OECD publishing

RIGHT HERE, RIGHT NOW? NEW EVIDENCE ON THE ECONOMIC EFFECTS OF SERVICES TRADE REFORM

OECD TRADE
POLICY PAPER
April 2023 **n°271**

OECD
BETTER POLICIES FOR BETTER LIVES

OECD TRADE AND AGRICULTURE DIRECTORATE

Right here, right now? New Evidence on the Economic Effects of Services Trade Reform

Sebastian Benz, Matteo Fiorini, Alexander Jaax, and Elisabeth van Lieshout

This paper provides evidence on the "when, how and where" of the effects of service trade policy reforms, discussing short-term impacts on services trade as well as on the performance of downstream manufacturing industries. A combination of novel methodological approaches is used to be able to track impacts over time and along the supply chain. The OECD Services Trade Restrictiveness Index serves as the measure for trade policy reform. Results show that reducing policy barriers to services trade can increase services imports already in the short run, and that benefits continue to grow over time. The impact of services trade reforms may still vary significantly depending on the nature of the policy change, the economic context, and the targeted mode of services supply. Finally, services trade reforms can have sizeable spillover effects on the productivity of manufacturing sectors that use services as intermediate inputs.

Keywords: Services trade policy; services imports; short-run gravity; synthetic control method; manufacturing productivity.

JEL codes: F13; F14; F61

Acknowledgement:

The authors would like to thank Andrea Andrenelli, Cemre Balaban, Charles Cadestin, John Drummond, Janos Ferencz, Frédéric Gonzales, Javier López González, Marion Jansen, Eleonora Mavroeidi, Sébastien Miroudot, Annabelle Mourougane, Ana-Maria Muresan, Julia Nielson, Silvia Sorescu, Frank van Tongeren, Enxhi Tresa, and Nicolas Woloszko for fruitful comments and suggestions. The excellent statistical support provided by Charles Cadestin, Charles Christmann, Frédéric Gonzales and Ana-Maria Muresan is gratefully acknowledged. The authors are particularly indebted to Annabelle Mourougane for her inspiring guidance during the design phase of this project. The work presented in Annex C benefitted from discussions at the OECD Study Group on Causal Inference and Machine Learning co-ordinated by the OECD Trade and Agriculture Directorate and the OECD New Approaches to Economic Challenges (NAEC) Innovation Lab. The authors are grateful to Stefan Wager for supervising the study group and providing guidance for the application presented in Annex C. This paper also benefitted from discussions in the OECD Working Party of the Trade Committee, which agreed to make it more widely available through declassification. The authors thank Laëtitia Christophe, Derya Isir, and Michèle Patterson for preparing this paper for publication.

Table of Contents

Figures

TABLES

BOX

Key messages

Background

The OECD Services Trade Restrictiveness Index (STRI) is a comprehensive database, recording applied services policies and policy changes in 50 economies and 22 sectors since 2014. The rich information included in this database allows for a quantitative assessment of potential economic impacts of changes in the regulatory framework for services trade.

Going beyond existing studies, this report uses the OECD STRI and combines different methodological approaches to provide evidence on the “when, how and where” of the effects of service trade policy reforms, discussing short-term impacts on services trade as well as on the performance of downstream manufacturing industries.

Using the OECD STRI to investigate the effects of services trade reforms does not come without caveats. The main STRI indicators, for example, capture most-favoured-nation regulation without taking into account commitments made in regional trade agreement. Moreover, the time series dimension in the STRI database is still relatively short. Continuous efforts to extend and update the OECD STRI database are key to strengthen the body of evidence on services trade reforms and their effects on economic outcomes.

