

# Data Flows Restrictiveness and (Un)equal Productivity and Jobs Effects\*

Ana Cusolito<sup>†</sup> Erik van der Marel<sup>‡</sup> Gaurav Nayyar<sup>§</sup> Regina Pleninger<sup>¶</sup>

August 5, 2025

## Abstract

This paper presents firm-level evidence on the uneven impact of cross-border and domestic digital data restrictions—hindering personal data transfers and data processing, respectively—on revenue-based productivity (TFPR) and employment, using a new database covering 32 European countries from 2010 to 2017. The findings show that, while the median marginal effect of both types of restrictions is negative, significant heterogeneous effects exist. Cross-border restrictions, such as local storage requirement or local processing requirements for personal data, reduce the productivity gap between leaders and laggards by disproportionately taxing high-productivity firms, whereas domestic restrictions, such as data retention or the requirement to have a data protection officer in place, widen the gap by burdening low-productivity firms more. Cross-border liberalization benefits service firms, while domestic liberalization favors manufacturing companies. Additionally, cross-border restrictions do not affect labor demand, but domestic restrictions increase it. Overall, the paper demonstrates that data policies neutral in design can nonetheless reshape market dynamics due to their differential impact on productivity across sectors and firms—even when the sector ranking by sensitivity to these restrictions remains consistent regardless of whether the regulations are cross-border or domestic.

*JEL Codes:* D22, D24, L25, O47.

*Keywords:* digital data restrictions, productivity, jobs, firm inequality.

---

\*The opinions expressed in this paper do not represent the views of the World Bank Group, its Board of Directors or the Governments they represent. All errors and omissions are the authors' responsibility. We would like to thank Jia Liu for excellent assistance in the preparation of the draft.

<sup>†</sup>World Bank. [acusolito@worldbank.org](mailto:acusolito@worldbank.org).

<sup>‡</sup>ECIPE. Corresponding author. [erik.vandermarel@ecipe.org](mailto:erik.vandermarel@ecipe.org).

<sup>§</sup>World Bank. [gnayyar@worldbank.org](mailto:gnayyar@worldbank.org).

<sup>¶</sup>World Bank; KOF Swiss Economic Institute at ETH Zurich. [rpleninger@worldbank.org](mailto:rpleninger@worldbank.org).

# 1 Introduction

The role of digital technologies has been central to discussions on productivity growth over the past few decades. While a growing body of literature provides evidence of a positive link between digitalization and productivity, the relationship between digital technologies and the gap between less and more productive firms remains relatively underexplored. This paper seeks to address this gap by examining the extent to which data-related restrictions—likely to hinder firms’ use of digital technologies—contribute to productivity dispersion.

Previous research shows that inequality in productivity between firms has increased across countries.<sup>1</sup> This widening gap between less productive and more productive firms has been attributed, at least in part, to the diffusion of digital technologies, although empirical evidence remains limited. Using data from across European countries, Gal et al. (2019) find that the positive association between digital adoption in an industry and productivity gains at the firm level is stronger for more productive firms. Similarly, analyzing firm-level data across 13 EU countries, Bloom, Sadun, and Reenen (2012) find that foreign affiliates of U.S. multinationals achieve higher productivity from their IT capital—compared to non-U.S. multinationals and domestic firms—due to better management practices. More broadly, there has been a decline in the intensity of knowledge diffusion from frontier to laggard firms (Akcigit and Ates 2023).

Moreover, the empirical literature offers even less discussion on how specific digital-related policies—although neutral in design—may produce differential effects in productivity across countries, sectors, and firms. One exception is Ferracane, Kren, and van der Marel (2020), who find that regulatory measures restricting the cross-border flow and commercial use of personal data are not only negatively and significantly associated with firm-level productivity but also disproportionately penalize smaller firms. This outcome, in turn, is likely to increase productivity dispersion among firms using digital technologies, given that access to data is a fundamental enabler of digitalization. To estimate the effect of interest, the authors implemented a two-step approach. In the first stage,

---

1. Research assessing this increase in inequality of productivity between firms across countries, without necessarily attributing this to digital technologies, include (Cusolito, Garcia-Marin, and Maloney 2023; De Loecker et al. 2016; Andrews et al. 2015; Atkin et al. 2015; Eslava et al. 2024)

they estimate productivity, omitting how regulatory data restrictions affect digitization, factor demand and, therefore, productivity. In the second stage, they regress the estimated productivity measures on the policy index. This approach is not only unrealistic from a business perspective but also invalidates the moment conditions needed to identify the production function elasticities. As a result, productivity measures are upwardly biased, given that stringent trade regulatory frameworks not only reduce the volume of trade but are also factor saving.

This paper overcomes these limitations in its examination of the differential effects of data-related restrictions on firm-level productivity in European countries to better understand the link between digitalization and productivity dispersion. The estimates indirectly capture the productivity gains associated with digital technologies because restrictions on the ability of firms to transfer and use personal data are correlated with digital technology adoption. If anything, estimating the productivity impacts of digital policy restrictions allows for a cleaner identification strategy compared to using measures of digital technologies directly. Put differently, in our case these policy changes related to data are more exogenous to the prevailing productivity dynamics than any measure of digital technology use, such as Gal et al. (2019). Our approach also addresses the methodological limitations of Ferracane, Kren, and van der Marel (2020) in several additional respects. First, we endogenize the productivity process to make it a function of the policy framework. Second, we estimate a semi-parametric flexible functional form that allows us to explore how the estimated marginal effect of regulatory data restrictions varies along the productivity distribution. Finally, we carefully show differences between services and manufacturing in the context of this relationship, which is also less explored in the previous literature. In addition, this paper quantifies, through counterfactual analysis, the aggregate productivity gains that countries can obtain by introducing regulatory reforms in data to catch up with the policy frontier.

To conduct the analysis, we develop a new database that combines data-related restrictiveness measures, sector-level digital elasticities, and firm-level productivity data from European countries. The data on data-related policies cover 15 specific types of regulations that affect how firms

conduct cross-border transfers and domestic processing of personal data.<sup>2</sup> It includes an aggregate restrictiveness index of data regulations, divided into two sub-indexes: one measuring regulations specific to cross-border data flows, and another covering domestic data use regulations. The former category encompasses regulatory restrictions such as outright bans on transferring personal data to third countries, local storage and infrastructure requirements, and highly stringent conditions under which personal data transfers may occur. These measures directly constrain firms' ability to move personal data across borders. The latter category consists of domestic regulations, including mandatory minimum data retention periods, requirements to conduct Data Protection Impact Assessments (DPIAs), obligations to appoint a Data Protection Officer (DPO), and provisions granting government access to personal data. These rules apply to all firms, regardless of their involvement in cross-border data transfers. While some of these measures are intended to safeguard individuals' privacy, they can also impose significant compliance burdens on businesses. The policy variables are based on the earlier work by Ferracane, Kren, and van der Marel (2020) and Ferracane and van der Marel (2021).

To capture the potential differential effects of digital trade restrictions across sectors, we develop sector-level digital elasticities—defined as a sector's data dependence—computed using input-output matrices. We then merge this database with firm-level balance sheet information from Orbis, which we use to calculate firm-level Total Factor Productivity (TFP). The final dataset covers 14 European countries over the period 2010–2017, the period before the EU's GDPR was introduced.

The results of our paper show that data policies affecting both cross-border and domestic digital trade reduce firm-level productivity. A one percentage point increase in cross-border or domestic restrictions has a negative median marginal effect of approximately 24 percent. However, 46.5 percent (for cross-border restrictions) and 45.2 percent (for domestic restrictions) of firms in the sample experience productivity gains from these restrictions. The marginal effects also vary along the productivity distribution. While low-productivity firms benefit from cross-border restrictions,

---

2. Personal data refers to personal information that firms use in the production process of delivering the good or service. See Goldfarb and Treffer (2018) for a discussion how restrictions on data affect firm performance and trade.

high-productivity firms experience productivity losses. In contrast, the opposite effect is observed for domestic data restrictions: low-productivity firms are negatively affected by more restrictive data privacy regulations. Finally, because the negative effect of cross-border restrictions outweighs that of domestic restrictions, data-related policies—as captured by our index—disproportionately burden high-productivity firms.

**Related Literature** Our paper relates to two separate strands of the research, namely one that focuses on estimating the impact of digital technologies on firm performance, and another that examines the effects of regulatory data-related regulations and restrictions on firm performance, including the quantification of aggregate productivity gains that countries can achieve by relaxing these constraints. First, on the relationship between digital technology use and firm performance, our paper is close to Gal et al. (2019) who assess the impact of digital technologies on firm-level productivity. We build on this work by analyzing productivity dispersion rather than focusing solely on average productivity effects. Additionally, we measure digitalization in an exogenous manner by using data-related policy restrictions, given that digital technologies are typically endogenous to economic performance. More broadly, our paper relates to the literature that assesses how digital technologies affect firm dynamics, especially smaller firms such as works by Sinclair and Livingston (2005), Lucking-Reiley et al. (2007), Melnik and Alm (2002), and Cabral and Hortacsu (2010), and more recently DeStefano, Kneller, and Timmis (2023) using cloud computing technology, as well as new literature how diffusion of more advanced technologies affect productivity performance, such as Graetz and Michaels (2018), Kleinberg et al. (2018), and Mullainathan and Obermeyer (2022) and Brynjolfsson, Li, and Raymond (2025).

