Do Data Policy Restrictions Inhibit Trade in Services?

Martina Francesca Ferracane and Erik van der Marel

European University Institute

Robert Schuman Centre for Advanced Studies

Global Governance Programme

Do Data Policy Restrictions Inhibit Trade in Services?

Martina Francesca Ferracane and Erik van der Marel

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies, created in 1992 and currently directed by Professor Brigid Laffan, aims to develop inter-disciplinary and comparative research on the major issues facing the process of European integration, European societies and Europe's place in 21st century global politics.

The Centre is home to a large post-doctoral programme and hosts major research programmes, projects and data sets, in addition to a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration, the expanding membership of the European Union, developments in Europe's neighbourhood and the wider world.

For more information: http://eui.eu/rscas

The Global Governance Programme

The Global Governance Programme is one of the flagship programmes of the Robert Schuman Centre. It is made of a community of outstanding professors and scholars, produce high quality research and engage with the world of practice through policy dialogue. Established and early career scholars research write on and discuss issues of global governance within and beyond academia, focusing on four broad and interdisciplinary areas: Global Economics, Europe in the World, Cultural Pluralism and Global Citizenship.

The Programme also aims to contribute to the fostering of present and future generations of policy and decision makers through its executive training programme: the Academy of Global Governance, where theory and 'real world' experience meet and where leading academics, top-level officials, heads of international organisations and senior executives discuss on topical issues relating to global governance. For more information: http://globalgovernanceprogramme.eui.eu

Abstract

This paper examines whether restrictive data policies impact trade in services over the internet. We have collected comparable information on a variety of policy measures that regulate data for a wide group of countries for the years 2006-2016. This information is compiled in a weighted index that assesses the restrictiveness of these countries' data policies. We distinguish between policies regulating the cross-border movement of data and policies regulating the domestic use of data. Using econometric estimations, we show that strict data policies negatively and significantly impact imports of data-intense services. Therefore, countries applying restrictive data policies, in particular with respect to the cross-border flow of data, suffer from lower levels of services traded over the internet. This negative impact is stronger for countries with better developed digital networks. The results of our analysis are significant and hold for various robustness checks.

Keywords

1. Introduction[*]

Trade in services over the internet has grown steadily in the last two decades and now represents a share of more than 20 percent of total trade worldwide (Loungani *et al*. 2017). Technological developments, in particular information and telecommunication technologies (ICT), and open regulatory regimes regarding services have both contributed to the expansion of services trade. A third factor has also played an important role in expanding the scope of cross-border trade in services, namely the global nature of the internet, which is the focus of this paper.

The global internet relies on the flow of data worldwide. Many digital services are data-intense as they employ a high amount of electronic data in their production process that cross borders multiple times before the service is consumed (Figure 1). This facilitates the trade of services over the internet. The free movement of data across borders allows services producers to source and send data where its value is best used, reinforcing their comparative advantage in digital services. However, over the years, countries have introduced policies that restrict the cross-border flow and the domestic use of data (Ferracane *et al*., 2018b). These policies threaten the global nature of the internet and are expected to increase the costs of trading services online. This paper is the first to assess whether data policies adversely impact trade in services over the internet.

We define data policies as those regulatory measures that restrict the commercial use of electronic data. We limit our analysis to policy measures which are implemented at the national or supranational level (such as the EU). Although there is a great number of data policies implemented by local public entities, these are not the policies on which we focus on this paper. We identify two main categories of data policies. The first category covers those policies that impact the cross-border transfer of data; the second category covers policies that apply to the use of data domestically. The former category deals with all measures that raise the cost of conducting business across borders by either mandating companies to keep data within a certain border or by imposing additional requirements for data to be transferred abroad. The latter category refers to all measures that impose certain requirements for firms to access, store, process or more generally make any commercial use of data within a certain jurisdiction.

More specifically, this paper employs an empirical approach to assess whether data policies implemented in 64 countries between 2006 and 2016 have a significant impact on imports of services over the internet.[1] For this analysis, we develop a so-called data policy index that measures how restrictive countries are in regulating both the domestic use and the cross-border movement of data. Through an econometric specification, we relate this index with cross-border trade in services to study whether indeed restrictive data policies reduce the imports of services traded over the internet. To analyse this question, we use an advanced identification strategy to evaluate whether more data-intense

[1] In WTO speak, the cross-border trade of services over the internet is commonly referred to as Mode 1 trade in services, which is defined as "services supplied from the territory of one Member into the territory of any other Member" pursuant to Article I:2 of the GATS. It is worth mentioning that WTO members have so far not agreed upon a clear determination of whether the electronic cross-border delivery of a service is a service supplied through GATS mode 1 (cross-border) or mode 2 (consumption abroad).

services are relatively more hurt by higher levels of restrictive data policies following the methodology pioneered by Arnold *et al*. (2015; 2011).

The topic of data policies and services trade is new. Previous literature on data policies is limited to analysing whether data policies have a significant effect on productivity. For instance, Ferracane *et al*. (2018a) use firm-level data to assess whether data policies inhibit firm-level productivity performance in data-intense goods and services. The authors conclude that restrictive data policies negatively affect productivity in these (downstream) sectors. The authors also find that domestic regulatory policies on the use of data tend to have a marginally stronger impact on the productivity of firms compared to restrictions on the cross-border flow of data. One potential explanation is that policies restricting the free flow of data first and foremost impact countries' ability to trade (i.e. import) services, which in turn affects productivity given that data is an input.[2]

This paper focuses on whether data policies inhibit imports of services traded over the internet. To robustly assess whether different types of data policies affect trade in services differently, this paper splits up the data policy index into two sub-indexes. One sub-index merely covers restrictions affecting the cross-border (CB) flow of data, and one takes up all other domestic regulatory (DR) policies affecting the domestic use of data. To estimate our model, the identification strategy we develop relies on the intuition that more data-intense services are proportionately more affected by stricter data policies.

One additional reason for employing this identification strategy is that simple correlations show a negative link between data policies and trade in services, particularly for some technological-intense services. A substantial number of technological-intense services are also data-intense. Table 1 reports regressions as correlations for the two data policy's sub-indexes that we later use in our regressions. The table shows that sectors such as postal and courier services, IPR, information services, R&D and audiovisual services show negative coefficients. This negative result means that more restrictive data policies are associated with less trade in these sectors. This pattern motivates us to develop a cleaner methodology to precisely identify which sectors are more data-dependent, and to eventually come up with more robust and generic conclusions regarding the impact of data policies on trade in services.[3] To make sure this channel of reduced performance is not spurious, we provide alternative indicators of data-intensities and include different control variables for services regulation, and use different sources of trade in services data.

The remainder of this paper is organised as follows. The next section provides a short literature review discussing all relevant previous works on how data policies impact the economy. Section 3 provides the empirical strategy of our econometrical assessment and presents the various data sources for trade in services and data-intensities. This section also discusses how we construct the data policy index. Section 4 presents the results of our empirical model and robustness checks. Finally, the last section concludes and puts the empirical results in a wider policy context.

---

[2]   Restricting data flows limits countries' ability to import digital services against lower prices and greater quality of new services and varieties, which then affects productivity. Moreover, the goods trade literature has shown that restricting intermediate inputs constrains countries' potential to reach greater levels of productivity. See, for example, Amiti and Konings (2007) and Goldberg *et al*. (2010; 2009) in the case of goods.

[3]   Note that some coefficients show a positive result which may be due to omissions of control variables such as services regulations. However, due to insufficient observations when including these controls, regressions become impossible. This also provides us with a pragmatic reason to use our preferred identification strategy of data-intensities by sector. See further in the paper.

2. Previous Literature

The economic literature that discusses the link between data policies and trade in services is scarce. This is probably due to the fact that this topic is relatively new. Yet some elements of the triangle relationship between data flows, data policies and trade in services have been researched to some degree.

A first set of research papers investigates the economic role of data from a generic point of view. Manyika *et al*. (2016) for instance claim that the contribution of cross-border data flows to GDP has overtaken that of flows in goods in the current wave of globalisation. The study states that data flows today account for $2.8 trillion of the total increased world GDP over the last decade, thereby exerting a larger impact on growth than traditional goods trade. Interestingly, this work does not dedicate special attention to the inter-linkages that exist between data flows and trade in services but takes the former as being a separate channel that impacts the economy independent from services. Earlier work from Freund and Weinhold (2002) did, however, point to the facilitating role of the internet on trade in services. They state that an increase in internet penetration by 10 percent has the effect of increasing the growth of services trade by 1.1 percentage point for imports and 1.7 percentage point for exports. These conclusions are closely related to the question of whether data flows influence trade in services to the extent that restrictions on data can be seen as a restriction on the use of the internet.

A separate strand of the literature focuses on the mapping of the various policies related to data. A first attempt was performed by Stone *et al*. (2015) which covers measures of data localisation requirements only. Their study also notes that data flows enhance the efficiency of trade for specialised services firms both domestically and across borders. These services firms include data hosting, processing and mining. However, the study does not underscore explicitly the economic importance of the free flow of data (and therefore also data policies) on many other data-intense services such as information and business services. Furthermore, work by Ferracane (2017) further categorises the different forms of existing data policies that affect the cross-border movement of data and surveys all data policies applied across 64 major economies to show that data restrictions are applied in many countries, in different forms, and on different types of data. Finally, Ferracane *et al*. (2018b) develop the Digital Trade Restrictiveness Index (DTRI) which assesses the level of restrictiveness for different types of data policies. An expanded version of this index is used in the empirical part of this paper.[4]

A final strand of the relevant literature assesses more specifically whether restrictions on the cross-border movement and the use of data have a knock-on effect on productivity. Bauer *et al*. (2016) is the first work on this link and it focuses on sector-level productivity. This work investigates how some data restrictions affect productivity in data-intense sectors for a basket of emerging economies and the European Union (EU). Their index of data restrictions is composed by "augmenting" an existing index of product market restrictions (PMR), as measured by the OECD's PMR database, with additional regulatory restrictions on data. A more rigorous assessment of this empirical relationship is provided by Ferracane *et al*. (2018a). Using firm-level data across a set of developed countries and by constructing a full-fledged restrictiveness index on data policies, the authors confirm the conclusion that restrictive data policies significantly harm the productivity of firms active in data-intense sectors, with stronger evidence for restrictions on the domestic use of data.

To date, however, there is no in-depth empirical examination on how data policies affect services traded over the internet. This is surprising given the extent to which trade in services today relies on

---

[4] Another recently developed report from the USITC (2017) has described and scrutinised the many ways in which digital trade takes place and ends this examination with a list of policy measures relevant for data flows. Examples include data protection and privacy and data localisation rules. Other examples the USITC report includes are more indirectly related to data flows such as cybersecurity measures, censorship and intellectual property rights measures. These measures are not included in our empirical assessment but are nonetheless picked up and discussed in Ferracane *et al*. (2018b).

flows of data (see Bauer *et al*., 2016) and considering the sizable portion of all trade in services being traded over the internet.[5]

Recent work by Goldfarb and Trefler (2018) does, however, discuss the potential theoretical implications of data policies, such as data localisation and privacy regulations, on international trade and how these policies relate to the existing models of international trade. Although this discussion is put in a wider context of Artificial Intelligence (AI), the authors make clear that an expanded AI industry, in which data flows are an important factor, would have clear implications for services trade. Similarly, Goldfarb and Tucker (2012) point out that privacy regulations may harm innovative activities, particularly in services. They present the results of previous case studies they undertook with respect to two services sectors, namely health services and online advertising. In short, both studies show that there are strong linkages between the effective sourcing and deployment of data, the services economy and trade in services.

3. Empirical Strategy

This section sets out the empirical strategy. We develop a so-called data linkage index, which accounts for the fact that some services sectors are more dependent on data than others, i.e. are more data-intense. Intuitively, we expect more data-intense sectors are proportionately more affected by changes in data policies. To reflect this consideration in the empirical setting, we weight our data policy index with a measure of data-intensity for each services sector. In a second step, we present our baseline specification for the regressions in which we use the data linkage variable.

3.1 Data Linkage

The data linkage index builds on the methodology pioneered by Arnold *et al*. (2011; 2015) and is presented in Ferracane *et al*. (2018a). Their empirical approach has been used in many other papers with the purpose of creating a so-called services linkage index. In our case, we develop a data linkage index. For each country, we interact the country-specific data policy index with the data-intensities of each downstream services sector. This identification strategy relies on the assumption that sectors more reliant on data are those that are more affected by changes in data policies. This weighted approach is, in our view, a better method to measure the impact of data policies on trade in services than a simpler unweighted analysis of data policies and trade in services as done in Table 1 (see Introduction). The reason is that not all services deploy an equal amount of data over the internet and neither are all services as easily tradable over the internet.

Hence, the country-specific data policies index we develop is multiplied with a measure of data-intensity for country *c*, from a set of data producing sectors *d*, for each downstream services sector *j*. This is how the data linkage (DL) variable is set up, in which the data-intensity is expressed as (D/L). We develop two kinds of data-intensities as presented in equation (1) and (2). The equations of data-intensities show a different numerator, namely ς_{jd} and φ_{jd}, respectively. In equation (1), the term ς_{jd} denotes the software usage for each sector *j* for which data is retrieved from the US Census ICT survey. In equation (2), the term φ_{jd} stands for each downstream sector *j*'s input use of data services which we obtain from the US Bureau of Economic Analysis (BEA) input-output Use Table. We use this latter numerator for our robustness checks. In both cases, the data-intensities are stated as a ratio over labour,

---

[5]   Other channels of trade in services are according to the WTO's GATS Article 1:2: Mode 2, which covers services supplied in the territory of one country to the service consumer of any other country (also known as "consumption abroad"); Mode 3, which includes services supplied by a service supplier of one country, through commercial presence, in the territory of any other country (also known as "commercial presence"); and finally Mode 4, services supplied by a service supplier of one country, through the presence of natural persons of a country in the territory of any other country (also known as "presence of natural persons").

called LAB_j, that is employed in each downstream sector j. The data for labour is retrieved from the US Bureau of Labour Statistics (BLS). As a result, we apply the following formula respectively:

$$Data\ Linkage\ (\text{DL})_{cjt} = \ln\frac{\sum_d \varsigma_{jd}}{LAB_j} * data\ policy\ index_{ct} \qquad (1)$$

$$Data\ Linkage\ (\text{DL})_{cjt} = \ln\frac{\sum_d \varphi_{jd}}{LAB_j} * data\ policy\ index_{ct} \qquad (2)$$

Note that in equation (1) and (2) we put the intensity indicators in logs, in line with previous literature on factor intensities. This expression of intensities is therefore closer to the literature of comparative advantage such as Chor (2011), Nunn (2007) and Romalis (2004). In equation (2), however, the data-intensity using φ_{jd} is more in line with the aforementioned literature such as Arnold *et al.* (2011; 2015) but also Bourlès *et al.* (2013) that create an indicator of dependency using input-output matrixes, although we use labour as a denominator instead of total input usage as most of these papers do. The economic literature also calls these latter types of input shares backward linkages.[6] Finally, in both equation (1) and (2), the data policy index refers to a country-year specific data policy index (see Section 3.3).