Findings

- Reducing policy barriers can raise services trade flows already in the short run, while additional effects continue to accrue over the longer run. While services reforms cutting 0.05 from a country’s STRI are associated with a short-run growth in services trade of less than 10% in all sectors, in the medium to long term trade values can increase by 20% to 50%, depending on the sector.
- The effect in the first years immediately after a reform is thus likely to be relatively small. Capturing the full effects of services reform requires adopting a long-run perspective, with additional impact accruing as production and investment decisions adjust to the new regulatory framework.
- The trade effect of services reforms can vary depending on their content and the existing economic environment. The magnitude of a reform’s impact and the pace at which it materialises depend on the policy instrument targeted by that particular reform. Moreover, a specific change in policy can have differential trade impacts across sectors and countries. The trade impact of reforms may also be different depending on the mode of services supply. Case studies of reform episodes in different countries illustrate considerable heterogeneity in the effects on services trade and suggest that other factors and policies can play a role in shaping the impact of services reforms.
- Downstream industries can benefit from policy reforms targeting upstream services sectors. Removing barriers to services trade imports has economically sizable spillover effects on the productivity of domestic manufacturing sectors.
- On average across the 17 manufacturing industries included in the analysis, simulated reforms of services trade policy – equivalent to a 0.05 reduction in the STRI score – are estimated to increase downstream manufacturing productivity by 8.4% (air transport reform scenario), 6.5% (telecommunications), and 2.3% (financial services). These findings reflect the essential role of services as intermediate inputs in production processes and provide a strong motive for a holistic approach when assessing the impact of services trade reforms.

1. Introduction

Services are transforming the global economy on a massive scale. The vast majority of workers are employed in the services sector. According to the World Bank World Development Indicators, in 2020 value added from services accounted for 65% of total output and 44% of gross exports globally. The introduction of new services changes consumption patterns, but it also has impacts on the production of goods. Services such as transport or logistics are the "glue" that holds value chains together (Low, 2013[1]). At the same time, R&D, distribution or marketing services represent important production stages in the manufacturing process themselves (Miroudot and Cadestin, 2017[2]). Overall, around 30% of the value added in global manufacturing exports comes from services sectors.

In the past years, countries have made significant endeavours towards reform of the regulatory framework for services sectors. These endeavours include domestic reforms that liberalise services trade unilaterally, which have been particularly ambitious in non-OECD countries (OECD, 2021[3]), but also plurilateral initiatives at the WTO, including the Joint Statement Initiatives on Services Domestic Regulation, on E-Commerce, and on Investment Facilitation for Development. In addition, services play an increasingly important role in bilateral negotiations, shown by the increasing number of regional trade agreements (RTAs) with services provisions and the emergence of digital economy agreements.

Notwithstanding these efforts, barriers to services trade remain high. In several important sectors, cross-border services trade stands at around 0.5% to 1.5% of what it could potentially be in a hypothetical, completely frictionless scenario (Benz, Jaax and Yotov, 2022[4]).[1] Regulatory restrictions to services trade are particularly prominent in non-OECD countries (Benz and Jaax, 2020[5]). Efforts towards the quantification of regulatory services trade barriers have long been hampered by the absence of adequate data.

The OECD Services Trade Restrictiveness Index (STRI) has filled this gap since its launch in 2014. The STRI is a comprehensive database that records applied services policies. Being updated annually and currently available from 2014 to 2021, it can be an important tool for the assessment of services trade reforms. It currently covers 50 economies, of which 38 are members of the OECD. The OECD STRI builds on earlier work related to the quantification of services trade policies by the World Bank (Borchert, Gootiiz and Mattoo, 2013[6]) and the Australian Productivity Commission (Findlay and Warren, 1990[7]).

Measuring barriers to trade in services entails methodological challenges and some caveats apply. The OECD STRI is a composite index on a scale from zero to one. It is calculated through a codified algorithm from the OECD STRI database. While binary scoring is applied to individual measures, there are hierarchies and inter-dependencies implemented in the scoring algorithm. The relative importance of different policy areas in each sector is determined on the basis of weights collected from an expert survey. The STRI is also a measure of MFN restrictions and does not take into account any specific commitments made in regional trade agreements or the existence of mutual recognition agreements (Geloso Grosso et al., 2015[8]). Answers in the STRI database are based on de jure rules, written down in the laws and regulations of each country. Hence, it may be an imperfect measure of *de facto* restrictiveness perceived by services firms.

