j + b_j t + c d_t + \varepsilon_{t,j})}}$$

If we assume that no other structural breaks happened in year Tb, parameters *c* and δ can be interpreted as a measurement of the magnitude of the break in series. Finally, trajectories are calculated using parameter *c* and δ to adjust the whole series to the most recent definition.

We have tested another approach, the Prediction error approach, to estimate the size of breaks in series. It starts by estimating the model using information only up to the last year before the break in series. Then, it uses these parameters' estimates to predict the value of the indicator in the years after the break in series. The size of the break is calculated as the difference between the predicted value using only information up to the last year before the break and the observed value for that same year. When there is more than one observation after the break in series, a simple average of the prediction error is taken. Finally, the whole series are adjusted by adding the estimated size of the break to have a time-comparable series, and the model is estimated using the whole adjusted series.

A1.2 Performance comparison

Although both approaches render similar results, the dummy variable one builds on the prediction error approach to provide more efficient results under classical assumptions and, using simulated data, we show that it can estimate the true size of the effect of the break in series more often and with a smaller mean error.

To see this, consider the following model:

$$Y_t = \alpha + \beta t + \delta d_t + \varepsilon_t$$

Where d is a break in the series that changes the equation intercept by δ units after time T^b and until the end of the series in time T. Now, suppose we do not include this term explicitly in the regression estimates. Then, the error term will capture the structural break and the equation would take the following form:

$$Y_t = \alpha + \beta t + u_t$$

Where $u_t = \delta d_t + \varepsilon_t$

If we consider classical assumptions (random sampling, normally distributed errors with 0 mean, no omitted variables, no collinearity), then the *Dummy approach* estimates δ using OLS and the *Prediction error* approach estimates δ as $E[\hat{Y}_{T^b+(T-T^b)|T^b} - Y_{T^b+(T-T^b)}]$. It can be shown that both approaches provide unbiased estimates if these assumptions hold, so the question bottoms down to which estimator is more efficient. Under classical assumptions, OLS is Best Linear Unbiased Estimator, which implies that it has minimum variance. Similar claims can be made for the Non-Linear Least Squares estimator used for S-curves. Hence, this theoretical better efficiency of using OLS and NLS justifies our choice.

Additionally, using simulations it can be shown that the *Dummy approach* provides more precise estimates than the *Prediction error* one. Let's create the following data generating process (DGP), where y0 is generated by a linear process with normal error with 0 mean and standard deviation of 1. We first generate data before there is any break in the series.

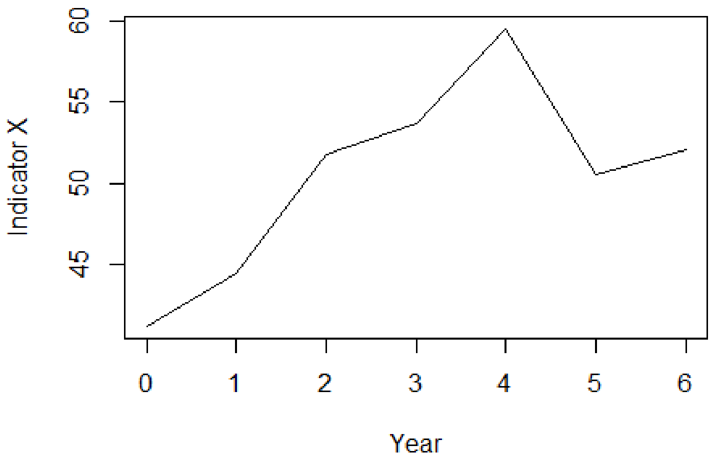
$$Y_t = \alpha_o + \beta t + \varepsilon_t$$

After t=4, we assume a break in series that reduces the level of the variable by 15 units. We generate two data points with this process, simulating the situation with the breaks in series in the real setting. Hence, the new DGP becomes the following:

$$Y_t = \alpha_1 + \beta t + \varepsilon_t$$

Where $\alpha_1 = \alpha_0 + \delta$ and δ represents the size of the break in series. Figure A. 1 shows the simulated series, where there is a visible structural break between year 4 and 5:

Figure A. 1 Example of series with a structural break using simulated data



Using this DGP, we run 10,000 simulations and estimate δ using both approaches – Dummy and Prediction error-, obtaining the following results:

Table A. 1 Metrics of the estimation of $\boldsymbol{\delta}$ with Dummy and Prediction error approaches – Linear model

Approach	Mean Absolute Error	Variance of Mean Absolute Error	Mean Squared Error	Variance Mean Squared Error	% cases with closer value predicted
Dummy variable	0.5463	0.1691	0.4683	0.4420	52.92%
Prediction error	0.5560	0.1737	0.4830	0.4683	47.08%

From Table A. 1, it is visible that overall the Dummy variable approach performs better than the Prediction error one. On the one hand, the Dummy variable approach presents smaller mean absolute and mean squared errors when estimating the real value of δ**.** On the other hand, it also have less error variance. Finally, it provides a closer value to the real size of the break slightly more often than the Prediction error approach. This evidence is consistent with the theoretical predictions outlined and suggests that the dummy approach is better suited to account for the size of breaks in series.

Analogously, we can repeat this exercise for the logistic curve model. Similar conclusions can be drawn from this model (Table A. 2). The Dummy approach provides a closer estimate to the real size of the break more often and has a slightly smaller mean absolute and squared error, although in this case the variance is lower in the Prediction error approach.

Table A. 2 Metrics of the estimation of **"c"** with Dummy and Prediction error approaches – Logistic model

Approach	Mean Absolute Error	Variance of Mean Absolute Error	Mean Squared Error	Variance Mean Squared Error	% Closer value predicted
Dummy variable	0.3349	0.0449	0.1570	0.0283	55.90%
Prediction error	0.3372	0.0447	0.1584	0.0273	44.10%

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

@eu_science

Publications Office
of the European Union