Second, on the relationship between data-related regulations and firm performance, our paper is very close to Ferracane and van der Marel (2021) and Ferracane, Kren, and van der Marel (2020), who show a negative association between stricter data policies and the performance of downstream firms in sectors reliant on electronic data. We build on this work by making the productivity process a function of the policy framework and estimating a functional form that allows us to explore how the estimated marginal effect of regulatory data restrictions varies along

the productivity distribution. Related work to ours also includes Sun and Trefler (2023), Cory and Dascoli (2021), and Cory (2017). Sun and Trefler (2023) find that cross-border restrictions on data flows halved the AI algorithm-related increases in international trade, while Cory and Dascoli (2021) and Cory (2017) show that data-localization policies that make it more expensive and time consuming, if not illegal, to transfer data overseas may decrease the economic gains through the trade, productivity, and price channels.

Finally, our paper also relates to an emerging set of empirical studies that examine firm-level effects of the EU’s General Data Protection Regulation (GDPR), such as Frey and Presidente (2024) and Demirer et al. (2024). These studies typically assess the impact of data-related regulations stemming from the GDPR, which has governed the data privacy of EU citizens—including cross-border transfers of personal data—since 2018. Our analysis differs from this literature in that it focuses on a period preceding the implementation of the GDPR. This earlier time frame allows us to examine cross-country variation in data-related policies before EU member states harmonized their data privacy regulations.<sup>3</sup> Moreover, our paper further departs from these two previous works by examining productivity dispersion among firms as a consequence of data-related policies, rather than focusing solely on whether firm-level productivity or any other firm-performance increased or decreased as result of the regulation.

Our paper is organized as follows. Section 2 describes the dataset we build to conduct the analysis. Section 3 presents the methodology. Section 4 discusses the estimation results. Section 5 displays the results from the counterfactual liberalization reform analysis. The last section concludes.

---

3. That said, although the GDPR aims to harmonize national data privacy regulations, significant variation in its implementation across countries persists due to differences in compliance levels and additional country-specific provisions introduced by national governments on top of the GDPR. See also Johnson (2022).

## 2 Data and Descriptive Statistics

To perform the analysis, we construct and merge three datasets: (1) the restrictiveness of the data policy index covering data-related restrictions, (2) the sectoral elasticities derived from the OECD Inter-Country Input-Output tables, and (3) the firm-level data obtained from the ORBIS database.

### 2.1 Data Flows Restrictiveness Index

The Data Flows Restrictiveness Index (DFRI) quantifies the degree of regulation governing cross-border and domestic data flows across countries, encompassing a wide spectrum of regulatory measures. This index is part of ECIPE's wider Digital Trade Restrictiveness Index (Ferracane and van der Marel 2021; Ferracane, Kren, and van der Marel 2020) and measures the full range of data-related regulations, both from a cross-border and domestic perspective. The former captures the extent to which firms are permitted to transfer personal data across borders without any restriction, whereas the latter covers regulatory requirements that domestic firms must comply for data processing in relation to data protection, such as subject rights and administrative requirements.

Cross-border restrictions cover regulatory measures such as outright bans on transferring personal data to third countries, local storage and infrastructure requirements (e.g., mandates to build data centers), and highly stringent conditions under which personal data transfers may occur. These measures are designed to increase the costs of conducting international business that requires the transfer of personal data across borders as part of their operations. Moreover, such restrictions have significant downstream effects on digital services that depend on the international movement of data. For example, mobile applications rely heavily on the cross-border transfer of data, as they often support international transactions. Limiting these flows through cross-border restrictions is therefore likely to reduce the usage of such apps by foreign users, as evidenced by Sun and Trefler (2023). Furthermore, as artificial intelligence relies on cross-border data flows, such restrictions also affect the digital delivery of services and transactions generated by AI algorithms—such as machine translation on platforms—as demonstrated by Brynjolfsson, Hui, and Liu (2019).

Domestic regulatory measures cover rules that firms must comply with to ensure the protection of data subjects (i.e., domestic citizens). Examples include mandatory minimum data retention periods for firms handling personal data or burdensome consent requirements by data subjects, which increase data processing costs. Other examples are the requirements to conduct Data Protection Impact Assessments (DPIAs) and the obligation to appoint a Data Protection Officer (DPO), which may necessitate the recruitment of additional personnel. While these measures aim to enhance data protection, they can also impose significant compliance costs on all businesses—even those not directly affected by cross-border data restrictions—as shown by Frey and Presidente (2024). Beyond these costs, when privacy regulations are particularly stringent, they are likely to hinder data-reliant business activities of both domestic and foreign firms when offering their services. Recent studies on the GDPR’s impact on digital transactions, such as those concerning web technologies and cloud computing services, demonstrate this effect as shown by Peukert et al. (2022); Demirer et al. (2024). It is worth noting that some domestic regulations may also require firms to grant government access to personal data. While these measures are often justified on national security grounds, they can undermine consumer trust and potentially deter businesses from engaging in data-driven transactions.

Our analysis considers both categories of data-related restrictions, as detailed in Table 1. The dataset comprises 15 distinct regulatory restrictions across 31 European countries, spanning an eight-year period from 2010 to 2017. Table 2 presents summary statistics for the DFRI, disaggregated into its two components. Both subindices are scaled from 0 to 1, where a value of 0 indicates full liberalization of data-related restrictions and a value of 1 corresponds to the most restrictive regulatory regime for data. On average, the cross-border subindex exhibits lower levels of regulatory restrictiveness, as indicated by its lower mean value. Furthermore, the cross-border component displays less variation across countries, suggesting lower levels of regulatory heterogeneity, relative to the domestic processing subindex, as reflected in a smaller standard deviation.

## 2.2 Sectoral elasticities

Because the effect of data-related restrictions is likely to be more pronounced in sectors relying more heavily on data-related services inputs, we employ a weighted approach of data restrictions using digital intensities. More specifically, we employ input-output tables to compute sectoral elasticities of digital intensities and multiply those with the two sub-indexes of data-related restrictions described above. In effect, this approach measures the extent to which each downstream sector in each country is reliant on digital inputs upstream and is computed as a proportion of each downstream total input usage.

We develop the sectoral elasticities by computing the input-output coefficients of the share of digital inputs from the three most data-intensive digital sectors—namely, computer, telecommunication, information services, and digital products—in the total input usage of each downstream sector. This measure serves as a proxy for the sensitivity of each sector to regulatory changes in data policy, as captured by the DFRI

The data is sourced from the OECD Inter-Country Input-Output tables (ICIO) tables and coefficients are computed across the full sample of European countries, following Borusyak, Hull, and Jaravel (2025). This approach captures cross-country variation in data dependency and enables identification based on heterogeneous exposure to common shocks. We further decompose the resulting sectoral elasticities into two components: one reflecting the share of digital inputs sourced domestically (domestic sectoral elasticities) and the other capturing the share of imported digital inputs (cross-border sectoral elasticities). The former is interacted with the domestic data restrictiveness subindex, while the latter is interacted with the cross-border data restrictiveness subindex. The sum of these two components constitutes the overall sectoral elasticity measure.

Table 3 reports both types of elasticities for each 1-digit sector in columns (2) and (3), respectively. Three key insights emerge from the data. First, service sectors exhibit higher elasticities than manufacturing sectors, indicating a greater reliance on data-related inputs sourced from upstream industries. This pattern is particularly pronounced in digitally deliverable services, such as financial and communication services. Second, across all sectors, elasticities are generally higher

with respect to domestic data inputs than to foreign ones, suggesting a greater sensitivity to domestic data regulations. Third, sectors including information and communication, finance, professional services, and wholesale and retail consistently rank highest in terms of both domestic and cross-border elasticities

### **2.3 ORBIS database**

The Orbis data includes both private and publicly listed firms. It includes raw firm-level financial files collected by Bureau van Dijk (BvD). The raw database comprises firm-level information from twenty-four European countries during the period 2010-2017.<sup>4</sup> We work with unconsolidated financial balance sheets, which better capture the activity of a firm in the country where it operates, to create the financial module. The financial and balance sheet information comes from national business registers, governed by country-specific legal and administrative filing requirements. We follow the approach in Kalemli-Özcan et al. (2024) to obtain a nationally representative firm-level dataset. Since this is not Census data, the approach ensures that firms are covering a large part of the aggregate economy and are representative in terms of the firm size distribution. We show this exercise for the manufacturing sector. First, we compute the ratio of the value of total output produced by firms in our sample relative to the value of total output from the official Eurostat SBS data, reported in Tables 4 and 5. Second, we conduct an additional validation exercise by comparing our data to the Eurostat-SBS in terms of the firm size distribution, reported in Tables 6 and 7.

We restrict our analysis to countries that have a sufficient number of country-year-sector-firm observations. The final Orbis dataset includes 14 countries<sup>5</sup> for 2010-2017, which is subsequently matched to the DFRI database, generating over 11,000,000 observations in total.

---

4. Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Finland, France, Germany, Hungary, Italy, Luxembourg, Montenegro, North Macedonia, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Serbia and Ukraine.

5. The countries are Belgium, Bulgaria, Croatia, Czech Republic, Finland, France, Germany, Hungary, Italy, Portugal, Romania, Slovakia, Slovenia, and Spain.

### 3 Empirical Methodology

The estimation strategy proceeds in two stages. The first stage focuses on estimating productivity by developing TFPR, while the second stage estimates labor demand.