We choose our two measures of data-intensity to vary at industry level and specific to one year only, namely 2010, to avoid endogeneity issues. This may occur in the event high data-intense services sectors with high trade volumes push for lower regulatory restrictions on data over time. Also, instead of country-sector specific intensities (i.e. ς_{cjd} or φ_{cjd}), we use common sector-specific data-intensities for one country, namely the US, which makes our data-input coefficients more exogenous. However, to alleviate the concern that this year may be lying in the middle of our panel period, we use the year 2007 to measure the data-intensities in equation (2). Yet one should be aware of the fact that ICT and internet technologies have had a tremendous impact on expanding the scope of services trade in more recent years with the help of free data flows.[7]

3.2 Data Intensities

The degree to which sectors are intense in the use of data is measured in two ways. For our preferred measure of data-intensity as defined in equation (1), we use information of data-usage from the 2011 US Census ICT Survey. This data is survey-based and records, at detailed 4-digit NAICS sector-level, how much each industry and services sector spend on inputs from the ICT-sector in terms of ICT equipment and computer software in Mn USD.

For our regressions, we take computer software expenditure as part of our preferred data-intensity. The US Census ICT Survey actually records two separate variables on software expenditure, namely capitalised and non-capitalised expenditure. We first select non-capitalised expenditure because this proxy is closer to our second measure of data-intensity for which we use input-output data. However, the regression results also hold when using capitalised software expenditures. Non-capitalised computer software expenditure is comprised of purchases and payroll for developing software and software licensing and service/maintenance agreements for software. Although this proxy does not entirely capture the extent to which sectors are using electronic data, it nonetheless is the closest kind of data-use variable we can publicly find. We take the year 2010 for our regressions, and as said we divide this software expenditure over labour, also for the year 2010, and use it in our baseline regression.

---

[6]   Moreover, one additional reason to look at the input-side of data and data-related services is that the recent economic literature connects the potential growth and productivity performance of countries notably to the input usage of data and digital services in the wider economy. See Jorgenson *et al.* (2011).

[7]   Note that 2007 is the most recent year the BEA report input-output tables for the US at a detailed level. In fact, these matrixes at 6-digit NAICS level are the most detailed in the world, which allows us to precisely determine which are the data-input sectors from where each services sector sources data from.

Figure 2 provides an overview of the data-intensities for each services sector calculated for capitalised and non-capitalised expenditures. Both intensities are computed at 4-digit NAICS but then concorded into the 2-digit BPM6 level. The reason for re-classifying the input-coefficients is that our services trade data is provided in BPM6. Since no concordance table exists between NAICS and BPM6, we have developed our own matrix and aggregated these data input coefficients at the 2-digit BPM6 level for both intensities by taking the simple average. Note that one category forms a mismatch between the two classification tables, which is Intellectual property / Royalties and license fees. This sector is not reported in the US Census nor in the BLS database. Nonetheless, it is important to include this sector as it covers, among other items, patents, trademarks and copyrights – all of which are data-intense items and for which also high amounts of services trade is recorded. Therefore, we have developed our own concordance method to include this sector. Details of this concordance can be found in Annex 1.[8]

Figure 2 also shows the ranking of sectors by data-intensity using our proxy for both capitalised and non-capitalised software expenditures. Not surprisingly, the telecom and computer service sectors are the most data-intense. They employ a high amount of data through their greater use of software. Information services such as data processing services and web search are also highly data-intense, in line with our expectations. An interesting finding is that both financial and insurance services also come out as very data-intense sectors. The two sectors are more broadly considered as very technological-intensive and internet technologies have massively increased in the financial services industry.[9] On the other side of the spectrum, the least data-intense sectors are construction and travel services. The middle-range of data-intense sectors is a mix of modern and traditional sectors such as R&D and transport services.

A second way to compute data-intensities is to use data from the US BEA Input Use Table, which is provided at the 6-digit level. This alternative method is defined in equation (2). For these intensities, we must first determine the sectors that provide data to other downstream (services) sectors before we are able to compute φ_{id}. Table 2 lists the sectors that we define as "data producers". These are sectors that generate a high amount of data when providing their services. As such, we determine that these sectors act as an input of data into all other downstream sectors in the economy. Hence, we compute input values based on these eight sectors. This list of sectors goes beyond what is covered by purely software expenditure. It follows Bauer *et al*. (2016) and is in line with Jorgenson *et al*. (2011) regarding their IT-producing industries. They include, *inter alia*, the telecommunication sector; data processing, hosting and related services; internet service providers and web search portals; software publishers; computer system design services; and other computer-related services.

We calculate the ratio between the BEA's input data services usage based on the purchaser's prices and labour for each downstream services sector at the 6-digit level. Labour is again sourced from the US BLS in NAICS for the same year and is matched with the US BEA input-output matrix which fits neatly. For this data-intensity, we take the year 2007 as the BEA latest available input-output Table is from this year. All intensities are re-concorded into BPM6 using our self-developed concordance table.[10]

---

[8] The concordance table between 4-digit NAICS and 2-digit BPM6 can be obtained upon request.

[9] Another non-ICT sector that is shown to be very data-intense is the retail sector. However, neither the US Census nor the BPM6 classification shows a separate entry for retail or wholesale distribution services, which is the reason why this sector is omitted in our analysis of intensities and is not covered in our regression analysis.

[10] An additional convenient motivation for using US tables is that the US is often used as a benchmark country in similar cross-country studies using sector intensities, which makes our input coefficients on data usage exogenous. However, there is a debate in the economic literature about whether one should use the assumption of equal industry (or sector) technologies across countries or not. Equal technology coefficients seem reasonable if one assumes that the countries selected in the sample are reasonably similar in their economic structures and technology endowments. On the other hand, our fixed effects in the econometrical specification should take care of these technology differences. Moreover, in our case, we don't have detailed country-specific IO tables for the 64 countries covered in our sample at such a disaggregated level, which we need. Practically, using US input-output shares only might as well form a convenient assumption if a suspicion exists that input-

As shown in Figure 3, telecom services, intellectual property, computer and financial services are again the most data-intense services sectors. However, two notable changes are visible compared to Figure 2. One is the remarkable increase in data-intensity for technical, trade and other business services. Whereas our preferred data-intensity indicator shows low data-intensity for this sector, using BEA data this sector is now the third-largest user of data-services. A second change is that insurance services have considerably dropped in the ranking and are now one of the least data-intense sectors using BEA data.

3.3 Data Policy Index

The second term of our data linkage variable is the data policy index, which is based on a quantifiable set of policy information on countries' regulatory framework on data. We draw on a comprehensive new database of data policies recently released by the European Centre for International Political Economy (ECIPE).[11] The policies used for the analysis are those considered to create a cost for firms relying on data for their businesses. The criteria for listing a certain policy measure in the database are the following: (i) it creates a more restrictive regime for online versus offline users of data; (ii) it implies a different treatment between domestic and foreign users of data; and (iii) it is applied in a manner considered disproportionately burdensome to achieve a certain policy objective.

Starting from the database, these policies are aggregated into an index using a detailed weighting scheme adapted from Ferracane *et al*. (2018b) and presented in detail in Annex 2.[12] We expand the index released by Ferracane *et al.* (2018b), which covered only the years 2016/2017, to create a panel for the years 2006-2017 that we can use in our regressions. In addition, the database and index have been updated with new regulatory measures found in certain countries.

To build up the index, each policy measure identified in any of the categories receives a score that varies between 0 (completely open) and 1 (virtually closed) according to how vast its scope is. A higher score represents a higher level of restrictiveness in data policies. While certain data policies can be legitimate and necessary to protect non-economic objectives such as the privacy of the individual or to ensure national security, these policies nevertheless create substantial costs for businesses and are therefore listed in the database.

After applying our weighting scheme, the data policy index also varies between 0 (completely open) and 1 (virtually closed). The higher the index, the stricter the data policies implemented in the countries. Moreover, the index is broken down into two sub-indexes that cover two main types of policy measures that we analyse in this paper: one sub-index that covers policies on the cross-border movement of data and another sub-index that covers policies on the domestic use of data. Analysing these two sub-indexes separately provides additional information on whether the impact of data policies on services trade varies according to the nature of the policies. The full data policy index is measured as the sum of these two sub-indexes. The list of measures included in the two sub-indexes is summarised in Table 3 and the specific weight for each measure is given in the last column.

---

output tables at the country level are not always very well measured for some economies. This could be the case for less developed countries which often suffer from weak reporting capacities. Our country selection includes a substantial number of less developed countries where this could be the case.

[11] The authors have contributed to the development of the database at ECIPE. The dataset comprises 64 economies and is publicly available on the website of the ECIPE at the link: www.ecipe.org/dte/database. Besides analysing the 28 EU member states and the EU economy as a single entity, this database also covers Argentina, Australia, Brunei, Canada, Chile, China, Colombia, Costa Rica, Ecuador, Hong Kong, Iceland, India, Indonesia, Israel, Japan, Korea, Malaysia, Mexico, New Zealand, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Russia, Singapore, South Africa, Switzerland, Taiwan, Thailand, Turkey, United States and Vietnam.

[12] The authors have previously used this categorisation in Ferracane *et al*. (2018a).

As shown in Table 3, the sub-indexes are measured as a weighted average of the different types of policy measures. The weights are intended to reflect the level of restrictiveness of the types of measures in terms of costs for digital trade. The first sub-index on cross-border data flows covers three types of measures, namely (i) a ban to transfer data or a local processing requirement for data; (ii) a local storage requirement, and (iii) a conditional flow regime. The second sub-index covers a series of subcategories of policies affecting the domestic use of data. These are: (i) data retention requirements, (ii) subject rights on data privacy, (iii) administrative requirements on data privacy, (iv) sanctions for non-compliance, and finally, (v) other restrictive practices related to data policies.

Figure 4 shows how the two sub-indexes and the overall data policy index have evolved over time between the years 2006 and 2016. Each line is computed as the average of the 64 countries covered in this study. As one can see, there is a clear upward trend reflecting the fact that all types of data policies are becoming stricter over time. Note that measures affecting the cross-border data flows can directly inhibit the free flow of data across countries and therefore can directly restrict trade in services. On the other hand, measures belonging to the second sub-index on domestic use of data only indirectly affect the flow of data across borders and therefore are expected to create costs for trade only indirectly.

3.4 Baseline Regression

Equations (1) and (2) are used in our baseline regression which is specified in equation (3) below. Equation (3) measures the causal impact of the data linkage index in the previous year on the log of cross-border imports of services (SM). We regress the logarithm of services imports in country *c,* in services sector *j,* in time *t,* and on the data linkage index with a one-year lag. We use a lagged variable because, on a wider scale, it takes time before downstream firms across all countries feel the regulatory consequences of a change in the data policies. In addition, applying the lag takes out further endogeneity concerns to the extent that reverse causality becomes less obvious. The baseline specification takes the following form:

$$ln(\text{SM})_{cjt} = \Phi + \theta \text{DL}_{cjt-1} + \delta_{ct} + \gamma_{jt} + \varepsilon_{cjt} \qquad (3)$$

In equation (3) the terms δ_{ct} and γ_{jt} refer to the fixed effects by country-year and sector-year, respectively. Sector fixed effects are applied at the 2-digit BPM6 level, which includes 18 sectors in total in line with Figure 2 and 3. Finally, ε_{icjt} is the residual term. Regressions are estimated with robust standard error clustered by country-sector-year and are performed over the period 2006-2016 with a one-year lag of the data linkage variable throughout.

For our dependent variable $ln(\text{SM})_{cjt}$ we use two different data sources that have recently been developed. One is the WTO-UNCTAD-ITC annual trade in services dataset, which covers exports and imports of total commercial services. This is our preferred source which we also use for our descriptive analysis below. This dataset covers 222 entities which include countries and regional aggregations/economic groupings from 2005-2016 at the 2-digit level. The data is in line with the sixth edition of the IMF Balance of Payments and International Investment Position Manual (BPM6) as well as the 2010 edition of the Manual on Statistics of International Trade in Services (MSITS 2010). This entails that, compared to the BPM5 classification, major changes for the Balance of Payments (BOP) classification for services have been introduced with regards to financial intermediation services, insurance services, intellectual property and manufacturing and maintenance services. The second dataset used in the analysis is the BaTIS dataset from the OECD-WTO. This data will be discussed as part of our robustness checks in Section 4.1.

3.5 Descriptive Statistics

Figure 5 shows that, depending on the country under consideration, services imports are positively or negatively correlated with a sector's data-intensity. The two top panels of Figure 5 show that, in the case of Ireland and the Netherlands, greater levels of per capita imports of services are positively associated with data-intensities. That is, the two countries show higher levels of per capita imports in services sectors that exhibit higher levels of data intensity. Conversely, Turkey and China, which are displayed in the bottom panel, show a negative association between per capita service imports and data-intensity. This means that both countries show lower levels of unit services imports in data-intense sectors. Various reasons may lay ground to this observation, including the fact the services sector plays a different role in the countries' respective economies. For example, in a country like Ireland, services have a greater relative economic importance than in Turkey.