Notwithstanding methodological caveats, the rich information in the STRI database has proven extremely valuable for empirical studies aimed at enhancing our understanding of the costs of regulatory barriers to services trade. Restrictions as measured in the STRI have been shown to hamper cross-border services trade as well as services exports via commercial presence (Nordås and Rouzet, 2015[9]; Rouzet, Benz and Spinelli, 2017[10]; Benz, 2017[11]; Benz and Jaax, 2020[5]; Khachaturian and Oliver, 2021[12]; Reverdy, 2022[13]). In addition to current regulatory barriers, uncertainty about future trade policy changes negatively affects services trade (Ahmad et al., 2020[14]). Restrictive regulations undermine competition, enabling shielded firms to charge higher markups (Rouzet and Spinelli, 2016[15]).

[1] In this frictionless trade scenario, borders are no longer an obstacle to trade and the costs of cross-border trade are the same as the costs of trade within national markets.

While important nuances – such as the disproportionate effect of trade barriers on small and medium-sized enterprises (Rouzet, Benz and Spinelli, 2017[10]) – were explored in existing STRI-based studies, several questions related to the “when, how and where” of the effects of service trade policy reforms remain underexplored. In particular, the time horizon over which a change in regulatory policies should be expected to affect trade flows remains unclear. Empirical studies seeking to evaluate impacts of trade policy changes often rely on the gravity model, but standard gravity regressions do not tell us how long it takes for a policy change to affect trade flows. Rather than focusing on the “when”, gravity models provide overall effects via a comparison of average outcomes in all observed years before and after the policy change.

Assumptions that firms and workers easily and instantaneously adjust to policy changes have increasingly been questioned in recent years (Dix-Carneiro and Kovak, 2017[16]; Anderson, Larch and Yotov, 2020[17]). A clearer picture of the time it takes for trade policy changes to change economic outcomes would support efforts to mitigate adjustment costs and optimise the coordination of relevant policies, e.g. with respect to trade facilitation and training measures.

Regarding the question where a services trade policy change is expected to affect economic outcomes, the existing STRI-based evidence clearly shows that a more restrictive regulatory framework for a given sector has a negative effect on trade flows in that specific sector (Nordås and Rouzet, 2015[9]; Benz, 2017[11]; Benz, Rouzet and Spinelli, 2020[18]; Benz and Jaax, 2022[19]; Reverdy, 2022[13]). Yet, services act as crucial inputs to manufacturing industries and play a key role in global value chains (Miroudot and Cadestin, 2017[2]; Cadestin and Miroudot, 2020[20]). This implies that the restrictions recorded in the STRI database are – despite its focus on services trade – likely to be of relevance to the economic performance of downstream manufacturing industries relying on inputs from upstream services sectors.

Combining three complementary methods, this paper draws on the OECD STRI database to investigate the effects of services trade policy changes over time and with respect to downstream manufacturing industries. First, an innovative structural gravity framework (Anderson and Yotov, 2020[21]) is used to distinguish between short run and long run effects of regulatory changes on cross-border services trade flows. Rather than assuming that exporters can immediately adjust to changes in trade barriers, this approach emphasises the need for investments in destination-specific capacity as a prerequisite for the expansion of exports.

The second analysis employs the synthetic control method (Abadie and Gardeazabal, 2003[22]) to explore how concrete examples of policy changes in four instances have affected services imports of the corresponding countries. Its focus on individual cases complements the aggregate perspective of most empirical work on the effects of trade restrictions, highlighting underlying heterogeneity in outcomes. Focusing on cross-border services trade as recorded in the balance of payments, the analysis compares two cases of liberalising reforms as well as two examples of restrictive policy changes. It also reports exploratory findings related to the impact on Mode 3 services imports, relying on data from the OECD AMNE database (Cadestin et al., 2018[23]).