#### 3.1 Productivity effects

The productivity variable to be estimated is revenue-based total factor productivity (TFPR). We estimate this measure instead of physical TFP (TFPQ) because ORBIS does not collect price information at the firm level. In order to construct proxy variables for output and inputs in comparable units across countries and over time, we use country deflators like the consumer price index to convert nominal variables into real ones. As a result, our measure of TFPR confounds variations in prices and efficiency at the firm level. It is therefore a measure of firm profitability.

To estimate TFPR, we first estimate a log-linearized Cobb-Douglas production function (PF), assuming that the PF elasticities vary at the 2-digit sector level. The estimation method follows Akerberg, Caves, and Frazer (2015), who rely on the control function approach (CFA) to deal with endogeneity of input choices. We use materials to make productivity observable. While the Akerberg, Caves, and Frazer (2015) method assumes an exogenous productivity process, we follow De Loecker (2013) and endogenize it.

Therefore, in our specification, TFPR is a function of the DFRI, lagged TFPR, and a set of country and time fixed effects. We do so, because assuming an exogenous TFPR process, by contrast, would have implied that the DFRI has no impact on efficiency or sales. This would have not only been unrealistic, but also would have invalidated the moment conditions needed to identify the production function elasticities, given that the productivity shock would not have been orthogonal to factor choices. In other words, if TFPR is a function of the DFRI, the PF elasticities will be biased. Since stringent data-related restrictions are expected to be factor-saving, lack of endogeneization of the Markov process would then lead to upward biased TFPR estimations.

After estimating the PF elasticities, we use equation 3.1 to estimate TFPR. Then, with unbiased

estimates of TFPR at the firm-level in hand, we pool all the observations and run an OLS regression of unbiased-TFPR on the variable of interest. The OLS coefficients are mathematically equivalent to the weighted average of the estimated coefficients obtained from the PF estimation, where the Markov coefficients vary at the sector-level (see section 2 on the Online Appendix for the proof).

Using the weighted average coefficients implies assuming homogeneous effects of the DFRI on TFPR instead of sector effects. We do this for two reasons. First, the type of data restriction we are analyzing is not sector-specific. But we allow it to have an heterogeneous marginal effect on TFPR depending on the sector, by multiplying the index for the sectoral elasticities, which measure the degree to which each sector is sensitive to data restrictions as captured by the DFRI. Second, by pooling all the observations we gain efficiency and increase the degrees of freedom in the estimation, especially with sectors that have few observations. Provided we focus the interpretation of the results (inference) on the entire sample, our approach eliminates imprecision coming from making estimations with small sub-samples.

As a result, our empirical strategy to estimate TFPR and the marginal effects of data restrictions is a three-step procedure. Step 1 and 2 are the standard Control Function Approach step, with the difference that we extend Akerberg, Caves, and Frazer (2015) and endogenize TFPR as a function of the DFRI as in De Loecker (2013). Step 3 recovers the marginal effects on TFPR at the firm-level. The following sub-sections provide further details about the specifications estimated in each stage.

### 3.1.1 TFPR estimation: CFA Step 1

We first estimate a log-linearized Cobb-Douglas production function at the sectoral level:

$$y_{ijct} = a_j + b_j l_{ijct} + c_j k_{ijct} + d_j m_{ijct} + tfpr_{ijct} + D_c + D_t + e_{ijct},^6 \quad (3.1)$$

where  $y_{ijct}$ ,  $l_{ijct}$ ,  $k_{ijct}$ , and  $m_{ijct}$  refer to output, labor, capital, and materials used by firm  $i$ , which operates in sector  $j$  of country  $c$ , at time  $t$ .  $e_{ijct}$  is an i.i.d error term that captures unanticipated

---

6. Henceforth  $x = \ln(X)$ , for  $X = \{Y, L, K, M, TFPR\}$ .

shocks to production or measurement error.  $D_c$  and  $D_t$  are country fixed-effects and time fixed-effects, respectively. Since productivity,  $TFPR_{ijct}$ , is unobservable, we follow Akerberg, Caves, and Frazer (2015) and use materials to make it observable:

$$m_{ijct} = h(l_{ijct}, k_{ijct}, tfpr_{ijct}, DFRI_{jct}, D_c, D_t), \quad (3.2)$$

Since materials are a strictly monotonic function of TFPR, we can invert function  $h(\cdot)$ , and express TFPR as a function of labor, capital, materials, and the DFRI.

$$tfpr_{ijct} = h^{-1}(l_{ijct}, k_{ijct}, m_{ijct}, DFRI_{jct}, D_c, D_t). \quad (3.3)$$

Inserting equation (3.3) into (3.1) yields:

$$\begin{aligned} y_{ijct} = & a_j + h^{-1}(l_{ijct}, k_{ijct}, m_{ijct}, DFRI_{jct}, X_{ijct}, D_c, D_t) \\ & + b_j l_{ijct} + c_j k_{ijct} + d_j m_{ijct} + D_c + D_t + e_{ijt}. \end{aligned} \quad (3.4)$$

Equation (3.4) can be estimated by OLS. We approximate function  $h(\cdot)$  using a third degree polynomial on labor, capital, and materials. Following Akerberg, Caves, and Frazer (2015), in the first step we cannot identify the coefficients of the PF. However, we can remove the estimated error term, and use output minus its predicted value to estimate the TFPR process and use the productivity shock for the moment conditions needed to estimate the PF elasticities.

We explore three specifications. The first one includes the aggregate DFRI as the main explanatory variable. However, given that the two sub-indexes covered in the aggregate DFRI variable are highly correlated, as countries with stringent domestic regimes are likely to also impose stringent cross-border restrictions, we decided to estimate their effect separately. The second specification therefore opens the DFRI into two sub-indexes, capturing the two categories of data-related restrictions: one that explores the cross-border dimension and one that covers the domestic dimension of data flows. We report the results from the regressions for the two sub-indexes separately.

### 3.1.2 TFPR estimation: CFA Step 2

As mentioned, the Akerberg, Caves, and Frazer (2015) CFA relies on an exogenous Markovian TFPR process to estimate the PF elasticities:

$$tfpr_{ijct} = g(tfpr_{ijct-1}) + \varepsilon_{ijct} \quad (3.5)$$

Following De Loecker (2013), the standard CFA can be extended by endogenizing TFPR as a function of the DFRI. Moreover, we adopt a flexible functional approach, which allows the DFRI marginal effect to vary with a firm's initial level of TFPR. Further, to deal with endogeneity concerns, we instrument the main explanatory variable(s) using the predictive value of the DFRI aggregate index, as well as the two sub-indexes of cross-border restrictions and domestic regulations, respectively, in the two respective estimations.

The extended specification nests other, more traditional approaches used in the literature such as OLS with fixed effects. This explains why De Loecker (2013), Braguinsky et al. (2015) and De Loecker et al. (2016) do not control for firm or plant fixed effects when endogenizing the Markov process. It is the timing assumption of the productivity shock coupled with the exogeneity of the data-related restrictions and the delay in the adoption of the technologies used to instrument the DFRI what ensures identification of the main effect.<sup>7</sup> The resulting estimation equation is:

$$tfpr_{ijct} = \alpha_j + \rho_{j1}tfpr_{ijct-1} + \rho_{j2}tfpr_{ijct-1}^2 + \rho_{j3}tfpr_{ijct-1}^3 + D_c + D_t \quad (3.6) \\ + \Psi(DFRI_{ijct}^{domestic}, DFRI_{ijct}^{cross-border}, tfpr_{ijct-1}) + \varepsilon_{ijct},$$

where  $\Psi$  is a function that includes the predicted value of  $DFRI_{ijct}$  as a free-standing variable(s),  $tfpr_{ijct-1}$ , and the interaction between both variables. The term  $\varepsilon_{ijct}$  is by assumption

---

7. Experimental evidence have evolved extensively, showing that low adoption rates can be explained by several factors. For technologies that do not display network effects, this includes preferences and behavioral bias (e.g., risk aversion, hyperbolic preferences, status-quo bias, and change-holding behavior), informational constraints and limitations to individual and social learning (e.g., lack of knowledge on how to use the new technology, uncertainty about its effectiveness, sustainability and returns, heterogeneous conditions, and ineffective knowledge transmission channels), weak demand (e.g., trade shocks, other demand shocks), and lack of incentives (e.g., principal-agent misalignment within the firm, and lack of product market competition).

uncorrelated with any lagged choice variable because the latter are in the firm’s information set. Given the timing of factor and input choices, the PF elasticities are estimated based on the following moment conditions: <sup>8</sup>

$$E \left[ \varepsilon_{ijct} (b_{jc}, c_{jc}, d_{jc}) \begin{pmatrix} l_{ijct-1} \\ k_{ijct} \\ m_{ijct-1} \end{pmatrix} \right] = 0. \quad (3.7)$$

In the current setting, it is the exogeneity of the policy shock coupled with the timing that it takes to endogenously change productivity, given its structural nature, what helps identify the main effect.

### 3.1.3 TFPR estimation: Step 3. Estimating global average digital upgrading premiums

With unbiased estimates of TFPR in hand, we pool all the observations and estimate equation (3.6) using OLS. Section D of Online Appendix shows that the estimated coefficients in the whole sample are a weighted average of the coefficients obtained across sub-samples.