However, data policies are also likely to play an important role. An illustration of how data policies are an important factor for countries to successfully trade in data-intense services is given in Figure 6. The figure sets out graphically the relationship that will be tested more formally in our next section. In Figure 6, the vertical axis displays the log of data-intense services imports in GDP for each of the 64 countries analysed. Data-intense services are here defined as the Top 5 most data-intense sectors following our (D/L) based on capitalised software expenditure as shown in Figure 2 (bottom panel). They are telecommunications, information services, intellectual property, computer services, and professional and financial services. The horizontal axis shows the data policy index. The figure shows a negative relationship, which means that countries with stricter (i.e. more restrictive) data policies also have relatively lower levels of imports in data-intense services.

3.6 Baseline Extension

In the next step, we expand the baseline specification to consider any differential impact of data policies on trade in services for countries that are technologically well-equipped to deal with data flows. The prime reason for doing so is that digital sectors have expanded rapidly in countries with a good digital-enabling environment. For instance, countries with qualitatively good telecom networks or better technology absorption by firms have stronger digital networks and are therefore likely to show greater data-related activities. In turn, therefore, countries with a more digital-enabling environment would be more negatively affected by more restrictive data policies. Moreover, Figure 6 also shows that some developing countries, such as Brazil and Thailand, have relatively low imports (in GDP) in data-intense services despite having low scores on the data policy index. The data linkage index in combination with the quality of the digital network environment may, therefore, explain this pattern.

Normally, fixed effects in Equation (3) control for various policy influences and unobserved shocks such as the digital enabling environment that varies at the level of country-year. However, with an interaction variable comprising the data linkage index that varies at industry-level and a country-level indicator that proxies for a country's digital environment over time, we can assess the specific differential effect because this interaction term varies by country-sector-year. A well-suited proxy that captures the development of the countries' digital enabling environment is the WEF's Network Readiness Indicator (NRI). This indicator measures the capacity of countries to leverage ICTs for increased competitiveness and well-being (WEF, 2016). This country-time-specific index summarises several sub-indicators measuring the extent to which individuals are using the internet, international bandwidth in kb/s per user, a country's availability of the latest technologies and the level of technology absorption by firms.

We include the index in our extended baseline regressions. We interact our data linkage, DL_{cjt-1}, with the NRI variable varying by country and time, which is demeaned. The extended baseline specification therefore becomes:

$$ln(\text{SM})_{cjt} = \Phi + \theta \text{DL}_{cjt-1} * \text{NRI}_{ct} + \delta_{ct} + \varsigma_{jt} + \varepsilon_{cjt} \qquad (4)$$

Equation (4) applies a similar fixed effects structure as equation (3) and has the same accompanying error term ε_{cjt}. Of note, our data linkage index increases with more restrictive data policies whilst higher levels of the NRI indicator represent more developed digital networks in a country. This implies that a negative sign on the interaction variable's coefficient outcome means that countries with more developed digital networks experience a stronger reduction in services imports with stricter data policies.[13] A positive sign would mean that stricter data policies have a lower impact on trade in services in countries that are better digitally equipped.

4. Results

The results of our baseline regressions are shown in Table 4. The regressions are performed with our preferred data-intensity from equation (1) in which software use per sector is computed over labour. Column 1 sets out the results for the full data policy index. Columns 2 and 3 report the results for the sub-indexes of data policy related to the cross-border data flows and data policies related to the domestic use of data, respectively. Column 1 shows that the full data policy index comes out highly significant with the expected negative coefficient sign. This indicates that overall more restrictive data policies negatively impact imports of data-intense services across countries over time. Columns 2 and 3 show that the sub-index for policies on cross-border data flows has a high significance while the sub-index covering the domestic data use component is only weakly significant. This means that the significance of the full data index is in mainly explained by the sub-index capturing policies on the cross-border data flow. When entering both sub-indexes together in column 4, the cross-border policies remain robustly significant while the domestic use component is not found significant.

Columns 1-4 report the data linkage index based on the preferred data-intensity variable using non-capitalised expenditures. As previously explained, this type of expenditure can be separated into expenditures on purchases and payroll for developing software and expenditure on software licensing and service/maintenance agreements. Column 5 and 6 show the coefficient results when using these two types of information respectively in our data linkage variable, as indicated by NC Pur and NC Lic in Table 4. In both cases again the coefficient of the sub-index of cross-border data flows is highly significant. The significance of the data linkage sub-index remains strong also when using the capitalised expenditure in the last column (indicated by C in Table 4). In terms of the coefficient size, purchases and payroll for software show the highest results. In all cases, the sub-index covering the domestic use of data shows low or no significance and therefore does not have any statistical importance although it shows a negative coefficient sign.

The results of the extended baseline regression following equation (4) are reported in Table 5. All data-linkage indexes, including their interactions, are regressed using non-capitalised software expenditure. The pattern of significance follows the one presented in Table 4. The full index and the cross-border sub-index show high significance, while the domestic regulatory sub-index shows low significance or no significance when regresses together with the former sub-index. In columns 1-2, the interaction variable remains insignificant, which means that for the full index and the cross-border sub-index no differential impact of stricter data policies on services trade can be found for countries with qualitatively better digital networks. However, column 4 reveals that the negative coefficient sign on the cross-border interaction term is now highly significant. This indicates that, when controlling for policy related to the domestic usage of data, the negative impact of stricter data policies on services

---

[13] Since our NRI variable is demeaned, the statistical interpretation is that a negative sign on the coefficient result of this interaction variable means that countries with stronger digital networks *compared to countries with an average NRI score* experience a stronger reduction in services imports when having stricter levels of policy frameworks for data. By centring our variable first, effects can therefore be made more interpretable.

imports is stronger in countries with a better digital-enabling environment. The interaction term with the alternative policy index of policies related to the domestic usage of data provides a surprisingly positive and significant sign.

4.1 Robustness checks

In this section, we provide several robustness checks for our analysis. These robustness checks are in particular meant to address concerns regarding: (i) the omission of other regulatory variables that restrict services trade, (ii) the use of an input-reliance coefficient for an intermittent year, and (iii) the fact that we only use one dataset for trade in services.

4.1.1 Services Restrictions

This robustness check mostly tackles the fact that many services are heavily regulated. This fact may cause concerns that if in the regressions this information is omitted, the results would fail to include a channel of services regulations that may be the prime channel to explain services trade. For this reason, we add the services trade restrictiveness index (STRI) as a control variable, which captures how much each services sector is restricted for each country in our dataset.

Both the OECD and the World Bank have created a version of the STRI. One constraint for us is that both indexes do not wholly cover the period in our analysis. We prefer to use the OECD's STRI for two reasons. First of all, it displays data for the years as of 2014 and therefore covers three years in our regressions. In addition, another advantage of the OECD's STRI is that it covers more sectors and has pre-defined groupings of the index according to different types of policy restrictions, such as policy measures related to the four modes of supply separately, or those related to market access and national treatment only (as opposed to measures related to domestic regulations only), or those which are discriminatory (as opposed to measures which are non-discriminatory – yet still affect the foreign service provider). We concord all STRI sectors into our BMP6 classification and use an average of the STRI there where multiple sectoral indexes are available. Table A3.1 in Annex 3 provides our self-constructed concordance table.

The results of the regressions including the STRI variable are reported in Table 6. Columns 1-4 report the STRI grouping that includes restrictions related to cross-border service supply (i.e. Mode 1). The reason for doing so is that this mode of supply covers trade in services that is performed on the internet, which is in line with our trade in services data as it mostly captures trade through this mode. Subsequent columns in Table 5 also report coefficient results when using alternative STRI groupings. Columns 1-4 show that, in all cases, the data linkage index is highly significant at the 1 percent level, whereas the STRI variable shows a negative yet insignificant coefficient. Surprisingly, the data linkage index containing policies for the domestic use of data is now also significant in column 3, even though the number of observations drops substantially. However, in column 4, the significance of this data linkage index drops dramatically when entered together with the data linkage for cross-border data policies.

The services literature points out that complementarities exist between modes of supply, particularly regarding computer services and information services imports (Kirkegaard, 2008).[14] If this were the

---

[14] In total there are four modes of supply of which (1) Mode 1 represents cross-border – services supplied from the territory of one Member into the territory of another, e.g. software services through e-mail to another country; (2) Mode 2 represents consumption abroad – services supplied in the territory of one Member to the consumers of another, e.g. education services in another country; (3) Mode 3 represents commercial presence – services supplied through type of business or professional establishment of one Member in the territory of another; and finally (4) Mode 4: Presence of natural persons – services supplied by nationals of one Member in the territory of another, e.g. doctors moving to foreign country to provide temporary their service.

case, then the inclusion of our Mode 1 STRI grouping might be too stringent. Therefore, we re-perform our regressions with various other groupings of the OECD's STRI. Column 5 first includes the widest form of the STRI that spans all four modes of supply. The results show that the full STRI is statistically significant. This suggests that barriers in Mode 1 for the various services sector for which we have sector-specific STRI do indeed not fully capture all restrictions affecting cross-border services trade online. Importantly, our cross-border data linkage index stays robustly significant while the domestic use data linkage index loses statistical importance. We also include STRI groupings in which we only take barriers affecting market access and national treatment (column 6) and discriminatory barriers (column 7). Both cases show a similar statistical significance as in the case of the full STRI.

4.1.2 Alternative Data-intensities

Further concerns may arise by the fact that, for our data-intensity measure, we rely on data from an intermittent year in our panel period, namely 2010. Using an earlier year could further exclude any potential endogeneity concern that would emerge because of political economy responses, that is that lower regulatory restrictions on data are the result of lobby efforts by data-intense sectors showing high services trade.

The US Census only provides data on software expenditure for earlier years at a very aggregate level, which is of little use for our analysis.[15] Therefore, we use the alternative data-intensities as developed in equation (2), which employ US input-output data from the BEA IO Use Tables. As explained, we consider the eight sectors in Table 2 to be data-producing sectors which provide data as inputs for every other downstream sector at a detailed 6-digit level. Data-intensities are then re-concorded and aggregated by taking the simple mean. As stated in the previous section, one advantage of using this data is that the latest input-output Use table is from the year 2007.

Figure 7 provides a first impression on how the two data-intensities between the US Census and BEA correlate. Overall, travel and construction services are the least data-intense whereas telecom, intellectual property, finance and computer are the most. The figure also shows that there are some small deviations between the two proxies. Telecom seems more data-intense according to the figures from the BEA whilst insurance, postal and courier, and air transport show greater data intensities (in the form of software) when using US Census data. Yet the overall correlation coefficient between these two data-intensities is 0.74, which substantiates their strong interrelationship. The regression results using this second data-intensity are shown in Table 7 and confirm the main message from our previous regression analysis – although some differences do appear.

First, compared to Table 6, the significance of the full data policy index in our linkage variable diminishes. A second difference in result is that also for the two sub-indexes of polices related to the cross-border flow and the domestic use of data come out weakly significant, as shown in columns 2-3 respectively. However, column 4 shows that, when entering the two sub-indexes together, the significance of both variables disappears. All these regressions are performed using the STRI control variable for Mode 1. Column 5 reports the results when excluding the STRI variable and provides a negative and significant coefficient for the cross-border flow sub-index. If we use alternative groupings for the STRI, such as the full STRI (column 6) or the one covering market access and national treatment measures only (column 7), the cross-border data policy linkage variable remains significant, whilst the domestic regulatory sub-index instead loses any significance. A similar outcome appears when using the grouping of discriminatory barriers alone (output omitted).

---

[15] Note as well that the US Census ICT Survey only provides ICT expenditure data at 4-digit NAICS for the years 2010, 2011 and 2013 and so earlier years before 2010 are not possible to take into our research at detailed level.

4.1.3 Alternative Trade in Services Data and STRI

The results so far have shown regression outcomes using the WTO-OECD-ITC dataset of trade in services. An alternative dataset is the OECD-WTO BaTIS database, which is laid out in Fortanier *et al*. (2017). This database provides a complete and consistent balanced dataset of services trade that originally served as input for the compilation of the Trade in Value-added (TiVA) database. The data covers the years 1995-2012 and includes 191 countries and 11 main EBOPS 2002 services sectors. Extensive efforts have been put into collecting the data from all available official sources, cleaning it and completing it using different methodologies to estimate missing information, including with the use of derivations, backcasting techniques, interpolation, and predictions derived from regression models. In this database, three different trade values are shown, namely the sheer reported values from sources, reported values including estimates, and the final balanced value. We present the results with the latter value, but we have also performed checks showing that results are consistent.

Using this data requires re-compiling our data-intensity variables. This is done using another self-constructed concordance table in which again the royalties and license sector follows the approach explained in Annex 1. Note that now we concord the sectors following a slightly different system compared to our preferred dataset since BaTIS is based on EBOPS 2002 whereas the WTO-UNCTAD-ITC database is based on BPM6. Nonetheless, the four most data-intense sectors using the BaTIS classification are in line with the rankings of data-intensities based on the BPM6 classification as presented in Figure A4.1 in Annex 4.

Apart from using different trade data, we also use the alternative STRI variable from the World Bank. There are two main reasons for doing so. One is that the OECD's STRI only starts in 2014, which is after the period covered by the BaTIS database, which goes only up to 2012. Second is that the sector classification of BaTIS provides fewer services sectors, which nicely coincides with the more aggregate sectors that the World Bank's STRI covers. Sectors included by the World Bank's STRI are finance, insurance, legal services, accounting and auditing services, as well as air, rail, road and maritime transport, and finally telecommunication.[16] One big disadvantage of using this index, however, is that it only provides information for one year, namely 2008/2009. Yet this year nicely overlaps with the early years of our data policy index as well as with the final years of the BaTIS trade data. We nonetheless do apply a one-year lag since our cross-section may not pick up any reform efforts in the same year.