The third analysis explores the relevance of services trade barriers to the performance of manufacturing industries. Panel regressions at the country-sector-year level are used to investigate how the regulatory framework of five services sectors (air transport, finance and insurance services, computer services, telecommunications, logistics) affects the labour productivity of 17 downstream manufacturing industries. Adopting an empirical strategy that has been used in related contributions (e.g. focused on product market regulation, Bourlès et al., 2013[18]), this analysis is inspired by contributions highlighting the pivotal role of services as agents supporting the upgrading of manufacturing activities and the coordination of global value chains (Nordås and Kim, 2013[24]; Miroudot and Cadestin, 2017[2]).

The gravity analysis and the investigation of spillovers from services trade reforms on downstream manufacturing industries illustrate the size of reform effects using a reduction of the STRI by 0.05 as a hypothetical reform scenario. While such a lowering of the STRI score can only be achieved through a combination of several regulatory changes, different bundles of reforms can lead to this reduction.[2]

[2] Regarding air transport, a 0.05 reduction in the STRI score roughly corresponds, for example, to the combination of the following three reforms: First, removing a 49% foreign equity cap for one type of traffic (e.g. domestic cargo) and

Main challenges for the robustness of all three methods include the limited country coverage of the OECD STRI and other data sets, including bilateral services trade data and sector-level productivity data. The analysis does not cover all modes of services trade and it does not consider services value added embodied in merchandise trade. Moreover, sector-level services trade data is not very granular, only allowing to study relatively broadly defined services sectors.

Three key results emerge from this work. First, reducing services barriers can increase cross-border services trade flows already in the short run, while additional effects continue to accrue over the longer term. Capturing the full effects of services reform requires adjustment of production and investment decisions in line with the new policies. Second, there are various dimensions of heterogeneity in the effect of services trade policy on services trade, including over time, across sectors, regarding different economic environments and regarding the direction of reform. Overall, heterogeneity emerges as an important feature characterising trade-responses to services trade reforms. Third, industries downstream the supply chain can benefit from policy reforms targeting upstream services sectors. Removing barriers to services trade imports has economically sizable spillover effects on the productivity of domestic manufacturing sectors.

The rest of the paper is structured as follows. Section 2 presents results of the gravity analysis focused on the distinction of short run and long run effects of services policies on trade. Section 3 discusses insights from the investigation of policy changes in three countries based on the synthetic control method. Section 4 examines the effects of services trade reforms on downstream manufacturing industries. Section 5 concludes.

2. The short-run impact of services policies on trade

2.1. Motivation

What is the impact of services trade policies on trade? Gravity models have been successfully providing answers to this question (Anderson et al., 2018[25]; Benz and Jaax, 2022[19]; Borchert and Di Ubaldo, 2021[26]; Reverdy, 2022[13]). But just how long does it take until the impact of a liberalising services reform can be observed in trade statistics? So far, there is no satisfactory answer.

The reason is simple: standard gravity models use observable determinants of trade cost, such as geographic distance, common language, and regional trade agreements (RTAs), to explain the value of bilateral trade flows. Results from these models indicate the extent to which a common language or the conclusion of an RTA is associated with higher values of bilateral exports and imports. However, these standard models are mostly silent about the time it takes for policy to deliver effects. Using the implementation of an RTA as an example, the gravity model compares bilateral trade flows before the policy change with bilateral trade flows after the policy change, controlling for confounding factors. If the model identifies a trade-promoting effect of the RTA, this effect might materialise in any year between implementation and the end of the observation period.

The impact of trade policies over time can be assessed by using lags and leads of trade policy variables, in addition to the contemporaneous variable. For merchandise trade, results suggest that the significant impact of an RTA on bilateral trade flows has fully materialised between 10-15 years after entry into force of the RTA (Baier and Bergstrand, 2007[27]; Egger, Larch and Yotov, 2021[28]). This result does not come as a surprise, given that many RTAs use so called phase-in periods for the introduction of commitments, such as new tariff schedules for manufactured goods.