## 3.2 Labor effects

Despite using a balanced panel dataset, our setting is static in nature and corresponds to partial equilibrium models. In our framework, data restrictiveness affects jobs through two different channels. A direct effect, which highly likely operates through trade in services, and an indirect effect that operates through productivity. Our framework thus disregards dynamic general equilibrium effects that emerge as a response of wages to changes in labor demand induced by changes in the

---

8. The Akerberg, Caves, and Frazer (2015) approach uses a value-added instead of output PF to estimate TFPR. It is intentionally done in this way to avoid estimating the elasticity corresponding to materials and therefore address the concern that lagged materials is not a valid instrument. Bond and Söderbom (2005) argue that materials are a flexible input, which implies that it does not follow an auto-regressive process. To explore this issue, we estimated an AR (1) model for materials and found that the coefficient of interest is equal to 0.86. We prefer this approach instead of the value-added approach, as the latter implicitly assumes an output elasticity with respect to materials equal to 1. The coefficients of the production functions are thus estimated by minimizing the sample analogue of equation (3.7) using GMM.

level of digital trade restrictiveness.<sup>9</sup> Given that stringent domestic regulatory constraints on data flows can enhance consumer trust and boost demand—and, in some cases, even require the recruitment of additional personnel—it is our premise that relaxing these constraints could, through this channel, reduce labor demand ( $\theta_2 > 0$ ); while a less restrictive cross-border regime will at the same time reduce jobs by increasing imports of services ( $\theta_3 > 0$ ). By contrast, a relaxation of regulatory constraints will make firms more productive and, *ceteris paribus*, increase labor demand ( $\theta_4 > 0$ ). To analyze the effect of data-related restrictions on labor demand, we estimate the following equation:

$$\Delta Labor_{ijct} = \theta_1 + \theta_2 \Delta DFR_{ijct}^{domestic} + \theta_3 \Delta DFR_{ijct}^{cross-border} + \theta_4 \Delta t f p r_{ijct} + D_{ct} + D_{jt} + v_{ijct}, \quad (3.8)$$

where  $\Delta Labor$  stands for changes in labor demand. Assuming a wage structure that has a fixed-term associated with workers’ skills, grade, performance, and characteristics of the local labor market, as well as a variable component related to inflation indexation, our specification controls for variations in wages, as it is unlikely that the fixed-term component will vary on an annual basis. This is because performance-driven wage upgrading and changes in local labor markets conditions often materialize over longer-term horizons. Wage indexation effects will be wiped out with the time fixed effects.

## 4 Results

This section presents semi-parametric estimates of the marginal effect of data-related restrictions on productivity and jobs.

---

9. Recent technological innovations have revamped an old concern about productivity-driven general equilibrium displacement effects on jobs, as well as shifts of the labor demand towards skilled workers. New task theories developed to understand the potential effects of automation on jobs depart from the skill-biased technological change models and show that the effect of technology-adoption on jobs is ambiguous. Under the new settings, robots compete against workers. Initially, machines replace workers in tasks previously performed by humans (Acemoglu and Restrepo 2018, 2019, 2020; Autor and Salomons 2018). However, as the economy grows, new tasks are introduced. Dynamically, in general equilibrium, the initial labor displacement effect pushes wages down and allows labor to regain a price advantage relative to machines. As a result, the new tasks are labor intensive. This effect is known as the reinstatement effect.

## 4.1 Productivity results

Figure C.1a displays the kernel distribution for the estimated marginal effect of cross-border data restrictions. As the figure shows, the marginal effect is negative for 53.5 percent of the sample, while it is positive for the rest. There are at least three channels through which stringent restrictions affect revenue-based productivity. The first one operates through a reduction in exports that decreases the domestic price of the goods sold nationally. The second channel functions through a contraction in imports that causes the opposite price effect, as it reallocates domestic demand towards domestic producers. The final mechanism materializes through a decline in imports of cheaper intermediate inputs that makes firms less productive.

Figure C.1b displays the kernel distribution for the estimated marginal effect of domestic data restrictions. As this second figure shows, the marginal effect is negative for 54.8 percent of the sample, while it is positive for the rest. Two opposite forces operate on prices and physical productivity when domestic regulatory restrictions are more stringent. One is the increase in transactions costs, which reduces the willingness to pay for a good or service and, therefore, its final price. Thus, domestic data restrictions make domestic firms less productive and less attractive. The other one is an increase in consumers' trust, due to data protection, which affects prices and TFPR in a contrary direction.

Table 8 displays the median marginal effect on productivity. As predicted by the theory, a one percentage point increase in data-related restrictions halves productivity for the average firm. Cross-border regulations reduce TFPR by 24.1 percent, while domestic ones by 23.8 percent. Figure C.2a shows that the median marginal cross-border effect varies along the productivity distribution. It is positive for firms located at the 50 percentile of the TFPR distribution, i.e. the very middle of the set of firms covered, or lower percentiles; while it is negative for the rest of firms. Furthermore, cross-border data restrictions tax disproportionately more high-productivity firms.

While our analysis does not allow us to identify the underlying mechanisms that could explain the heterogeneity of the main effect, it is our conjecture that stringent cross-border data regimes tend to favor unproductive firms by restricting imports, while constraining exports for more pro-

ductive firms. However, the reverse is observed for domestic data restrictions, which increase the productivity of top performers but leaves bottom performers behind, as shown in Figure C.2b, thereby widening the productivity gap between leaders and laggards. The behavior of the net marginal effect resembles that of the cross-border effect, as it is this margin that dominates. Overall, the unequal effect of data-related regulatory restrictions creates a trade-off between aggregate real gains and inequality for the case of domestic restrictions. While data-related regulatory liberalization boosts aggregate productivity, it also enlarges the performance gap between leaders and the rest.

## **4.2 Sector-specific results**

Table 9 presents summary statistics of the marginal effects across sectors. The median marginal effect of cross-border data restrictions on productivity is negative and high for digitally deliverable services, including financial, ICT, professional, and administrative services. This is consistent with the fact that mandating companies to keep data within the national boundaries or imposing additional requirements for data to be transferred abroad reduces imports of these services (Ferracane and van der Marel 2021) which, in turn, reduces competition from foreign service providers and productivity. The median marginal effect of cross-border restrictions on productivity is non-negative for traditional services and the manufacturing sector, except for construction.

The opposite is observed for domestic data restrictions where the median marginal effect is positive for firms operating in digitally deliverable services sectors. This may reflect an expansion of more productive domestic firms that are more likely to be able to meet the requirements imposed on firms to access, store, process data, or more generally make any commercial use of data within a certain jurisdiction. However, the median marginal effect is negative for firms operating in traditional services, such as transportation and trade services. Even though these sectors are measured as less data-intensive, they nonetheless are required to comply with the regulation, given that all firms are required to implement domestic regulatory requirements, which hampers their productivity performance. The median marginal effect of domestic data restrictions is also negative for

manufacturing enterprises. This can be attributed, at least in part, to the fact that traditional services, such as transportation and distribution, are important upstream inputs with cascading effects on the productivity of downstream manufacturing activity (Arnold, Kiel, and Voigt 2016).

### **4.3 Labor results**

Table 10 presents the results from estimating the effect of changes in data restrictions on labor demand changes, controlling for endogenous policy restrictions-induced changes in total factor productivity. Our findings shows that a one percentage point increase in domestic data restrictions increases, as expected, labor demand, probably by restricting imports of services. However, cross-border data restrictions do not seem to affect jobs, while reductions in firm-level productivity caused by stringent policy regimes reduce jobs. This is highly likely due to increases in the transaction cost of selling goods and services, which reduce the willingness to pay for the good or service, while demand contraction leads to a reduction in labor demand.

## **5 Simulating productivity gains from the liberalization of data flows**

With the estimated coefficients at hand, this section quantifies the productivity gains that firms can obtain when governments relax regulatory constraints for data, taking year 2017 as the baseline for the exercise. Given that the level of regulatory harmonization is relatively high in a region like Europe, we focus our analysis on a full liberalization scenario instead of taking the countries to the regional frontier.

We conduct two counterfactual exercises. In the first one, a non-targeted scenario, where we universally liberalize data restrictions. In the second one, a targeted scenario, where we liberalize data restrictions for firms that display positive marginal effects. Although this scenario is unrealistic because reforms like such are difficult to be implemented, we want to present the results to compare them with the universal reform, where there are winners and losers.

Table 11 presents summary statistics of the firm-level productivity gains from the universal reform. With the exception of Belgium, the rest of the countries exhibit on average a positive productivity gain that ranges from 0.314 percent for a country like Spain and up to 2.423 percent for France. By contrast, half of the economies in the sample display, on average, productivity losses from liberalization of domestic data restrictions. The largest positive gains are observed for Romania (0.83 percent) and Bulgaria (0.769).

Table 12 presents summary statistics of the firm-level productivity gains from the targeted reform. For several economies, the cross-border gains are larger than the domestic ones, similar to the previous scenario. For liberalization of cross-border restrictions, the gains range from a low of 0.091 percent for Belgium and up to 3.834 percent for Finland. Gains from liberalizing domestic restrictions range from 0.0641 for Belgium and up to 2.190 for Finland.

## 6 Conclusions

This paper estimates the effect of cross-border and domestic data restrictions on productivity and jobs and quantifies the productivity gains from liberalizing digital trade. The paper shows that on average both types of regulatory restrictions decrease productivity. However, the effects vary along the productivity distribution. While cross-border restrictions tax disproportionately more high-productivity firms, domestic restrictions do so for low-productivity companies. There are interesting differences across sectors when it comes to the effect of data restrictions, depending on their scope. Cross-border restrictions have a negative median marginal effect for the majority of the sectors, but a positive one for manufacturing ones. While the opposite is observed for domestic restrictions. So service firms benefit the most, in productivity terms, from liberalizing cross-border data flows, while manufacturing companies experience the largest gains with liberalization of domestic data restrictions. Finally, while the effect of cross-border restrictions on labor demand is statistically insignificant, domestic restrictions are associated with higher labor demand, as restrictions shield domestic firms from foreign competition.