The results of these robustness checks are shown in Table 8. In all entries, none of the coefficient results are significant, though the data linkage indexes do mostly show a negative sign. Columns 1-4 show the regression results using the World Bank's STRI for Mode 1, in line with Table 6 and 7. The full STRI is also used as a control variable of which the results are presented in column 5, which only shows weak significance. This may be due to the fact that, as a one-year observation, the restrictiveness of the index for 2008/2009 does not have any impact in 2010 as many countries may have reformed their services markets before. To correct for this, we use the same years in which the trade data is recorded, which is actually year 2009 and which therefore means that results should now be interpreted as correlations. Columns 6 and 7 show that for the two cases, only when entering the full range of services restrictions, a weak significance of the STRI variable is found. The lack of any strong significance on any services regulatory variable as well as data policy index may be due to the cross-section nature of these regressions.

---

[16]   A final services sector included in the World Bank's STRI is the retail sector. However, since the EBOPS 2002 manual provides explanation of the difficulties of classifying trade in this sector (and their associated challenges of actually measuring trade in retail), we have omitted this sector in our analysis.

5. Conclusion

This paper aims to make three contributions to our understanding of the impact of data policies on services traded over the internet in the context of a world economy that is relying increasingly on data. First, we develop a novel time-varying country-specific index that measures the extent to which different countries restrict the cross-border movement and the domestic use of electronic data. Using this index, our second contribution is to estimate the effects of strict data policies on imports of services traded online. We find that more restrictive data policies, in particular with respect to the cross-border movement of data, result in lower imports in data-intense services for countries imposing them. While it is possible that countries with greater levels of trade in data-intense services intentionally avoid restricting the cross-border flow of data, we provide various robustness checks to tackle this endogeneity problem. Finally, we show that strict cross-border data policies have a stronger impact on services imports in those countries with stronger digital networks.

Our results are in line with an expanding literature that shows that restrictive regulatory trade barriers in services have a negative and significant impact on trade in services. However, our work differs from this previous research as most of the antecedent work looks at sector-specific regulatory barriers in services whereas our findings apply to economy-wide barriers that are specifically targeted at the internet. We show that, in addition to the more commonly known trade barriers that affect the services sectors and services trade, stricter policies on cross-border data flows also restrict trade in services online. As most services are highly data-intense, and as many countries are moving into an increasingly digital direction, these policies restricting data flows across borders are likely to impede countries to reap the efficiency gains stemming from services imports. In addition, exports of data-intense services would, in turn, decrease towards countries that impose strict data policies.

One important question for future research relates to restrictive data policies affect developing countries and their opportunities for growth. First, following our results, the imposition of data policies by developed countries could deter significantly the opportunities of developing countries to export data-intense services to developed economies, something alluded to in Mattoo and Meltzer (2018). India is a clear case in point as the country exports a lot of data-intense services to developed markets such as the EU. Second, it remains to be seen whether the decision by developing countries to impose strict data policies will benefit their economies in the long-run. An argument often put forward is that developing countries require policy space to establish expertise and specialisation in industries in which they subsequently can export or participate in value chains. Yet, data and data-intense services are important inputs for downstream sectors, also for manufacturing. When efficiently supplied, they help countries develop. Strict data policies may therefore reduce the opportunity for poorer countries to reap efficiency gains and ultimately grow.

References

Amiti, M. and J. Konings (2007) "Trade Liberalization, Intermediate Inputs and Productivity." *American Economic Review*, Vol. 97, No. 5, 1611-1638.

Arnold, J., B. Javorcik and A. Mattoo (2011) "The Productivity Effects of Services Liberalization: Evidence from the Czech Republic", *Journal of International Economics*, Vol. 85, No. 1, pages 136-146.

Arnold, J., B. Javorcik, M. Lipscomb and A. Mattoo (2015) "Services Reform and Manufacturing Performance: Evidence from India", *The Economic Journal*, Vol. 126, Issue 590, pages 1-39.

Bauer, M., H. Lee-Makiyama, E. van der Marel and B. Verschelde (2016) "A Methodology to Estimate the Costs of Data Regulation", *International Economics*, Vol. 146, Issue 2, pages 12-39.

Bourlès, R., G. Cette, J. Lopez, J. Mairesse and N. Nicoletti (2013) "Do Product Market Regulations in Upstream Sectors Curb Productivity Growth? Panel Data Evidence for OECD Countries", *The Review of Economics and Statistics*, Vol. 95, No. 5, pages 1750-1768.

Chor, D. (2011) "Unpacking Sources of Comparative Advantage: A Quantitative Approach", *Journal of International Economics*, Vol. 82. No. 2, pages 152-167.

Ferracane, M.F., J. Kren and E. van der Marel (2018a) "Do Data Policy Restrictions Impact the Productivity Performance of Firms?", ECIPE Working Paper No. 2018/1, Brussels, ECIPE.

Ferracane, M.F., H. Lee-Makiyama and E. van der Marel (2018b) "Digital Trade Restrictiveness Index", European Center for International Political Economy, Brussels: ECIPE.

Ferracane, M. F. (2017), "Restrictions on Cross-Border data flows: a taxonomy", ECIPE Working Paper No. 2017/1, Brussels, ECIPE.

Fortanier, F., A. Liberatore, A. Maurer, G. Pilgrim and L. Thomson (2017) "The OECD-WTO Balanced Trade in Services Database", OECD-WTO Working Paper, November.

Freund, C. and D. Weinhold (2002) "The Internet and International Trade in Services", *American Economic Review,* Vol. 92, No. 2, pages 236-40.

Goldberg, P., A. Khandelwal, N. Pavcnik and P.Topalova (2010) "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India", *Quarterly Journal of Economics*, Vol. 125, Issue, 4, pages 1727-67.

Goldberg, P., A. Khandelwal, N. Pavcnik and P.Topalova (2009) "Trade Liberalization and New Imported Inputs", *American Economic Review*, Vol. 99, No. 2, pages 494-50.

Goldfarb, A. and D. Trefler (2018) "AI and International Trade", NBER Working Paper Series No. 24254, National Bureau of Economic Research, Cambridge MA: NBER.

Goldfarb, A. and C. Tucker (2012) "Privacy and Innovation," in Innovation Policy and the Economy (eds.) Josh Lerner and Scott Stern, / University of Chicago Press, pages 65–89. See also NBER Working Paper Series No. 17124, National Bureau of Economic Research, Cambridge MA: NBER.

Jorgenson, D.W., M.S. Ho and J.D. Samuels (2011) "Information Technology and US Productivity Growth: Evidence from a Prototype Industry Production Account", *Journal of Productivity Analysis*, Vol. 2, Issue 36, pages 159-175.

Kirkegaard, J.F. (2008) "Distance Isn't Quite Dead: Recent Trade Patterns and Modes of Supply in Computer and Information Services in the United States and NAFTA Partners" PIIE Working Papers WP 08-10, Washington DC: Peterson Institute for International Economics.

Loungani, P., S. Mishra, C. Papageorgiou and K. Wang (2017) "World Trade in Services: Evidence from A New Dataset", IMF Working Paper Research Department, Washington DC: IMF.

Manyika, J., S. Lund, J. Bughin, J. Woetzel, K. and D. Dhingra (2016) "Digital Globalization: The New Era of Global Flows", McKinsey Global Institute, Washington DC: McKinsey and Company.

Mattoo, A. and J.P. Meltzer (2018) "International Data Flows and Privacy: The Conflict and Its Resolution", Policy Research Working Paper No. 8431, Washington, D.C.: World Bank Group.

MIT (2015) "A Business Report on Big Data Gets Personal", MIT Technology Review, Cambridge MA: MIT.

Nunn, N. (2007) "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade", *The Quarterly Journal of Economics*, Vol. 122, Issue 2, pages 569-600.

Romalis, J. (2004) "Factor Proportions and the Structure of Commodity Trade", American Economic Review, Vol. 91, No. 1, pages 67-97.

Stone, S., J. Messent and D. Flaig (2015) "Emerging Policy Issues. Localisation Barriers to Trade", OECD Trade Policy Papers No. 180. OECD Publishing, Paris.

USITC (2017) "Global Digital Trade 1: Market Opportunities and Key Foreign Trade Restrictions", Publication Number: 4716, United States International Trade Commission, USITC: Washington DC.

WEF (2015) "Global Information Technology Report 2015: ICTs for Inclusive Growth", Davos: World Economic Forum

Tables and Figures

Figure 1: Global data traffic
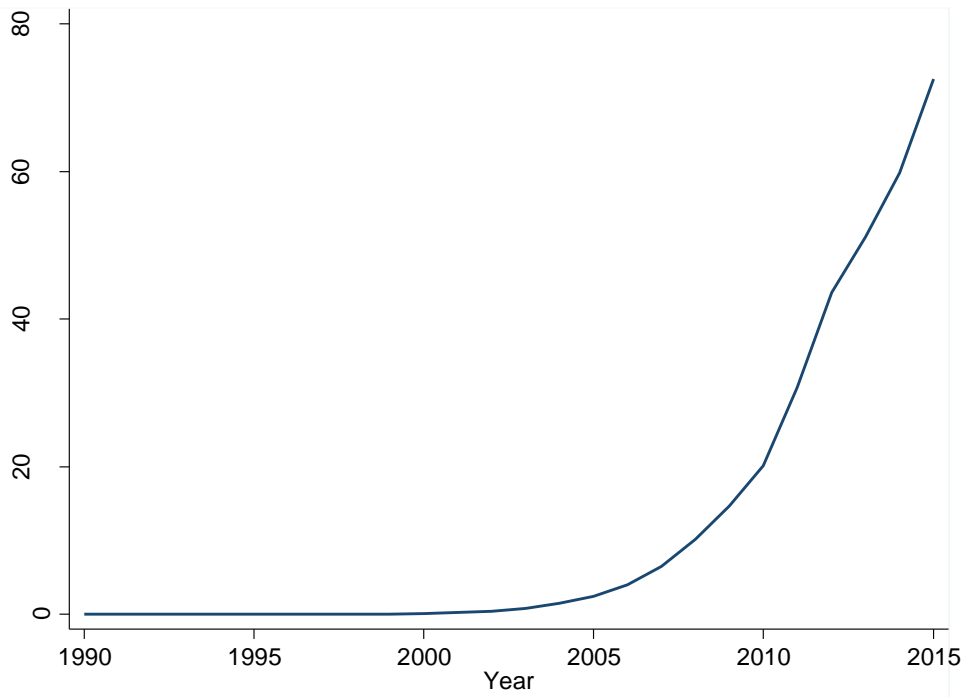
Source: Cisco Visual Networking Index; IP stands for Internet Protocol, BP stands for petabyte which is a multiple of the unit byte for digital information, i.e. 1000^5 bytes.

Table 1: Regressions as correlations for data policies and trade in services

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)
	MAINT	SEATRA	AIRTRA	OTHTRA	PO&CO	TRAVEL	CONSTR	INSUR	FINA
Data policy CB	0.274	0.572*	0.551**	-5.170**	-2.650**	0.805***	-0.904*	0.838**	1.860***
	(0.660)	(0.091)	(0.024)	(0.046)	(0.016)	(0.006)	(0.056)	(0.038)	(0.002)
Data policy DR	0.438	-0.678	-0.672	0.755	1.027	-0.130	2.332***	0.130	-1.607*
	(0.599)	(0.182)	(0.146)	(0.332)	(0.481)	(0.708)	(0.001)	(0.773)	(0.077)
FE country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	435	348	335	278	290	552	446	547	537
R2A	0.934	0.991	0.987	0.982	0.928	0.982	0.916	0.962	0.936
R2W	0.001	0.015	0.014	0.135	0.024	0.023	0.015	0.011	0.033
RMSE	0.510	0.169	0.171	0.260	0.481	0.198	0.590	0.315	0.434

Note: * $p<0.10$; ** $p<0.05$; *** $p<0.01$. The dependent variable the log of services imports ln(SM) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the year level. Fixed effects for sector is applied at 2-digit BPM level. CB denotes Cross-Border and covers all policies outlined under 1.1 in Annex 2. DR denotes Domestic Regulations and covers all policies outlined under 1.2 in Annex 2. Both data policy variables are lagged with 1 year. MAINT: Maintenance and Repair services, SEATRA: Sea transport, AIRTRA: Air transport, OTHTRA: Other modes of transport, PO&CO: Postal and Courier services, TRAVEL: Travel, CONSTR: Construction, INSUR: Insurance and Pension services, FINA: Financial services. Please note, due to insufficient observations, any variable controlling for services trade restrictions within each services sector is not possible.