While these contributions provide useful insights on economic effects in the long term, they remain silent on short-run effects. Even without phase-in periods of trade policy reforms, there are many reasons why changes in trade policy changes might not have an immediate effect on bilateral trade values. The aggregate trade effect is driven by the decisions of numerous businesses, including large multinationals,

allowing up to 100% foreign equity share. Second, removing for cargo as well as passenger traffic any limits to the proportion of shares that can be acquired by foreign investors in publicly controlled firms. Third, removing labour market tests for intra-corporate transferees, contractual service suppliers, and independent services suppliers.

as well as small and medium-sized enterprises (SMEs). In other words, an instantaneous impact on aggregate trade values would hinge upon businesses' capacity to react quickly to liberalising policy changes and optimally adjust their operations to the new situation.

However, this reaction might require recruitment of new staff or training of existing staff to exploit the new business opportunity created from liberalisation in a foreign market.[3] Moreover, it might hinge on an increase in the capacity for digital communication, such as installation of new hardware. It might also entail business travel to the foreign market, which could be limited because airlines are not able to anticipate increased demand.[4] In addition, network effects and learning across firms could imply that some firms start exporting only after a sufficient number of exporters is already present in a market. Such network effects could be related to information regarding the identification of buyers, foreign prices, market selection, or standards and testing requirements (Roberts and Tybout, 1997[29]).[5] The lack of knowledge and the existence of network effects might also limit the number of firms that are able to source from foreign providers, leading to constraints on the importer side. All this suggests that the adjustment of trade values after a services policy change is unlikely to happen right away.

2.2. Estimation framework

The analysis in this section is based on recent progress regarding the modelling of bilateral trade flows in the short run (Anderson and Yotov, 2020[21]). The study relies on the idea of frictions regarding the adjustment of trade. Technically, this idea builds on the assumption that bilateral trading capacities adjust to a long-run equilibrium when bilateral trade costs remain unchanged. Consequently, trade volumes in the long run are identical compared to a standard gravity model without trading capacities.

However, trading capacities cannot be raised in the short-run, even if regulatory liberalisation makes bilateral trade less costly.[6] Trade liberalisation will still lead to an increase in bilateral exports. However, the impact is smaller than in a situation where there is no comparable capacity constraint. Comparing results from a traditional gravity model and a gravity model with capacity constraint, it can be shown that the impact of RTAs and tariff reductions on manufacturing trade in the short run is around one quarter of its long-run impact.[7] The adjustment and expansion of trading capacities across country pairs raised world manufacturing trade by 75% between 1988 and 2006 (Anderson and Yotov, 2020[21]). To the best of our knowledge, there exist no comparable estimates for services trade so far.

The analysis described in this section replicates main components of the Anderson and Yotov (2020[21]) model. Instead of manufacturing trade, this study focuses on bilateral services trade flows in five sectors: business services, communications services, financial services, insurance and transport.

The gravity model used in this analysis is based on data for international services trade, as well as domestic consumption of domestically produced services, also referred to as "domestic trade". This framework can identify the impact of country-specific policy variables even when controlling for country-specific trade costs ("multilateral resistance") using exporter-year and importer-year fixed effects. The model also encompasses year-specific border dummies to control for an overall trend towards services globalisation

---

[3] Recruitment of managers with export experience in previous jobs is positively associated with the overall probability of exporting (Sala and Yalcin, 2014[74]). Specific market knowledge can double the probability of successful exports to that market (Mion, Opromolla and Sforza, 2022[76])

[4] Depending on the sector, a 10% expansion of bilateral air travellers is associated with growth of exports by between 1% and 3% (Benz, Jaax and Yotov, 2022[4]).

[5] Such network effects can be explicit modelled in a trade model with heterogeneous firms (Krautheim, 2012[75])

[6] This idea is also consistent with network dynamics (Chaney, 2014[77]) or managerial experience in a specific market (Mion and Opromolla, 2014[78]).

[7] The short run is defined as the period during which trading capacities do not adjust. In reality, trading capacities depend on different types of constraints, which potentially adjust at varying speed. Therefore, this model is silent about a specific time horizon for the short run.

that is not driven by policy factors, but other determinants, including growth of air traffic and ICT adoption (Benz, Jaax and Yotov, 2022[4]).[8]

Bilateral capacities for exports and imports cannot be easily captured in a standard gravity model. A main challenge is that there is no data for bilateral trade capacities. In addition, there is no fully developed theory regarding firms' investment in and expansion of bilateral trade capacities.