## References

- Acemoglu, Daron, and Pascual Restrepo. 2018. “The race between man and machine: implications of technology for growth, factor shares, and employment.” *American Economic Review* 108 (6): 1488–1542.
- . 2019. “Demographics and automation.” *National Bureau of Economic Research Working Paper No.* 24421.
- . 2020. “Robots and jobs: evidence from US labor markets.” *Journal of Political Economy* 128 (6): 2188–2244.
- Akerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. “Identification properties of recent production function estimators.” *Econometrica* 83 (6): 2411–2451.
- Akcigit, Ufuk, and Sina T Ates. 2023. “What happened to US business dynamism?” *Journal of Political Economy* 131 (8): 2059–2124.
- Andrews, Dan, Chiara Criscuolo, Peter Gal, Carlo Menon, et al. 2015. “Firm Dynamics and Public Policy: Evidence from OECD Countries— Conference–2015.”
- Arnold, Christian, Daniel Kiel, and Kai-Ingo Voigt. 2016. “How the industrial internet of things changes business models in different manufacturing industries.” *International Journal of Innovation Management* 20 (08): 1640015.
- Atkin, David, Azam Chaudhry, Shamyra Chaudry, Amit K Khandelwal, and Eric Verhoogen. 2015. “Markup and cost dispersion across firms: Direct evidence from producer surveys in pakistan.” *American Economic Review* 105 (5): 537–544.
- Autor, David H, and Anna Salomons. 2018. “Is automation labor share-displacing? Productivity growth, employment, and the labor share.” *The Brookings Institution* 49 (1): 1–87.

- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. “Americans do IT better: US multinationals and the productivity miracle.” *American Economic Review* 102 (1): 167–201.
- Bond, Stephen, and Måns Söderbom. 2005. “Adjustment costs and the identification of Cobb Douglas production functions.” *Institute for Fiscal Studies IFS Working Papers* W05/04.
- Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2025. “A Practical Guide to Shift-Share Instruments.” *Journal of Economic Perspectives* 39, no. 1 (February): 181–204.
- Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson. 2015. “Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry.” *The American Economic Review* 105 (7): 2086–2119.
- Brynjolfsson, Erik, Xiang Hui, and Meng Liu. 2019. “Does machine translation affect international trade? Evidence from a large digital platform.” *Management Science* 65 (12): 5449–5460.
- Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. “Generative AI at work.” *The Quarterly Journal of Economics*, qjae044.
- Bureau van Dijk. n.d. *Orbis [Company financial database]*. Orbis. Retrieved from <https://www.bvdinfo.com>.
- Cabral, Luis, and Ali Hortacsu. 2010. “The dynamics of seller reputation: Evidence from eBay.” *The journal of industrial economics* 58 (1): 54–78.
- Cory, Nigel. 2017. *Cross-border data flows: Where are the barriers, and what do they cost?* Technical report. Information Technology and Innovation Foundation.
- Cory, Nigel, and Luke Dascoli. 2021. *How barriers to cross-border data flows are spreading globally, what they cost, and how to address them*. Technical report. Information Technology and Innovation Foundation.

- Cusolito, Ana Paula, Alvaro Garcia-Marin, and William F Maloney. 2023. “Proximity to the frontier, markups, and the response of innovation to foreign competition: Evidence from matched production-innovation surveys in Chile.” *American Economic Review: Insights* 5 (1): 35–53.
- David, Joel M, Venky Venkateswaran, Ana Paula Cusolito, and Tatiana Didier. 2021. “Capital allocation in developing countries.” *The World Bank Economic Review* 35 (4): 1102–1121.
- De Loecker, Jan. 2013. “Detecting learning by exporting.” *American Economic Journal: Microeconomics* 5 (3): 1–21.
- De Loecker, Jan, Pinelopi Goldberg, Amit Khandelwal, and Nina Pavcnik. 2016. “Prices, Markups, and Trade Reform.” *Econometrica* 84 (2): 445–510.
- Demirer, Mert, Diego J Jiménez Hernández, Dean Li, and Sida Peng. 2024. “Data, privacy laws and firm production: Evidence from the GDPR.” *NBER Working Paper* 32146.
- DeStefano, Timothy, Richard Kneller, and Jonathan Timmis. 2023. “Cloud computing and firm growth.” *Review of Economics and Statistics*, 1–47.
- Eslava, Marcela, Marcela Meléndez, Gabriel Ulyssea, Nicolás Urdaneta, and Ignacio Flores. 2024. “Firms and Inequality in Latin America.” LACIR III Working Paper 138.
- Ferracane, M., J. Kren, and E. van der Marel. 2020. “Do Data Policy Restrictions Impact the Productivity Performance of Firms and Industries?” *Review of International Economics* 28 (3): 676–722.
- Ferracane, M.F., and E. van der Marel. 2021. “Do Data Flows Restrictions Inhibit Trade in Services?” *Review of World Economics* 157 (4): 727–776.
- Frey, Carl Benedikt, and Giorgio Presidente. 2024. “Privacy regulation and firm performance: Estimating the GDPR effect globally.” *Economic Inquiry* 62 (3): 1074–1089.

- Gal, Peter, Giuseppe Nicoletti, Theodore Renault, Stéphane Sorbe, and Christina Timiliotis. 2019. “Digitalisation and productivity: in search of the holy grail—firm-level empirical evidence from EU countries.” *OECD Economics Department Working Paper* No. 1533.
- Goldfarb, Avi, and Daniel Trefler. 2018. “AI and international trade.” *NBER Working Paper* 24254.
- Graetz, Georg, and Guy Michaels. 2018. “Robots at work.” *Review of economics and statistics* 100 (5): 753–768.
- Johnson, Garrett. 2022. “Economic research on privacy regulation: Lessons from the GDPR and beyond.”
- Kalemli-Özcan, Şebnem, Bent E Sørensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcen Yeşiltaş. 2024. “How to construct nationally representative firm-level data from the orbis global database: New facts on SMEs and aggregate implications for industry concentration.” *American Economic Journal: Macroeconomics* 16 (2): 353–374.
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. “Human decisions and machine predictions.” *The quarterly journal of economics* 133 (1): 237–293.
- Lucking-Reiley, David, Doug Bryan, Naghi Prasad, and Daniel Reeves. 2007. “Pennies from eBay: The determinants of price in online auctions.” *The journal of industrial economics* 55 (2): 223–233.
- Melnik, Mikhail I, and James Alm. 2002. “Does a seller’s ecommerce reputation matter? Evidence from eBay auctions.” *The journal of industrial economics* 50 (3): 337–349.
- Mullainathan, Sendhil, and Ziad Obermeyer. 2022. “Diagnosing physician error: A machine learning approach to low-value health care.” *The Quarterly Journal of Economics* 137 (2): 679–727.

Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer. 2022. “Regulatory spillovers and data governance: Evidence from the GDPR.” *Marketing Science* 41 (4): 746–768.

Sinclair, Joseph T, and Stanley Livingston. 2005. *eBay photography the smart way: creating great product pictures that will attract higher bids and sell your items faster*. AMACOM/American Management Association.

Sun, Ruiqi, and Daniel Treffer. 2023. *The impact of AI and cross-border data regulation on international trade in digital services: A large language model*. Technical report. National Bureau of Economic Research.

## A Historical Financial Data Cleaning Procedure

Following Bureau van Dijk (n.d.), David et al. (2021), and Kalemlı-Özcan et al. (2024), we document the steps we apply to clean the financial information.

1. *Fill time-invariant data gaps*: for a given BvD.ID-year combination, with BvD.ID standing for firm unique identifier, replace missing highly likely time-invariant information with information available for previous years (e.g., US SIC code, NAICS, NACE, NACE main sector, company name, city, region, postal code, legal form, incorporation date). To perform this step, the team first worked with auxiliary raw tables, which collect legal and sectoral information of the firm, and collapsed the time-invariant variables at the BvD.ID level.
2. *Harmonize timeframe*: convert variable closedate from string to numeric format. Then create a new variable, name it year, and assign a year to the observation according to the following rule. If closing month corresponding to the observation is June or any other month after June, then make Year take the year reported in closedate. Otherwise, make Year the year reported in closedate minus 1.
3. *Drop duplicates*: the raw database presents a large number of duplicates at the BvD.ID-year level. The team noticed that the information was the same, except in the SIC primary code variable. Thus, we collapsed all the SIC primary codes reported by the same BvD.ID-year in one variable, using semicolons to list all the SIC primary codes, and eliminated duplicates.
4. *Drop firms with missing relevant information*: drop all the firms with no information for the following set of variables: US SIC code, NAICS, NACE core code, NACE main sector.
5. *Drop observations with missing information for the currency code*: eliminate observations with missing information for the currency code.
6. *Drop observations with missing information for variable closedate*: eliminate observations with missing information for the close date of the financial statement.