Table 1: Regressions as correlations for data policies and trade in services (cont'd)

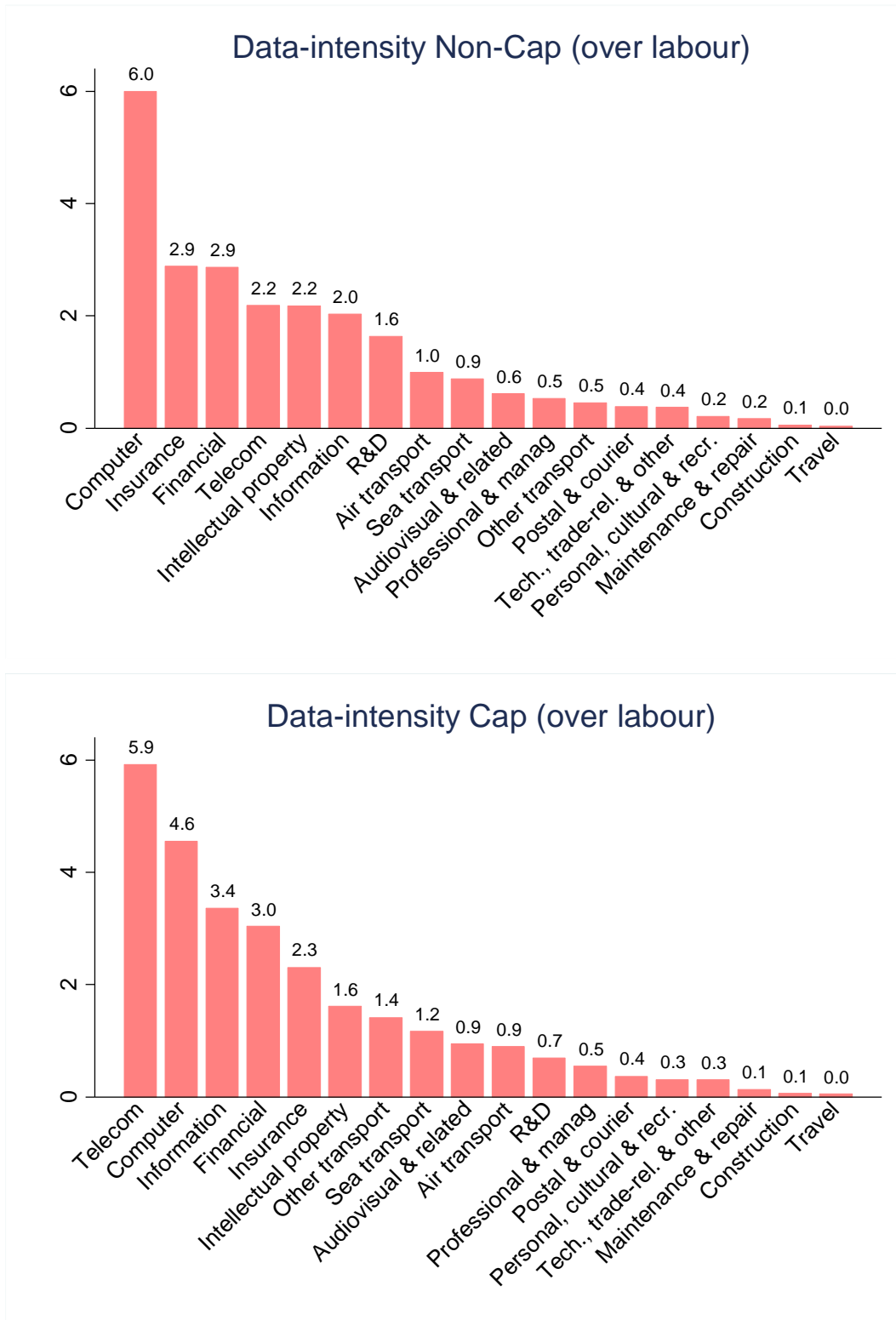|	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)	ln(SM)
	IPR	TELEC	COMP	INFO	R&D	PRO&M	TECH	AUDIO	OTHER
Data policy CB	-0.625**	1.010***	-0.292	1.508	-1.736*	0.143	-1.178	-1.941*	-0.090
	(0.018)	(0.009)	(0.421)	(0.176)	(0.088)	(0.783)	(0.160)	(0.069)	(0.965)
Data policy DR	-1.035**	-0.000	1.042	-3.739***	-1.363	0.191	0.672	0.086	0.508
	(0.021)	(0.999)	(0.160)	(0.003)	(0.106)	(0.765)	(0.144)	(0.832)	(0.526)
FE country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	537	411	363	345	333	374	441	315	362
R2A	0.974	0.961	0.979	0.915	0.970	0.970	0.970	0.940	0.886
R2W	0.016	0.017	0.011	0.040	0.021	0.000	0.014	0.011	0.001
RMSE	0.311	0.265	0.296	0.562	0.436	0.369	0.312	0.462	0.626

Note: * $p<0.10$; ** $p<0.05$; *** $p<0.01$. The dependent variable the log of services imports ln(SM) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the year level. Fixed effects for sector is applied at 2-digit BPM level. CB denotes Cross-Border and covers all policies outlined under 1.1 in Annex 2. DR denotes Domestic Regulations and covers all policies outlined under 1.2 in Annex 2. Both data policy variables are lagged with 1 year. IPR: Charges for the use of Intellectual Property Rights, TELEC: Telecommunication services, COMP: Computer services, INFO: Information services, R&D: Research and Development services; PRO&M: Professional services and & Management consulting; TECH: Technical, Trade-related and Other business services, AUDIO: Audiovisual and related services, OTHER: Other personal, Cultural and Recreational services. Please note, due to insufficient observations, any variable controlling for services trade restrictions within each services sector is not possible.

Figure 2: Data intensities as a ratio over of labour (D/L), US Census (2010)

Data-intensity Non-Cap (over labour)

Computer 6.0, Insurance 2.9, Financial 2.9, Telecom 2.2, Intellectual property 2.2, Information 2.0, R&D 1.6, Air transport 1.0, Sea transport 0.9, Audiovisual & related 0.6, Professional & manag 0.5, Other transport 0.5, Postal & courier 0.4, Tech., trade-rel. & other 0.4, Personal, cultural & recr. 0.2, Maintenance & repair 0.2, Construction 0.1, Travel 0.0

Data-intensity Cap (over labour)

Telecom 5.9, Computer 4.6, Information 3.4, Financial 3.0, Insurance 2.3, Intellectual property 1.6, Other transport 1.4, Sea transport 1.2, Audiovisual & related 0.9, Air transport 0.9, R&D 0.7, Professional & manag 0.5, Postal & courier 0.4, Personal, cultural & recr. 0.3, Tech., trade-rel. & other 0.3, Maintenance & repair 0.1, Construction 0.1, Travel 0.0

Source: Author's calculations using US Census, US BLS and BPM6 classification.

Table 2: Data producers

IO code	Sector description
511200	Software publishers
517110	Wired telecommunications carriers
517210	Wireless telecommunications carriers (except satellite)
518200	Data processing, hosting, and related services
519130	Internet publishing and broadcasting and Web search portals
541511	Custom computer programming services
541512	Computer systems design services
541513	Other computer related services, including facilities management

Source: BEA 2007 IO Use Table. IO codes follow NAICS 6-digit codes.

Figure 3: Data intensities as a ratio over labour (D/L), using BEA IO Use Table (2007)
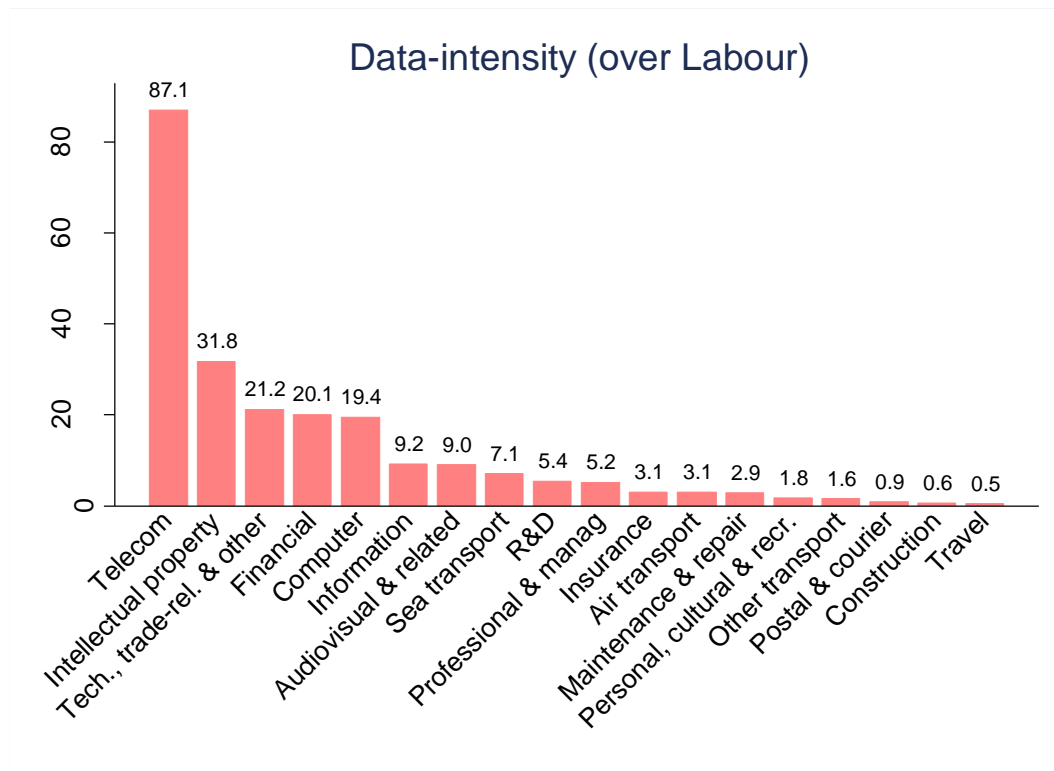

Source: Author's calculations using BEA, US BLS and BPM6 classification.

Table 3: Categories covering data policy index and weights

Categories	Type of measures	Weights
1. Cross-border flow measures		**0.5**
1.1	Ban to transfer or local processing requirement	0.5
1.2	Local storage requirement	0.25
1.3	Conditional flow regime	0.25
2. Domestic regulatory measures		**0.5**
2.1 Data retention		**0.15**
2.1.1	Minimum period	0.7
2.1.2	Maximum period	0.3
2.2 Subject rights on data privacy		**0.1**
2.2.1	Burdensome consent requirement	0.5
2.2.2	Right to be forgotten	0.5
2.3 Administrative requirements on data privacy		**0.15**
2.3.1	Data protection impact assessment (DPIA)	0.3
2.3.2	Data protection officer (DPO)	0.3
2.3.3	Data breach notification	0.1
2.3.4	Government access to personal data	0.3
2.4 Sanctions for non-compliance		**0.05**
2.4.1	Monetary fine above 250.000 EUR or set as a percentage of revenue	0.5
2.4.2	Jail time	0.5
2.5 Other restrictive practices related to data policies		**0.05**
2.5.1	Other restrictive practices related to data policies	1

.

Figure 4: Data policy index over time, all countries (2006-2016)
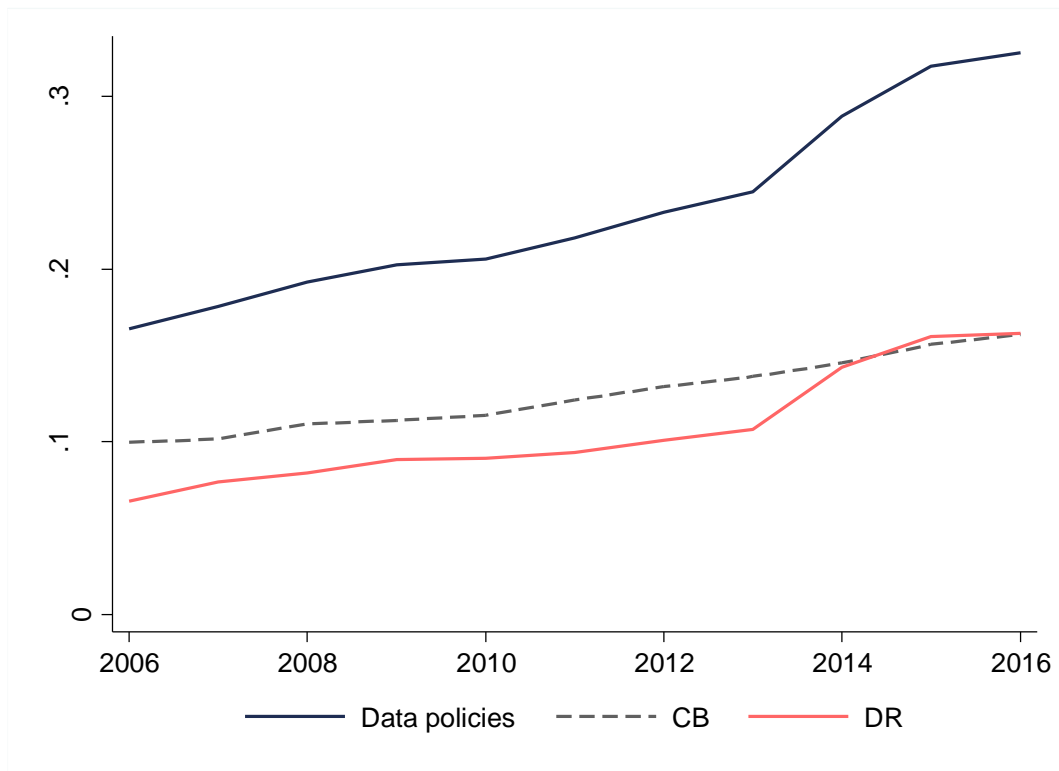
Source: ECIPE; the line for Data Policies covers both Cross-border (CB) and Domestic Regulatory (DR) policies, which is therefore composed of the two items. All three indexes are a weighted average across all countries using GDP constant (2010) as weight.

Figure 5: Per capita services imports and data-intensity (D/L), (2015)
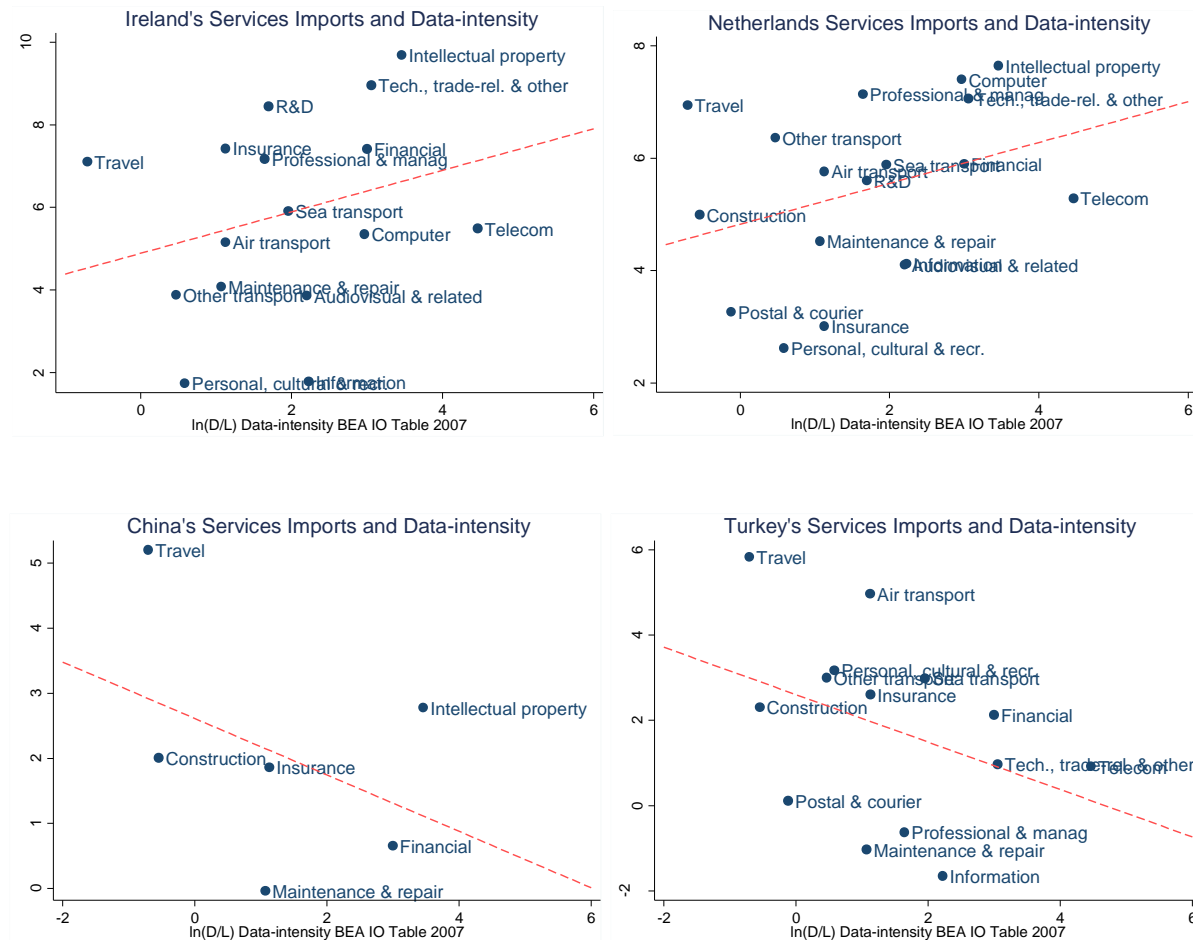
Source: Author's calculations using WTO-UNCTAD-ITC BPM6, World Bank World Development Indicators (WDI), BEA and US BLS.