These challenges can be addressed through a structural approach to economic dynamics (Lucas and Prescott, 1971[30]). The methodology uses the observable distribution of sales across different export destinations as proxies for unobserved bilateral trading capacities. In the gravity analysis, this boils down to adding lagged bilateral exports as additional explanatory variable in the regression. This approach is based on the idea of a long-run equilibrium, in which bilateral export capacity is on a level that is efficient in the long run for a given value of bilateral trade flows. Trade policy changes lead to deviations from this equilibrium. Capacity allocation is assumed to move back toward the long run efficient level over time.[9] Hence, export shares from the previous period can be used as a proxy for the unobservable capacity based on the efficient allocation theory.

This approach does not only provide estimates for the importance of capacity constraints for bilateral trade flows, but it also can quantify the impact of services policy changes on services trade flows in the short run.[10] All other variables are included contemporaneously, i.e. data on the determinants of trade flows, including those related to policy indicators, refer to the same year in which a trade flow is recorded.

Main results are derived from this model using a specification in which exporter-year and importer-year fixed effects control for each country's multilateral resistance to trade, while standard gravity variables, such as bilateral distance, contiguity, or common language, control for time-invariant country-pair-specific trade cost. A potential downside of this approach is the risk that these variables might not capture all determinants of bilateral trade costs, due to the existence of unobservable time-invariant trade costs shaping trade relations of a specific country pair. For example, these might be related to differences in consumer preferences across countries. Results from a specification using symmetric pair-specific fixed effects are reported in Annex A as robustness check.[11]

2.3. Data and description of variables

The study uses data for the period between 2014 and 2019.[12] Trade flows between the 45 countries included in the sample jointly accounted for roughly two thirds of global trade in these five sectors in 2019.[13]

All trade data refer to cross-border services exports according to the definition in the balance of payments. The balance of payments measures the value of transactions between a resident and a non-resident

[8] These year-specific border dummies equal one if the corresponding trade flow is international and occurs in a specific year.

[9] In the absence of adjustment cost it could also move back instantaneously. However, this is a special case that is not relevant for the estimation of the model.

[10] Technical details of the estimation framework are described in Annex A.

[11] This specification only exploits changes in trade policies over time for the identification of trade effects. Such variation is limited due to the relatively short time coverage of the dataset used in this chapter. Hence, it is not surprising that some of the regression coefficients in this robustness check are not statistically different from zero (see Annex A for more details).

[12] The coverage is mostly determined by the availability of data, with 2019 being the most recent year available at the time of writing.

[13] More specifically, bilateral trade flows recorded in the database used for the analysis accounted for 60% of global exports of communications services in 2019, for 65% in the case of financial services, 73% for insurance services, 61% of transport services, and 70% regarding business services.

institutional unit.[14] In the context of services, this includes Mode 1 (cross-border services trade), Mode 2 (consumption abroad) and Mode 4 (movement of people). A disaggregation is possible by the type of service included in the transaction, allowing to categorise Mode 2 as travel services, which is excluded from the analysis. Throughout the rest of the document, “cross-border services trade” is used in reference to the value of services trade measured in the balance of payments.

This section does not analyse Mode 3 services trade, recorded in foreign affiliate trade statistics (FATS). Therefore, it can only provide an incomplete picture of the impact of policy reform on services trade. However, existing cross-country evidence using a gravity framework suggests that the association between the OECD STRI and bilateral Mode 3 services trade is similar to the STRI’s association with cross-border services trade flows (Benz and Jaax, 2020[5]).

Data on international cross-border services trade come from the OECD international trade in services statistics and the OECD-WTO Balanced Trade in Services dataset (BaTIS). As customary in the related literature, information on bilateral exports is the main data source. In cases where exports are not reported at a sufficiently disaggregate level, these data are complemented by import values reported by a country’s trading partners.