7. *Drop observations with relevant missing information* eliminate observations that at the BvD.ID-year level have missing information in all the following variables: operating revenue (turnover), sales, employment, total assets.
8. *Drop duplicates and keep most updated information*: keep observations with the most recent closing date if there are duplicates at the BvD.ID-year-first letter of consolidation code (e.g., C, U) level.
9. *Drop duplicates and keep information from annual reports*: keep observations with annual report in *Use FillingType* variable if there are still duplicates and keep the standardized information. Using annual reports (IFRS preferred, instead of local reports) guarantees standardization of reporting protocol at international level.
10. *Eliminate firms with noisy data*: drop all the observations corresponding to a specific BvD.ID if any of the following variables has a negative value in a specific year – total fixed assets, tangible fixed assets, intangible fixed assets, other fixed assets, current assets, sales, and employment.
11. *Deflate values*: use country GDP deflators from the World Bank database to deflate nominal variables and set year 2005 as the base year.<sup>10</sup>
12. *Harmonize currencies*: convert values in local currency to USD dollars, using the average of the monthly exchange rate for year 2005.

---

10. <https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS>.

## B Tables

Table 1: Data-related restrictions and regulations

Restrictions cross-border on data flows
Is transfer of data prohibited or is there a local processing requirement?
Is there a local storage requirement?
Are there conditions to data flow?
Domestic regulations on data
<i>Subject rights on data privacy</i>
Request of consent for the collection and use of data?
Right to be forgotten?
<i>Administrative requirements on data privacy</i>
Perform an impact assessment (DPIA)?
Appoint a data protection officer (DPO)?
Notify the authorities in case of a data breach?
Allow the government to access personal data collected?
<i>Sanctions for non-compliance</i>
Are there fines?
If yes, up to what amount (in USD)?
Jailtime?
<i>Other</i>
Are there other instances in which data policies represent a barrier for digital trade?
<i>Data retention</i>
Is there a minimum period for data retention?
Is there a maximum period for data retention?

Table 2: Cross-border and Domestic Data Flow Restrictiveness Index (DFRI)

<b>Variable</b>	<b>Obs</b>	<b>Mean</b>	<b>Std. dev.</b>	<b>Min</b>	<b>Max</b>
DA_CB	11,049,716	0.1531	0.0421	0.125	0.375
DA_DR	11,049,716	0.2006	0.0684	0.060	0.385

Table 3: Data-sensitive Elasticities

<b>NACE main section</b>	<b>iocIMP</b>	<b>iocDOM</b>
C - Manufacturing	0.0089	0.0163
D - Electricity	0.0083	0.0312
E - Water supply	0.0101	0.0357
F - Construction	0.0067	0.0139
G - Wholesale and retail trade	0.0177	0.0459
H - Transportation	0.0143	0.0515
I - Accommodation	0.0081	0.0276
J - Information and communication	0.1015	0.2214
K - Financial and insurance	0.0354	0.1127
L - Real estate	0.0096	0.0349
M - Professional services	0.0269	0.0645
N - Administrative services	0.0127	0.0382
<b>Total</b>	<b>0.0169</b>	<b>0.0428</b>

*Note:* Please refer to the narrative in the report for a description of the method used to calculate elasticities.

Table 4: Coverage of the Manufacturing Sector (2010–2017)

Year	AT	BE	BA	BG	HR	CZ	EE	FI	FR	DE	HU	IT	LV
2010	0.02	1.03	–	0.79	1.07	0.86	0.50	0.63	0.71	0.58	0.83	0.90	0.07
2011	0.05	1.02	0.81	0.83	1.11	0.93	0.50	0.64	0.63	0.60	0.83	1.03	0.07
2012	0.27	1.01	0.56	0.84	1.13	0.93	0.48	0.64	0.51	0.60	0.86	0.99	0.06
2013	0.53	0.98	0.83	0.83	1.09	0.96	0.48	0.62	0.62	0.60	0.88	1.00	0.06
2014	0.54	0.97	0.78	0.81	1.08	0.92	0.48	0.65	0.77	0.59	0.85	1.02	0.07
2015	0.58	1.03	0.78	0.80	1.06	0.99	0.45	0.60	0.78	0.56	0.99	1.01	0.07
2016	0.61	0.95	0.73	0.79	0.99	0.94	0.46	0.61	0.81	0.54	0.88	1.01	0.07
2017	0.58	0.95	0.53	0.77	0.94	0.72	0.43	0.69	0.76	0.46	0.82	0.98	0.07
Avg (TFP)	0.40	0.99	0.72	0.81	1.06	0.91	0.47	0.64	0.70	0.57	0.86	0.99	0.07
Avg (ALL)	0.59	0.94	0.67	0.78	0.97	0.87	0.69	0.90	0.90	0.58	0.95	0.94	0.83
Year	LU	NL	MK	NO	PL	PT	RO	RS	SK	SI	ES	SE	CH
2010	0.83	0.01	–	0.02	0.42	1.10	0.78	–	0.72	1.03	1.00	0.39	0.00
2011	0.79	–	–	0.02	0.29	1.12	0.80	–	0.88	1.03	1.00	0.36	0.00
2012	0.71	–	1.00	0.02	0.22	1.13	0.84	–	0.92	1.03	1.03	0.33	0.00
2013	0.73	–	1.07	0.00	0.14	1.13	0.80	–	0.94	1.00	1.04	0.34	0.00
2014	0.70	–	1.01	0.02	0.12	1.13	0.80	–	0.95	0.99	1.05	0.36	0.00
2015	0.37	0.00	0.99	0.90	0.11	1.14	0.76	–	0.98	1.00	1.05	0.43	0.00
2016	0.34	–	–	0.99	0.29	1.09	0.76	0.79	0.99	0.95	1.06	0.40	0.00
2017	0.26	–	–	0.93	0.44	1.08	0.75	0.80	0.00	0.97	1.01	0.39	0.00
Avg (TFP)	0.59	0.01	1.02	0.36	0.25	1.11	0.79	0.79	0.80	1.00	1.03	0.37	0.00
Avg (ALL)	0.69	0.33	1.05	0.85	0.86	1.04	0.73	0.69	0.88	0.95	0.95	0.73	0.00

Notes: AT = Austria, BE = Belgium, BA = Bosnia and Herzegovina, BG = Bulgaria, HR = Croatia, CZ = Czech Republic, EE = Estonia, FI = Finland, FR = France, DE = Germany, HU = Hungary, IT = Italy, LV = Latvia, LU = Luxembourg, NL = Netherlands, MK = North Macedonia, NO = Norway, PL = Poland, PT = Portugal, RO = Romania, RS = Serbia, SK = Slovakia, SI = Slovenia, ES = Spain, SE = Sweden, CH = Switzerland. TFP = Average coverage where TFP data available (TFP sample). ALL = Average coverage.

Table 5: European Coverage of the Manufacturing Sector

Year	European Unweighted Average	European GDP-weighted Average
2010	0.62	0.65
2011	0.67	0.69
2012	0.67	0.67
2013	0.69	0.69
2014	0.69	0.71
2015	0.70	0.69
2016	0.71	0.73
2017	0.64	0.69

Notes: Average coverage where TFP data available (TFP sample).

Table 6: Firm Size Distribution: 2010

Size Class	Panel A: Based on TFP																					
	AT	BE	BG	HR	CZ	EE	FI	FR	DE	HU	IT	LV	NL	NO	PL	PT	RO	SK	SI	ES	SE	CH
<i>Orbis</i>																						
1 to 19 employees	0.03	0.03	0.07	0.12	0.05	0.15	0.06	0.04	0.00	0.01	0.10	0.05	0.02	0.21	0.01	0.12	0.07	0.09	0.12	0.10	0.17	NaN
20 to 249 employees	0.25	0.28	0.36	0.36	0.31	0.48	0.27	0.22	0.15	0.22	0.43	0.46	0.64	0.62	0.32	0.43	0.33	0.36	0.35	0.36	0.39	0.11
250+ employees	0.72	0.69	0.58	0.52	0.64	0.38	0.68	0.73	0.85	0.77	0.47	0.49	0.34	0.17	0.67	0.45	0.60	0.56	0.53	0.54	0.44	0.89
<i>Eurostat SBS</i>																						
1 to 19 employees	0.07	0.09	0.08	0.16	0.10	0.13	0.07	0.11	0.06	0.07	0.21	0.12	0.10	0.11	0.08	0.15	0.08	0.08	0.13	0.14	0.09	0.08
20 to 249 employees	0.34	0.28	0.33	0.35	0.28	0.55	0.20	0.28	0.22	0.20	0.40	0.57	0.34	0.38	0.26	0.45	0.30	0.21	0.35	0.36	0.28	0.35
250+ employees	0.59	0.64	0.59	0.49	0.62	0.32	0.73	0.61	0.72	0.73	0.39	0.31	0.57	0.52	0.66	0.40	0.62	0.71	0.52	0.50	0.63	0.57
Panel B: Based on Employment																						
<i>Orbis</i>																						
1 to 19 employees	0.00	0.02	0.07	0.16	0.06	0.22	0.11	0.10	0.00	0.01	0.13	0.15	0.00	0.24	0.01	0.25	0.12	0.08	0.17	0.23	0.29	NaN
20 to 249 employees	0.23	0.41	0.57	0.42	0.42	0.55	0.40	0.32	0.20	0.35	0.47	0.39	0.52	0.54	0.41	0.53	0.41	0.41	0.39	0.44	0.42	0.27
250+ employees	0.77	0.57	0.35	0.42	0.52	0.23	0.49	0.58	0.80	0.64	0.40	0.45	0.48	0.22	0.58	0.22	0.47	0.51	0.45	0.33	0.29	0.73
<i>Eurostat SBS</i>																						
1 to 19 employees	0.16	0.19	0.19	0.25	0.22	0.21	0.17	0.22	0.16	0.21	0.40	0.24	0.24	0.19	0.21	0.32	0.15	0.26	0.21	0.32	0.19	0.16
20 to 249 employees	0.37	0.38	0.49	0.38	0.37	0.55	0.34	0.34	0.32	0.37	0.37	0.53	0.44	0.43	0.38	0.49	0.41	0.31	0.38	0.41	0.33	0.43
250+ employees	0.47	0.43	0.32	0.37	0.41	0.24	0.50	0.44	0.52	0.43	0.23	0.23	0.32	0.38	0.41	0.18	0.44	0.43	0.41	0.28	0.48	0.40

Note: Country codes: AT = Austria, BE = Belgium, BG = Bulgaria, HR = Croatia, CZ = Czech Republic, EE = Estonia, FI = Finland, FR = France, DE = Germany, HU = Hungary, IT = Italy, LV = Latvia, NL = Netherlands, NO = Norway, PL = Poland, PT = Portugal, RO = Romania, SK = Slovakia, SI = Slovenia, ES = Spain, SE = Sweden, CH = Switzerland. ‘—’ indicates data not available. Average coverage where TFP data available (TFP sample).