Figure 6: Digital-intensive services imports in GDP and data policy index, (2015)

Table 4: Baseline regression results

	(1) ln(SM)	(2) ln(SM)	(3) ln(SM)	(4) ln(SM)	(5) ln(SM)	(6) ln(SM)	(7) ln(SM)
ln(D/L) * Data policy	-0.226***						
	(0.000)						
ln(D/L) * Data policy CB		-0.387***		-0.383***	-0.273***	-0.506***	-0.355***
		(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
ln(D/L) * Data policy DR			-0.204*	-0.012	0.075	-0.125	-0.007
			(0.086)	(0.924)	(0.463)	(0.386)	(0.954)
FE year-country	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE year-sector	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Type software in (D/L)	NC	NC	NC	NC	NC Pur	NC Lic	C
Observations	7250	7250	7250	7250	6959	6959	7250
R2A	0.780	0.780	0.779	0.780	0.777	0.777	0.780
R2W	0.002	0.002	0.000	0.002	0.001	0.004	0.002
RMSE	1.053	1.053	1.054	1.053	1.043	1.042	1.053

Note: * $p<0.10$; ** $p<0.05$; *** $p<0.01$. The dependent variable the log of services imports ln(SM) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry-year level. Fixed effects for sector is applied at 2-digit BPM level. The term (D/L) denotes the data intensity over labour using data from US Census on software and US BLS. Columns 1-4 show results using non-capitalized software expenditures (NC), column 5 shows results for using non-capitalized expenditures on purchases and payroll for developing software (NC Pur), column 6 shows the results for using non-capitalized expenditures on software licensing and service/ maintenance agreements for software licenses (NC Lic), column 7 shows results for using capitalized software expenditures (C). CB denotes Cross-Border and covers all policies outlined under 1.1 in Annex 2. DR denotes Domestic Regulations and covers all policies outlined under 1.2 in Annex 2.

Table 5: Extended baseline regression results

	(1) ln(SM)	(2) ln(SM)	(3) ln(SM)	(4) ln(SM)
ln(D/L) * Data policy	-0.199***			
	(0.004)			
ln(D/L) * Data policy * NRI	0.026			
	(0.457)			
ln(D/L) * Data policy CB		-0.329***		-0.271**
		(0.001)		(0.010)
ln(D/L) * Data policy CB * NRI		-0.012		-0.252***
		(0.822)		(0.007)
ln(D/L) * Data policy DR			-0.237*	-0.143
			(0.087)	(0.360)
ln(D/L) * Data policy DR * NRI			0.158*	0.458***
			(0.073)	(0.002)
FE year-country	Yes	Yes	Yes	Yes
FE year-sector	Yes	Yes	Yes	Yes
Observations	6327	6327	6327	6327
R2A	0.782	0.783	0.782	0.783
R2W	0.001	0.002	0.001	0.003
RMSE	1.050	1.049	1.050	1.049

Note: * p<0.10; ** p<0.05; *** p<0.01. The dependent variable the log of services imports ln(SM) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry-year level. Fixed effects for sector is applied at 2-digit BPM level. The term (D/L) denotes the data intensity over labour using data from US Census on non-capitalized software and US BLS. CB denotes Cross-Border and covers all policies outlined under 1.1 in Annex 2. DR denotes Domestic Regulations and covers all policies outlined under 1.2 in Annex 2. NRI stands for the Network Readiness Indicator sourced from the WEF and is demeaned.

Table 6: Baseline regression results with STRI

|	(1) ln(SM)	(2) ln(SM)	(3) ln(SM)	(4) ln(SM)	(5) ln(SM)	(6) ln(SM)	(7) ln(SM)
ln(D/L) * Data policy	-1.459***						
	(0.002)						
ln(D/L) * Data policy CB		-2.078***		-1.803**	-1.082***	-1.058***	-1.067***
		(0.007)		(0.019)	(0.003)	(0.003)	(0.003)
ln(D/L) * Data policy DR			-1.614**	-1.066	-0.481	-0.546	-0.543
			(0.028)	(0.155)	(0.236)	(0.183)	(0.185)
STRI	-3.216	-2.967	-3.662	-3.097	-0.942**	-1.050**	-1.095**
	(0.503)	(0.536)	(0.448)	(0.520)	(0.020)	(0.030)	(0.026)
FE year-country	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE year-sector	Yes	Yes	Yes	Yes	Yes	Yes	Yes
STRI classification	M1	M1	M1	M1	ENTIRE	MA&NT	DISCR
Observations	430	430	430	430	950	886	886
R2A	0.756	0.755	0.752	0.756	0.763	0.753	0.753
R2W	0.030	0.026	0.014	0.031	0.026	0.027	0.027
RMSE	0.966	0.968	0.974	0.967	0.951	0.949	0.948

Note: * p<0.10; ** p<0.05; *** p<0.01. The dependent variable the log of services imports ln(SM) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry-year level. Fixed effects for sector is applied at 2-digit BPM level. The term (D/L) denotes the data intensity over labour using data from US Census on non-capitalized software and US BLS. CB denotes Cross-Border and covers all policies outlined under 1.1 in Annex 2. DR denotes Domestic Regulations and covers all policies outlined under 1.2 in Annex 2. STRI refers to the OECD's Services Trade Restrictiveness Index; M1 refers to Mode 1, ENTIRE refers to the entire STRI with all its sub-components, MA&NT refers to the STRI that covers sub-components Market Access and National Treatment measures only, and DISCR refers to the STRI that covers sub-components discriminatory measures only.

Figure 7: Correlation between data-intensities ln(D/L) US Census 2011 and BEA IO Use Table 2007
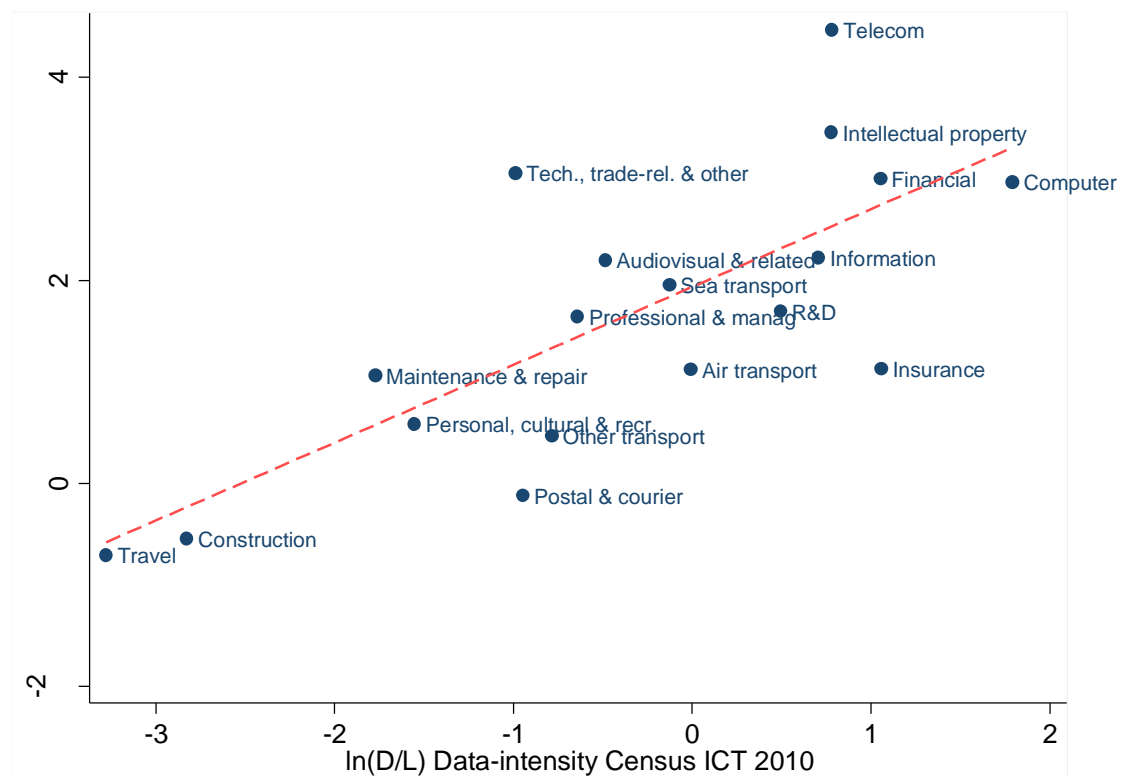
Source: US BEA and US Census. Note: The term ln(D/L) denotes the log of data intensity over labour. The vertical axis constructs this data intensity using data from the BEA IO Use Tables and US BLS for the year 2007. The horizontal axis constructs this data intensity using data from the US Census ICT survey based on non-capitalized software expenditure and US BLS. Data is originally given at 4 and 6-digit NAICS level and is concorded into BPM 2-digit sector.

Table 7: Extended regression results using BEA Use IO Table (D/L) for 2007

	(1) ln(SM)	(2) ln(SM)	(3) ln(SM)	(4) ln(SM)	(5) ln(SM)	(6) ln(SM)	(7) ln(SM)
ln(D/L) * Data policy	-0.815*						
	(0.051)						
ln(D/L) * Data policy CB		-1.023*		-0.841	-0.198**	-0.731**	-0.755***
		(0.094)		(0.157)	(0.034)	(0.013)	(0.009)
ln(D/L) * Data policy DR			-1.050*	-0.781	0.217*	-0.224	-0.242
			(0.097)	(0.200)	(0.066)	(0.465)	(0.423)
STRI	-4.055	-3.451	-4.331	-4.026		-0.900**	-1.076**
	(0.399)	(0.476)	(0.371)	(0.409)		(0.027)	(0.027)
FE year-country	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE year-sector	Yes	Yes	Yes	Yes	Yes	Yes	Yes
STRI class	M1	M1	M1	M1	NO	ENTIRE	MA&NT
Observations	430	430	430	430	7250	950	886
R2A	0.753	0.752	0.751	0.752	0.780	0.760	0.750
R2W	0.016	0.012	0.010	0.016	0.001	0.017	0.018
RMSE	0.973	0.975	0.976	0.974	1.054	0.956	0.953

Note: * $p<0.10$; ** $p<0.05$; *** $p<0.01$. The dependent variable the log of services imports ln(SM) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry-year level. Fixed effects for sector is applied at 2-digit BPM level. The term (D/L) denotes the data intensity over labour using data from the BEA IO Use Table 2007 and US BLS. CB denotes Cross-Border and covers all policies outlined under 1.1 in Annex 2. DR denotes Domestic Regulations and covers all policies outlined under 1.2 in Annex 2. STRI refers to the OECD's Services Trade Restrictiveness Index; M1 refers to Mode 1, ENTIRE refers to the entire STRI with all its sub-components, MA&NT refers to the STRI that covers sub-components Market Access and National Treatment measures only.

Table 8: Cross-country regression results using OECD-WTO BaTIS

	(1) ln(SM)	(2) ln(SM)	(3) ln(SM)	(4) ln(SM)	(5) ln(SM)	(6) ln(SM)	(7) ln(SM)
ln(D/L) * Data policy	-0.162						
	(0.646)						
ln(D/L) * Data policy CB		-0.331		-0.416	-0.366	-0.189	-0.229
		(0.523)		(0.493)	(0.535)	(0.739)	(0.680)
ln(D/L) * Data policy DR			-0.086	0.214	-0.147	-0.719	-0.891
			(0.909)	(0.806)	(0.855)	(0.335)	(0.207)
STRI	0.002	0.002	0.002	0.002	-0.004*	0.001	-0.005*
	(0.433)	(0.417)	(0.422)	(0.411)	(0.099)	(0.527)	(0.067)
FE country	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE sector	Yes	Yes	Yes	Yes	Yes	Yes	Yes
STRI class	M1	M1	M1	M1	ENTIRE	M1	ENTIRE
Year	2010	2010	2010	2010	2010	2009	2009
Observations	192	192	192	192	240	192	240
R2A	0.910	0.910	0.910	0.909	0.904	0.912	0.906
R2W	0.008	0.009	0.006	0.009	0.011	0.016	0.020
RMSE	0.573	0.573	0.574	0.575	0.581	0.560	0.570

Note: * $p<0.10$; ** $p<0.05$; *** $p<0.01$. The dependent variable the log of services imports ln(SM) using data from OECD-WTO BaTIS. Robust standard errors are clustered at the country-industry level. Fixed effects for sector is applied at 2-digit EBOPS 2002 level. The term (D/L) denotes the data intensity over labour using data from US Census on non-capitalized software and US BLS. CB denotes Cross-Border and covers all policies outlined under 1.1 in Annex 2. DR denotes Domestic and covers all policies outlined under 1.2 in Annex 2. STRI refers to the World Bank's Services Trade Restrictiveness Index; M1 refers to Mode 1, ENTIRE refers to the entire STRI with all its sub-components for Mode 1, 3 and 4. Year indicates for which year the cross-section is regressed.