Table 7: Firm Size Distribution: 2017

Size Category	Panel A: Based on Gross Output																							
	AT	BE	BA	BG	HR	CZ	EE	FI	FR	DE	HU	IT	LV	LU	MK	NO	PL	PT	RO	RS	SK	SI	ES	SE
<i>Orbis</i>																								
1–19 employees	0.00	0.01	0.12	0.14	0.17	0.04	0.22	0.11	0.03	0.00	0.01	0.21	0.21	0.01	0.17	0.14	0.04	0.21	0.11	0.15	0.01	0.17	0.20	0.26
20–249 employees	0.22	0.40	0.45	0.50	0.43	0.37	0.56	0.39	0.31	0.11	0.28	0.47	0.31	0.40	0.43	0.49	0.42	0.53	0.38	0.39	0.10	0.38	0.45	0.47
250+ employees	0.78	0.59	0.43	0.36	0.40	0.59	0.23	0.50	0.65	0.89	0.72	0.32	0.49	0.59	0.41	0.37	0.54	0.26	0.51	0.46	0.89	0.44	0.35	0.27
<i>Eurostat SBS</i>																								
1–19 employees	0.14	0.19	0.21	0.19	0.23	0.19	0.21	0.17	0.17	0.12	0.17	0.36	0.25	—	0.21	0.19	0.20	0.28	0.13	0.16	0.23	0.23	0.27	0.17
20–249 employees	0.36	0.37	0.44	0.46	0.38	0.34	0.54	0.36	0.24	0.32	0.33	0.37	0.54	0.36	0.42	0.44	0.35	0.50	0.38	0.36	0.31	0.36	0.42	0.35
250+ employees	0.50	0.44	0.35	0.35	0.39	0.46	0.25	0.48	0.59	0.55	0.50	0.27	0.21	0.56	0.38	0.37	0.45	0.22	0.49	0.47	0.46	0.41	0.31	0.48
Panel B: Based on Employment																								
<i>Orbis</i>																								
1–19 employees	0.00	0.01	0.12	0.14	0.17	0.04	0.22	0.11	0.03	0.00	0.01	0.21	0.21	0.01	0.17	0.14	0.04	0.21	0.11	0.15	0.01	0.17	0.20	0.26
20–249 employees	0.22	0.40	0.45	0.50	0.43	0.37	0.56	0.39	0.31	0.11	0.28	0.47	0.31	0.40	0.43	0.49	0.42	0.53	0.38	0.39	0.10	0.38	0.45	0.47
250+ employees	0.78	0.59	0.43	0.36	0.40	0.59	0.23	0.50	0.65	0.89	0.72	0.32	0.49	0.59	0.41	0.37	0.54	0.26	0.51	0.46	0.89	0.44	0.35	0.27
<i>Eurostat SBS</i>																								
1–19 employees	0.14	0.19	0.21	0.19	0.23	0.19	0.21	0.17	0.17	0.12	0.17	0.36	0.25	—	0.21	0.19	0.20	0.28	0.13	0.16	0.23	0.23	0.27	0.17
20–249 employees	0.36	0.37	0.44	0.46	0.38	0.34	0.54	0.36	0.24	0.32	0.33	0.37	0.54	0.36	0.42	0.44	0.35	0.50	0.38	0.36	0.31	0.36	0.42	0.35
250+ employees	0.50	0.44	0.35	0.35	0.39	0.46	0.25	0.48	0.59	0.55	0.50	0.27	0.21	0.56	0.38	0.37	0.45	0.22	0.49	0.47	0.46	0.41	0.31	0.48

Note: Country codes: AT = Austria, BE = Belgium, BG = Bulgaria, HR = Croatia, CZ = Czech Republic, EE = Estonia, FI = Finland, FR = France, DE = Germany, HU = Hungary, IT = Italy, LV = Latvia, NL = Netherlands, NO = Norway, PL = Poland, PT = Portugal, RO = Romania, SK = Slovakia, SI = Slovenia, ES = Spain, SE = Sweden, CH = Switzerland. ‘—’ indicates data not available. Average coverage where TFP data available (TFP sample).

Table 8: Estimated Productivity Marginal Effect from Cross-Border and Domestic Data Restrictions

Model	N	Median	SD	Min	Max	% over 0	F-value	p-value
Cross-border DTR	7,855,418	-0.2416	4.2398	-35.0459	22.5884	46.5	901.3849	0.000
Domestic DTR	7,855,418	-0.2379	2.3841	-14.1664	19.2279	45.2	887.8304	0.000
Full model							854.2705	0.000

*Note:* This table presents the estimated median marginal effect of data flow restrictions on productivity based on the estimation of equation (3.6).

Table 9: Summary Statistics: Estimated Marginal Effect of Cross-border and Domestic DFRI on Productivity

NACE Main Section	Cross-border DFRI				Domestic DFRI			
	Median	SD	Min	Max	Median	SD	Min	Max
C - Manufacturing	0.1794	3.4357	-21.8865	20.6167	-0.4243	2.1239	-12.9088	13.8473
D - Electricity	8.8683	2.9952	-20.3750	17.9445	-5.9945	1.8571	-11.5643	12.6328
E - Water supply	0.9962	4.8950	-19.4484	4.5044	-1.1299	3.0263	-3.2665	11.3661
F - Construction	-0.1743	1.0894	-16.8346	6.4820	-0.2505	0.6806	-4.4941	10.4647
G - Wholesale	2.1389	4.4118	-28.9664	22.5884	-1.8916	2.6553	-14.1664	17.0116
H - Transportation	4.7171	5.7204	-35.0459	12.3368	-3.3974	3.5904	-8.0632	19.2279
I - Accommodation	1.6110	1.9259	-19.7605	9.3827	-1.4612	1.1984	-6.2524	11.5022
J - Information	-8.0065	4.0497	-29.6281	5.5615	0.8735	1.2282	-3.9507	6.3613
K - Financial	-7.9703	3.1785	-18.0430	1.5030	3.8916	1.9779	-1.9674	9.9351
L - Real estate	-3.8797	1.7535	-15.7518	13.9377	1.8775	1.0300	-8.9458	9.1466
M - Professional	-3.7598	2.4227	-21.1221	17.5778	1.4202	1.2330	-11.4536	11.1955
N - Administrative	-6.9000	3.3582	-34.6076	5.4204	3.4786	1.9065	-3.8350	18.4289
<b>Total</b>	-0.2416	4.2398	-35.0459	22.5884	-0.2379	2.3841	-14.1664	19.2279

*Note:* This table presents the results from estimating equation (3.6). The statistics include the estimated median marginal effects of both cross-border and domestic DFRI on productivity across NACE main sectors.

Table 10: Estimated Marginal Effect of Changes in DFRI on Changes in Jobs

Variable	Beta Hat	Std. Error	t-ratio
Cross-border DFRI	-0.09288	0.19644	-0.4728
Domestic DFRI	0.19756	0.06517	3.0317
TFPR	0.02529	0.00102	24.7277
<b>N</b>		5,634,827	
<b>R<sup>2</sup></b>		0.2988	

*Note:* This table presents results from estimating equation (3.8).

Table 11: Summary Statistics of Firm-Level Productivity Gains from Fully Liberalizing Data Flow Restrictions (in percentage terms)

<b>Country</b>	<b>Mean</b>	<b>Median</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
<i>Cross-border DFRIs</i>					
Belgium	-1.16	-1.33	0.81	-2.26	1.26
Bulgaria	0.32	0.02	2.76	-6.60	21.99
Croatia	0.25	-0.06	1.68	-3.94	13.50
Czech Republic	0.18	-0.08	2.38	-4.42	15.18
Finland	2.42	0.11	6.44	-1.93	55.53
France	1.80	0.42	6.19	-2.07	87.80
Hungary	1.10	1.04	0.24	0.35	2.19
Italy	1.27	0.07	4.06	-1.90	37.15
Portugal	0.32	-0.02	1.50	-1.23	19.72
Romania	0.15	-0.31	2.85	-15.34	17.81
Slovenia	0.72	0.02	3.19	-2.10	24.55
Spain	0.31	0.00	1.36	-2.12	17.87
<b>Total</b>	0.70	0.02	3.22	-15.34	87.80
<i>Domestic DFRIs</i>					
Belgium	0.60	0.70	0.35	-0.94	1.00
Bulgaria	0.77	0.15	2.45	-8.27	10.38
Croatia	0.47	0.36	1.13	-4.70	7.44
Czech Republic	0.27	0.07	0.94	-4.73	3.42
Finland	-0.37	-0.04	3.54	-16.71	6.77
France	-0.65	-0.09	3.40	-34.24	9.91
Hungary	-1.16	-1.08	0.32	-2.65	-0.15
Italy	-0.99	-0.04	3.47	-35.81	20.25
Portugal	-0.06	0.09	1.79	-16.20	2.71
Romania	0.83	0.36	2.06	-7.73	14.34
Slovenia	0.01	0.01	1.50	-12.04	3.67
Spain	-0.10	0.09	2.23	-17.50	18.76
<b>Total</b>	-0.22	0.07	2.77	-35.81	20.25

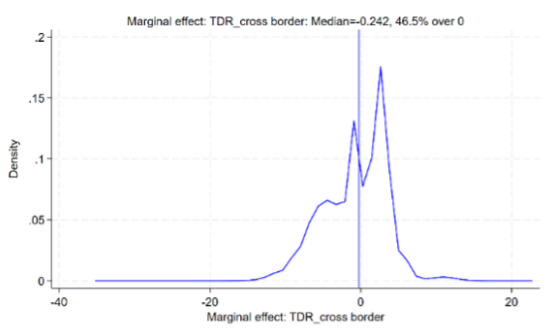
*Note:* This table presents summary statistics (mean, median, standard deviation, minimum, and maximum) of firm-level productivity gains from fully liberalizing data flow restrictions across 2017.

Table 12: Summary Statistics for Targeted Scenario

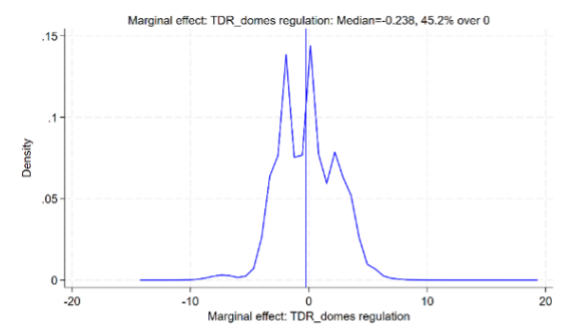
<b>Country</b>	<b>Mean</b>	<b>Median</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
<i>Cross-border DFRIs</i>					
Belgium	0.091	0.062	0.142	0.0016	1.255
Bulgaria	1.916	1.525	2.775	0.000004	21.988
Croatia	1.263	0.657	2.100	0.000026	13.498
Czech Republic	1.583	0.658	3.020	0.000032	15.181
Finland	3.834	0.145	7.283	0.000403	55.529
France	3.613	1.354	7.616	0.000849	87.803
Hungary	1.099	1.042	0.238	0.355	2.195
Italy	2.216	0.555	4.890	0.000002	37.154
Portugal	1.365	1.297	1.794	0.000001	19.725
Romania	2.501	2.351	2.120	0.000051	17.810
Slovenia	1.794	0.938	3.922	0.000034	24.555
Spain	0.891	0.640	1.726	0.0000002	17.873
<b>Total</b>	1.887	0.782	3.952	0.0000002	87.803
<i>Domestic DFRIs</i>					
Belgium	0.641	0.711	0.320	0.000713	0.998
Bulgaria	1.856	0.605	2.464	0.000094	10.383
Croatia	1.051	0.662	0.912	0.000022	7.435
Czech Republic	0.756	0.760	0.736	0.000031	3.424
Finland	2.190	1.918	1.864	0.000041	6.774
France	2.055	2.486	1.375	0.000151	9.906
Italy	1.462	1.313	1.157	0.000003	20.245
Portugal	1.053	0.885	0.864	0.000013	2.705
Romania	1.904	1.258	1.797	0.000019	14.339
Slovenia	0.882	0.436	0.894	0.000012	3.671
Spain	1.290	0.848	1.347	0.000004	18.756
<b>Total</b>	1.446	1.033	1.449	0.000003	20.245

*Note:* This table presents summary statistics (mean, median, standard deviation, minimum, and maximum) of firm-level productivity gains from fully liberalizing data flow restrictions only for firms that display negative marginal effects. We use 2017 as the baseline year.

# C Figures

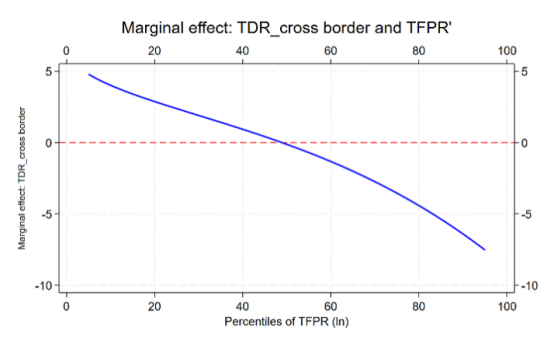


(a) Cross-border DFRIs lower productivity for slightly more than half of the firms

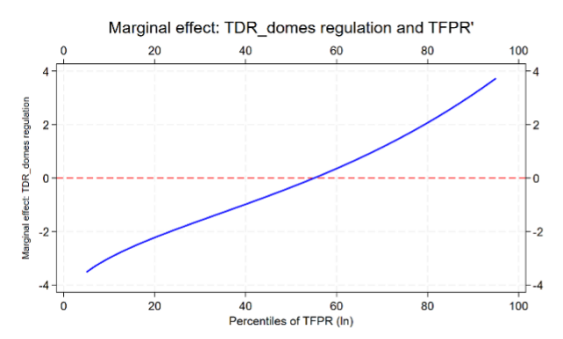


(b) Domestic DFRIs lower productivity for slightly more than half of the firms

Figure C.1: Kernel distribution for estimated marginal effects



(a) Cross-border DFRIs tax disproportionately more high-productivity firms



(b) Domestic DFRIs widen the productivity gap between leaders and laggards

Figure C.2: Marginal effects by level of productivity

## D Proof: Estimates of Homogeneous Coefficients Reflect Sectoral-Weighted Averages of Estimates of Heterogeneous Coefficients

Suppose we have the following GMM estimator:

$$\hat{\theta} = \underbrace{\left[ \left( \sum_{i=1}^N X_i' Z_i \right) A \left( \sum_{i=1}^N Z_i' X_i \right) \right]^{-1}}_{\Psi} \underbrace{\left[ \left( \sum_{i=1}^N X_i' Z_i \right) A \left( \sum_{i=1}^N X_i' y_i \right) \right]}_{\Phi}, \quad (\text{D.1})$$

where  $X$  is a  $K \times N$  matrix of regressors,  $Z$  is a  $Q \times N$  matrix of instruments,  $A$  is a  $K \times K$  weighting matrix, and  $y$  is a  $1 \times N$  vector.

The first bracket is a matrix of dimension  $K \times K$ , which is constructed based on a sample of size  $N$ :

$$\left[ \left( \sum_{i=1}^N X_i' Z_i \right) A \left( \sum_{i=1}^N Z_i' X_i \right) \right]^{-1} = \kappa_1 \underbrace{\left[ \left( \sum_{i=1}^{N_1} X_i' Z_i \right) A_1 \left( \sum_{i=1}^{N_1} Z_i' X_i \right) \right]^{-1}}_{\Psi_1} \quad (\text{D.2})$$

$$\left[ \left( \sum_{i=1}^N X_i' Z_i \right) A \left( \sum_{i=1}^N Z_i' X_i \right) \right]^{-1} = \kappa_2 \underbrace{\left[ \left( \sum_{N_1+1=1}^N X_i' Z_i \right) A_2 \left( \sum_{N_1+1=1}^N Z_i' X_i \right) \right]^{-1}}_{\Psi_2} \quad (\text{D.3})$$

$N$ -size sample can be divided into two samples of size  $N_1$  and  $N_2$ , where  $N = N_1 + N_2$ .  $\kappa_1$  and  $\kappa_2$  are matrices mapping each component in  $\Psi$  to each component in  $\Psi_1$  and in  $\Psi_2$ .

The second component,  $\Phi$ , is additive. Therefore, it can be written as the sum of the two

components corresponding to the two sub-samples:

$$\left[ \left( \sum_{i=1}^N X_i' Z_i \right) A \left( \sum_{i=1}^N X_i' y_i \right) \right] = \underbrace{\left[ \left( \sum_{i=1}^{N_1} X_i' Z_i \right) A_1 \left( \sum_{i=1}^{N_1} X_i' y_i \right) \right]}_{\Phi_1} + \underbrace{\left[ \left( \sum_{N_1+1=1}^N X_i' Z_i \right) A_2 \left( \sum_{N_1+1=1}^N X_i' y_i \right) \right]}_{\Phi_2}. \quad (\text{D.4})$$

Replacing D.2 - D.4 into D.1, we can write  $\hat{\theta}$  in the following way:

$$\hat{\theta} = \kappa_1 \Psi_1 \Phi_1 + \kappa_2 \Psi_2 \Phi_2.$$

That is, the GMM estimator corresponding to the full sample can be written as a weighted sum of the two estimators corresponding to subsamples 1 and 2.