Annex 1: The category of Royalties and licenses & Intellectual property

The category of Royalties and license and Intellectual property are two different names that refer to the same variable which are found in WTO-UNCTAD-ITC and OECD-WTO (BaTIS) trade in services databases. In the WTO-ITC-UNCTAD database, to which we refer as BPM6, this category is called Intellectual property whereas in the BaTIS database this category is denoted as Royalties and licenses.

Unfortunately, no direct connection between the NAICS 2007 classification and the sectors Royalties and licenses nor Intellectual property can be made from where we have computed our data intensities, i.e. (D/L). Equally unfortunate is that no concordance table exists between NAICS and BPM6 and NAICS and EBOPS more generally. Therefore, we have constructed our own concordance tables and build them up from an extremely detailed 6-digit level. This is not too difficult when mapping each 6-digit NAICS code into a 2-digit BPM6 or EBOPS code. However, since no clear 6-digit NAICS code can be directly linked to the services category of Royalties and license or Intellectual property, we have extended our concordance scheme to include this sector. We have done so in an indirect way through other concordance systems. The result of this concordance process can be seen in Table A1.1 below.

The way to do so is not clear-cut and some assumptions need to be made. For starters, the WTO-UNCTAD-ITC trade in services database designates Intellectual Property as chapter "SH" following the 6th edition of the Balance of Payments (BPM6) while the OECD-WTO BaTIS denotes this category as S266 following EBOPS 2002. As said, both overlap and are therefore indicated as "SH / S266" in Table A1.1. To eventually arrive at the NAICS 2007 code, two sequential sources are needed. First, the Annex III of the MSITS 2002 EBOPS classification provides a concordance table between EBOPS and CPC 1.0, which is used as a first step. Four sectors are classified under 266 Royalties and license fees, namely Patents, Trademarks, Copyrights and Other non-financial intangible assets. With the help of the United Nations correspondence tables website (https://unstats.un.org/unsd/classifications), a concordance can be made between CPC 1.0 and finally NAICS 2007 through five successive steps as outlined in Table A1.1.

Many different NAICS 2007 codes fall into one of the four original CPC 1.0 codes and therefore not all of them are equally relevant for Royalties and licenses or Intellectual property services. For that reason, we are not taking all 6-digit NAICS 2007 which eventually trace back to the two BPM6 and EBOP 2002 sectors as given in Table A1.1. The reason is that not all NAICS 2007 sectors are fully covered by the two intangible sectors. We only identify those which are not partially covered. These sectors are given in bold in column "NAICS 2002 / 07" of Table A.11 and are not given an * under the column "P" (which stands for partial). The information on whether an item is covered partially or not also comes from the United Nations correspondence tables. To come up with 2-digit BPM6 and EBOPS 2002 sector intensities, we take the unweighted average of each data intensity of these designated non-partial NAICS 2007 sectors, which should give us eventually a good approximation of the level of data used in the two sectors of Royalties and license and Intellectual property.

As one can see, a mix of services sector fall under the two sectors, namely R&D services, some financial services, as well as cultural services such as motion pictures and sound recording. Also trust funds are fully covered under this category of Royalties and license / Intellectual property. Of note, the NAICS sector 515120 is not included under EBOPS, but is covered under BPM6 following their respective manuals.

Table A1.1: Concordance for Intellectual property & Royalties and license fees

BPM6 / EBOPS	CPC 1.0	CPC 1.0 description	CPC Prov.	CPC Ver.1.1	ISIC Rev.3	ISIC Rev.3.1	NAICS 2002 / 07	P	NAICS 2002 / 07 description
SH / S266	51210	Patents	89210	81110 / 81120 / 81130 / 81140 / 81150 / 81190 / 81300	7310	7310	**541710**		R&D in agriculture, electronics, environmental, biology, botany, biotechnology, computers, chemistry, etc.
SH / S266	51220	Trademarks	89220	73340	6599	6599	**523110**		Acting as a principle in the underwriting or dealing of securities
SH / S266							523130	*	Acting as a principal in the buying or selling of commodity futures or spot market contracts
SH / S266							523910	*	Venture capital companies and own account investors
SH / S266							523991	*	Administration of trusts and private estates
SH / S266							**533110**		Patent holders, franchise issuers, and holders of similar intangible assets receiving royalties or licensing fees
SH / S266							**551111**		Bank holding companies
SH / S266							551112	*	Other financial holding companies (except bank holding companies)
SH / S266							**813211**		Foundations or charitable trusts raising funds and providing grants to applicants with specific qualifications or specific institutions
SH / S266	51230	Copyrights	89230	73310	7220	7221	**511210**		Software publishing
SH / S266				73320	9211 / 9249	9211	512110	*	Motion picture production
SH / S266							**512120**		Motion picture and video distribution to theatres, television and cable networks and stations, and exhibitors
SH / S266							**512191**		Post production services for motion pictures and video, such as editing, dubbing, subtitling, titling, animation, special effects, closed captioning, and similar
SH / S266							512199	*	Motion picture film laboratories and booking agencies
SH / S266							**512240**		Sound recording studios
SH / S266							512290	*	Recording books on tape (except publishers)
SH / S266					9213	9213	512110	*	Production of television shows, commercials, etc., not done in broadcast facilities
SH / S266							512290	*	Production of radio programs (except live) not done in broadcast facilities

SH / S266							515111	*	Radio networks, including transmission of purchased or self-produced content
SH / S266							515112	*	Radio stations, including transmission of purchased or self-produced content
SH / S266							**515120**	*	Television broadcasting, including transmission of purchased or self-produced content
SH / S266							515210	*	Cable networks, including transmission of purchased or self-produced content
SH / S266				9214	9214		561599	*	Ticket agencies, theatrical
SH / S266							711110	*	Theatre companies, comedy troupes, opera companies, live theatrical production
SH / S266							**711120**		Dance companies, groups, and ballets
SH / S266							**711130**		Musicians and musical groups
SH / S266							711190	*	Other performing arts companies, such as magic shows, ice skating shows, and similar
SH / S266							711310	*	Arts events organizers with facilities, arts facilities operators
SH / S266							711320	*	Arts events organizers without facilities
SH / S266							711510	*	Artists, writers, speakers, journalists, actors, cartoonists, dancers, producers, art restorers (all independent).
SH / S266	51290	Other non-financial intangible assets	89290	73390		6599	**523110**		Acting as a principle in the underwriting or dealing of securities
SH / S266							523130	*	Acting as a principal in the buying or selling of commodity futures or spot market contracts
SH / S266							523910	*	Venture capital companies and own account investors
SH / S266							523991	*	Administration of trusts and private estates
SH / S266							**533110**		Patent holders, franchise issuers, and holders of similar intangible assets receiving royalties or licensing fees
SH / S266							**551111**		Bank holding companies
SH / S266							551112	*	Other financial holding companies (except bank holding companies)
SH / S266							**813211**		Foundations or charitable trusts raising funds and providing grants to applicants with specific qualifications or specific institutions

Source: United Nations, BPM6 and EBOPS 2002.

Annex 2: Methodology for the data policy index

The data policy index covers those data policies considered to impose a restriction on the cross-border movement and the domestic use of data. The methodology to build the index follows Ferracane *et al.* (2018b) and covers the measures listed in the Digital Trade Estimates (DTE) database which is available on the ECIPE website (www.ecipe.org/dte/database). Starting from the DTE database, these policies are aggregated into an index using a detailed weighting scheme adapted from Ferracane *et al*. (2018b). We expand the index released by Ferracane *et al.* (2018b), which covered only the years 2016/2017, to create a panel for the years 2006-2016 that we can use in our regressions. In addition, the database and index are updated with new regulatory measures found in certain countries.

While certain policies on data flows can be legitimate and necessary to protect the privacy of the individual or to ensure national security, these policies nevertheless create a cost for trade and are therefore included in the analysis. The criteria for listing a certain policy measure in the DTE database are the following: (i) it creates a more restrictive regime for online versus offline users of data; (ii) it implies a different treatment between domestic and foreign users of data; and (iii) it is applied in a manner considered disproportionately burdensome to achieve a certain policy objective.

Each policy measure identified in any of the categories receives a score that varies between 0 (completely open) and 1 (virtually closed) according to how vast its scope is. A higher score represents a higher level of restrictiveness in data policies. The data policy index also varies between 0 (completely open) and 1 (virtually closed). The higher the index, the stricter the data policies implemented in the countries.

The index is composed of two sub-indexes that cover two main types of policy measures that we analyse in this paper: one sub-index covers policies on the cross-border movement of data and one sub-index covers policies on the domestic use of data. Analysing these two sub-indexes separately provides additional information on whether the impact of data policies on services trade varies according to the nature of the policies. The full data policy index is measured as the sum of these two sub-indexes. This annex presents in detail how the two sub-indexes are composed. It shows which policy measures are contained in each of the sub-index and the scheme applied to weigh and score each measure.

The list of measures included in the two sub-indexes is summarised in Table 3. As shown in the table, the sub-indexes are measures as a weighted average of different types of measures. The weights are intended to reflect the level of restrictiveness of the types of measures in terms of costs for digital trade. The first sub-index on cross-border data flows covers three types of measures, namely (i) a ban to transfer data or a local processing requirement for data; (ii) a local storage requirement, and (iii) a conditional flow regime. The second sub-index covers a series of subcategories of policies affecting the domestic use of data. These are: (i) data retention requirements, (ii) subject rights on data privacy, (iii) administrative requirements on data privacy, (iv) sanctions for non-compliance, and finally, (v) other restrictive practices related to data policies.

The main sources used to create the database are national data protection legislations. Otherwise, information is obtained from legal analyses on data policies and regulations from high profile law firms and from OECD (2015). Moreover, occasionally corporate blogs and business reports were also taken into consideration, as they can provide useful information on the de facto regime faced by the company when it comes to movement of data.

1.1 Sub-index on cross-border data flows

The first sub-index covers those policy measures restricting cross-border data flows. These measures are also referred to as "data localisation" measures and can be defined as government imposed measures

which result in the localisation of data within a certain jurisdiction. Measures related to data localisation come in various forms and have different degrees of restrictiveness depending on the type of measure itself, but also on the sector and type of data affected.

We identify three types of measures, namely (i) a ban to transfer data or a local processing requirement for data; (ii) a local storage requirement, and (iii) a conditional flow regime. As shown in table 3, the category of bans to transfer and local processing requirements has a score of 0.5, while the other two categories have a score of 0.25 each. The sum of the scores of these categories can go up to 1, that reflects a situation of virtually closed regime on cross-border data flows. This score is multiplied by 0.5 to create the final sub-index on cross-border data flows. The sub-index therefore goes from 0 (completely open) to 0.5 (virtually closed).

Bans to transfer data across the border and local processing requirements are the most restrictive measures on cross-border flow of data. In case of a ban to transfer data or a local processing requirement, the company needs to either build data centres within the implementing jurisdiction or switch to local service providers with a consequent increase in costs if these domestic service providers are less efficient than foreign providers. The difference between bans to transfer and local processing requirements is quite subtle. In the first case, the company is not allowed to even send a copy of the data cross-border. In the second case, the company can still send a copy of the data abroad - which can be important for communication between subsidiary and its parent company and in general for exchange of information within the company. In both cases, however, the main data processing activities need to be done in the implementing jurisdiction.

For the scoring of these measures, both the sectoral coverage of the measure as well as the type of data affected are taken into account. If the ban to transfer or local processing requirement applies to a specific subset of data (for instance, when it applies to health records or accounting data only), this measure receives a scoring of 0.5. A similar score is also assigned when the restriction only applies to specific countries (for instance, when data cannot be sent for processing only to a specific country). On the other hand, when the measure applies to all personal data or data of an entire sector (such as financial services or telecommunication sector), then a score of 1 is given. Measures targeting personal data also receive the highest score because it is often hard to disentangle personal information versus non-personal information, and therefore measures targeting personal data often end up covering the vast majority of data in the economy (MIT, 2015). The score, as always, goes from 0 (completely open) to 1 (virtually closed). Therefore, if there are two measures scoring 0.5, the score is 1. If there are more additional measures, the score for this category still remains one. This score is then weighted by 0.5 which is the weight assigned to the category of bans and local processing requirements (as presented in Table 3).

The second category covers local storage requirements. These measures require a company to keep a copy of certain data within the country. Local storage requirements often apply to specific data such as accounting or bookkeeping data. As long as the copy of the data remains within the national territory, the company can operate as usual. As for the scoring, when data storage is only for specific data as defined above, this measure receives a score of 0.5, whereas when the data storage applies to personal data or to an entire sector, it receives a score of 1. As mentioned before, the score goes up to 1 maximum and is then weighted by 0.25 which is the weight assigned to the category of local storage requirements (as presented in Table 3).

The third category of cost-enhancing measures related to cross-border flow of data is the case of conditional flow regime. These measures forbid the transfer of the data abroad unless certain conditions are fulfilled. If the conditions are stringent, the measure can easily result in a ban to transfer. The conditions can apply either to the recipient country (e.g. some jurisdictions require that data can be transferred only to countries with an "adequate" level of protection) or to the company (e.g. a condition might consist in the need to request the consent of the data subject for the transfer cross-border of his/her data). In terms of scoring, if a conditional flow regime is found, it receives a score of 0.5 in case it

applies to specific data, but it receives a score of 1 in case conditions apply for personal data and or the entire sector. The final score is then weighted by 0.25, which is the weight assigned to the category of conditional flow regimes.

Of note, in certain cases it is not easy to discern whether a measure is a ban to transfer, a local processing requirement or a conditional flow regime. In fact, often cases of ban to transfer and local processing requirements have certain exceptions which might de facto result in a conditional flow regime. When the exceptions are quite wide (for example, if they include the request for consent from the data subject), then the measure has been categorized as a conditional flow regime.

Figure A2.1 shows a graphical representation of the various levels of data localisation measures taken up in this sub-chapter. The direction of the arrow indicates the increased level of restrictiveness. Note that conditional flow regime is put outside this conventional sequence of restrictiveness because it prevents the flow of data only when the conditions are not fulfilled. Also, note that in Table 3 the ban to transfer is put together with local processing requirements although these two measures have actually been separated in Figure A2.1. The point is that the impact of those measures on trade is very similar and they are not always easy to discern. Yet, a ban to transfer is generally more restrictive than a local processing requirement.

Figure A2.1: Graphical overview of data policies
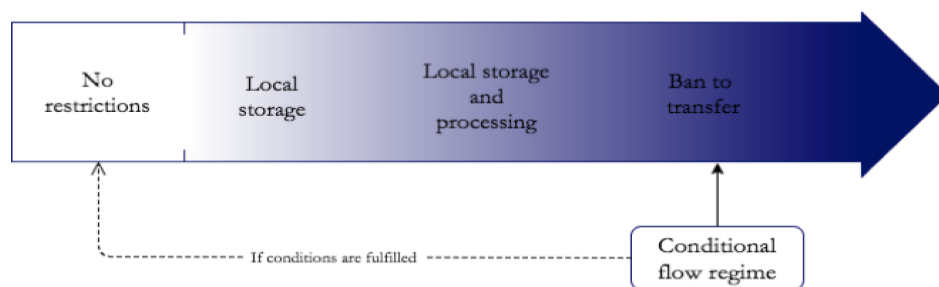
Source: Ferracane (2017)

1.2 Sub-index on domestic use of data

The sub-index on domestic use of data index covers a series of subcategories of policies affecting the domestic use of data. These are: (i) data retention requirements, (ii) subject rights on data privacy, (iii) administrative requirements on data privacy, (iv) sanctions for non-compliance, and finally, (v) other restrictive practices related to data policies. Given that each of these sub-categories contains, in turn, additional sub-categories, they will be presented separately. For the calculation of the sub-index, the weights assigned to the categories are shown in Table 3. The categories with the highest weights (and therefore those which are considered to create higher costs for digital trade) are data retention and administrative requirements on data privacy, which are assigned a weight of 0.15 each. The category of subject rights on data privacy is assigned a score of 0.1, while the other two categories of sanctions for non-compliance and other restrictive practices are assigned a score of 0.05.

The sum of the scores of these categories can go up to 0.5 that reflect a situation of virtually closed regime on domestic use of data. The sub-index therefore goes from 0 (completely open) to 0.5 (virtually closed). As mentioned above, the data policy index is measured as the sum of the two sub-indexes and therefore the score for the final data policy index goes from 0 to 1.

1.2.1 Data retention

The first category belonging to the sub-index on domestic use of data deals with measures related to data retention, which are measures regulating how and for how long a company should keep certain data within its premises. Data retention measures can define a minimum period of retention or a maximum period of retention. In the first case, the companies (often telecommunication companies) are required to retain a set of data of their users for a certain period, which can go up to two years or more in some cases. These measures can be quite costly for the companies and they are assigned a weight of 0.7. On the other hand, the measures imposing a maximum period of retention are somewhat less restrictive and prescribe the company not to retain certain data when it is not needed anymore for providing their services. They are therefore given a weight of 0.3. The country receives a score of 1 in each of the two sub-categories when there is a one or more measures implemented, while 0 is assigned in case of absence of these measures. Therefore, if a country implements one or more data retention requirements for a minimum period of time and no data retention requirements for a maximum period of time, the score will be 0.7. Alternatively, if the country only implements one requirement of maximum period of data retention, the score will be 0.3.

1.2.2 Subject rights on data privacy

The second category belonging to the sub-index on domestic use of data includes measures related to subject rights on data privacy. The rights of the data subject are often a legitimate goal in itself, but they can nonetheless represent a cost for the firm when they are implemented disproportionately or in a discriminatory manner. This is the reason why they are covered in the database. However, they only form a smaller part of the sub-index with a weight of 0.1 as their cost on business is significantly low compared with other measures. Two categories of measures are identified regarding data subject rights, which are (i) the need for consent for the collection and use of data (with a weight of 0.5) and (ii) the right to be forgotten (with also a weight of 0.5).

If one of the measures applies, a score of 1 is given whereas a score of 0 is assigned otherwise. Regarding the first measure on the request of consent for the collection and use of data, a score of 1 is given only when the process for requesting consent is considered as disproportionately burdensome. This is the case when the consent has to be always written and explicit or when consent is required not only for the collection of data, but also for any transfer of data outside the collecting company. If this is not the case, then a score of 0 is assigned. Additionally, important to note is that, if the consent is required only in case of transfer across borders, this measure is instead reported in the first sub-index under conditional flow regime and scored accordingly.

1.2.3 Administrative requirements on data privacy

The third category belonging to the sub-index on domestic use of data covers administrative requirements on data privacy. Measures included in this category are (i) the requirement to perform a data privacy impact assessment (DPIA) (with a weight of 0.3), (ii) the requirement to appoint a data protection officer (DPO) (with as well a weight of 0.3), (iii), the requirement to notify the data protection authority in case of a data breach (with a weight of 0.1), and finally (iv) the requirement to allow the government to access the personal data that is collected (with also a weight of 0.3).

For the scoring, the first two measures receive a score of 1 when a measure applies and 0 otherwise. In the case of the fourth measure, which is the requirement to allow government to access collected personal data, a full score of 1 is assigned only when the government has an open access to data in at least one sector of the economy. However, if a government has only access to escrow or encryption keys, but still notifies access to the data, an intermediate score of 0.7 is assigned. Government direct access to data handled by the company or the use of escrow keys may in fact create remarkable consumer

dissatisfaction that may lead to the user's termination of service demand. Finally, if the government has to follow the same procedure that it would follow for offline access to data - that is, the presence of a court decision or a warrant, or when the request follows a judicial investigation process - then the score is 0.

1.2.4 Sanctions for non-compliance

The fourth category belonging to the sub-index on domestic use of data examines measures which impose a sanction for non-compliance. These measures cover both pecuniary and penal sanctions with a weight of 0.5 for each of them. The pecuniary sanctions are not considered a restriction per se, but their presence is listed in the database and accounted for in the sub-index when (i) they are above 250.000 EUR; (ii) companies have explicitly complained about disproportionately high fines or discriminatory enforcement of sanctions; (iii) they are expressed as a percentage of a company's domestic or global turnover. In fact, in all these cases, the sanctions have the capacity of putting a company out of business and might play an important role in the economic calculation of a company. We also list under this section those instances in which the infringement of data privacy rules can be sanctioned by closing down the business. On the other hand, the application of penal sanctions such as jailtime as a result of infringement of data privacy rules is included in the database as a restriction. Instances in which penal sanctions are assigned as a result of identity theft and similar illegal actions are obviously not included. For what concerns the scoring, if these cases are identified, a score of 1 is assigned.

1.2.5 Other measures

Finally, the last category takes up all those measures which are related to domestic use of data, but do not fit under any of the aforementioned categories. All these measures are assigned a score of 1.

Figure A2.2: Density graphs for Policy index and Policy index * ln(D/L), all countries (2015)
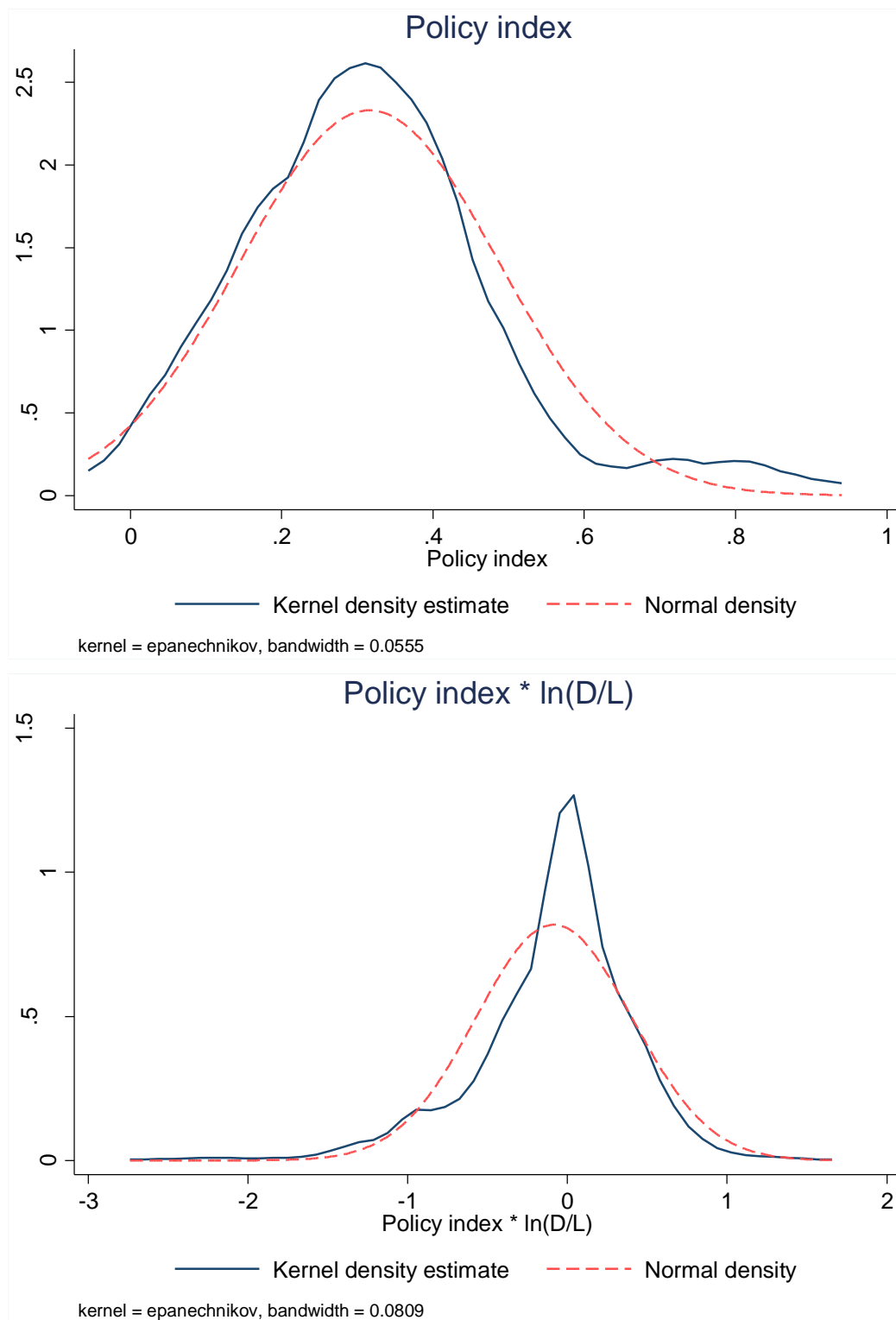
Source: Ferracane *et al.* (2018b). The Kernel density estimate is done for the overall Data Policy variables as taken up in all regressions with N=64 for all countries covered in the DTRI.

Annex 3: Concordance table between OECD STRI and BPM6 services classification

Figure A3.1: Concordance table between OECD STRI sector codes and BPM6 codes of the WTO-UNCTAD-ITC database.

BPM6	Sector code	Sector description	OECD STRI					
SB	MAINT	Maintenance and repair services n.i.e.						
SC1	SEATRA	Sea transport	TRMAR					
SC2	AIRTRA	Air transport	TRAIR					
SC3	OTHTRA	Other modes of transport	TRRAI	TRROF	LSCAR	LSCUS	LSFGT	LSSTG
SC4	PO&CO	Postal and courier services	CR					
SD	TRAVEL	Travel						
SE	CONSTR	Construction	CO					
SF	INSUR	Insurance and pension services	FSINS					
SG	FINA	Financial services	FSBNK					
SH	IPR	Charges for the use of intellectual property n.i.e.	ASMOT	ASSOU	ASBRD	CS	DS	
SI1	TELEC	Telecommunications services	TC					
SI2	COMP	Computer services	CS					
SI3	INFO	Information services						
SJ1	R&D	Research and development services						
SJ2	PRO&M	Professional and management consulting services	PSACC	PSLEG				
SJ3	TECH	Technical, trade-related, and other business services	PSARC	PSENG				
SK1	AUDIO	Audiovisual and related services	ASMOT	ASSOU	ASBRD			
SK2	OTHER	Other personal, cultural, and recreational services						

Note: For abbreviations in column "Sector code", see the two footnotes in Table 1. TRMAR: Maritime transport, TRAIR: Air transport, TRAI: Train freight transport, TFFOF: Road freight transport, LSCAR: Logistics cargo handling, LSCAR: Logistics customs brokerage, LSFGT: Logistics freight forwarding, LSSTG: Logistics storing and warehouse, CR: courier services, CO: Construction, FSINS: Insurance services, FSBNK: Commercial banking services, TC: Telecommunication, CS: Computer services, PSACC: Accounting services, PSLEG: Legal services, PSARC: Architectural services, PSENG: Engineering services, ASMOT: Motion picture, ASSOU: Sound recording, ASBRD: Broadcasting.

Annex 4: Data intensities with WTO-OECD BaTIS services classification

Figure A4.1: Data intensities as a ratio over of labour (D/L), US Census (2010) for BaTIS

Source: Author's calculations using US Census, US BLS and EBOPS 2002 classification.

Author contacts:

Martina Francesca Ferracane

Hamburg University, European University Institute, European Centre for International Political Economy (ECIPE)

Email: martina.ferracane@gmail.com

Erik van der Marel (corresponding author)

Senior Economist at ECIPE & Associate Professor at Université Libre de Bruxelles (ULB) & ECARES

Avenue des Arts 40

B-1050 Brussels

Email: erik.vandermarel@ecipe.org