The export variable corresponds to within-country trade when the exporter is the same country as the importing country. Within-country trade is defined as the value of gross services production in the corresponding sector that is consumed domestically. Data for within-country trade are constructed by deducting a country’s sector-level exports to the world (taken from the OECD EBOPS 2010 database) from gross production in that sector, using data from the OECD national accounts and OECD STAN. Information on the sectoral coverage, including definition and correspondence across data sources is reported in Annex A.

Other data used for this analysis include regional trade agreements (RTAs) from the WTO. An RTA is coded as a services RTA if it includes a legally enforceable services provision, as defined in the World Bank Deep RTAs database (Mattoo, Rocha and Ruta, 2020[31]).[15] All other RTAs are coded as goods RTAs, controlling for the potentially trade-creating effect of services exports linked to merchandise trade. Data for standard gravity variables comes from Centre d’Etudes Prospectives et d’Informations (CEPII).

The OECD STRI and the intra-EEA STRI are employed as measures of services trade restrictiveness. The STRI is a measure of most favoured nation (MFN) restrictions and does not take into account any specific concessions such as regional trade agreements or existence of mutual recognition agreements (Geloso Grosso et al., 2015[8]). The intra-EEA STRI database allows for the measurement of services trade restrictiveness within the preferential regime of the EEA, characterised by the freedoms of the internal market (free movement of goods, people, services and capital) and a harmonisation of rules in areas such as competition policy or regulatory transparency (Benz and Gonzales, 2019[32]). This distinction is important when looking at bilateral services trade between a specific pair of countries.

Bilateral services trade data and the OECD STRI both have a limited country coverage. The sample used for analysis in this section encompasses 45 countries, including 37 OECD countries[16] plus Brazil, The People’s Republic of China (hereafter “China”), Indonesia, India, Malaysia, The Russian Federation (hereafter “Russia”), Thailand, and South Africa.

[14] The residence of each institutional unit is the economic territory with which it has the strongest connection, expressed as its centre of predominant economic interest. A household is resident in a territory where it has been present or intends to be present for one year or more (International Monetary Fund, 2009[79]).

[15] A services provision is defined as any provision covering services. Sectoral heterogeneity regarding the coverage of specific services in individual RTAs is not taken into account.

[16] Costa Rica is not covered due to limited availability of data.

2.4. Results

This section describes the main results of the short-run gravity analysis and uses results from a standard gravity model as benchmark. Results displayed in this section focus on the impact of key policy variables on bilateral trade flows, including the STRI and services RTAs. A full set of results, including regression coefficients of control variables, is reported in Annex A.

Benchmarking short-run gravity estimates against a standard gravity model indicates a relatively smaller impact of services trade policies in the short run in almost all specifications. This result suggests that only a proportion of the overall trade-creating impact of services reform materialises in the short term. A period of adjustment is required before full benefits of a services liberalisation can materialise. For manufacturing trade, most of the adjustment happens within ten years after a policy reform (Baier and Bergstrand, 2007[27]; Anderson and Yotov, 2020[21]; Egger, Larch and Yotov, 2021[28]), providing a tentative indication for the meaning of the term "long-run" effect.[17]

Results in this section are based on a specification where exporter-year and importer-year fixed effects absorb each country's multilateral resistance to trade, while observable gravity variables such as bilateral distance, contiguity and common language to control for bilateral trade costs between two countries. This approach can exploit all variation in the applied measures of services trade policies, including variation across countries.[18]

The impact of domestic services regulation on bilateral trade flows is shown graphically in Figure 2.1. The chart uses diamonds to indicate the increase in cross-border services trade associated with a reduction of the STRI by 0.05 points.[19] The chart applies upwards-pointing arrows to depict 90% confidence intervals. Results from the short-run gravity model are contrasted against equivalent results from a standard gravity analysis.

Regression coefficients of the STRI variable in the standard gravity analysis are all negative and highly statistically significant. All results are in line with existing estimates from the literature (Benz and Jaax, 2020[5]). Short-run gravity coefficients are also all negative and statistically significant, but much smaller in absolute terms.[20]

[17] Efforts towards identifying the precise time horizon for the long-term effects of services liberalisation are hampered by the limited time coverage of the STRI.

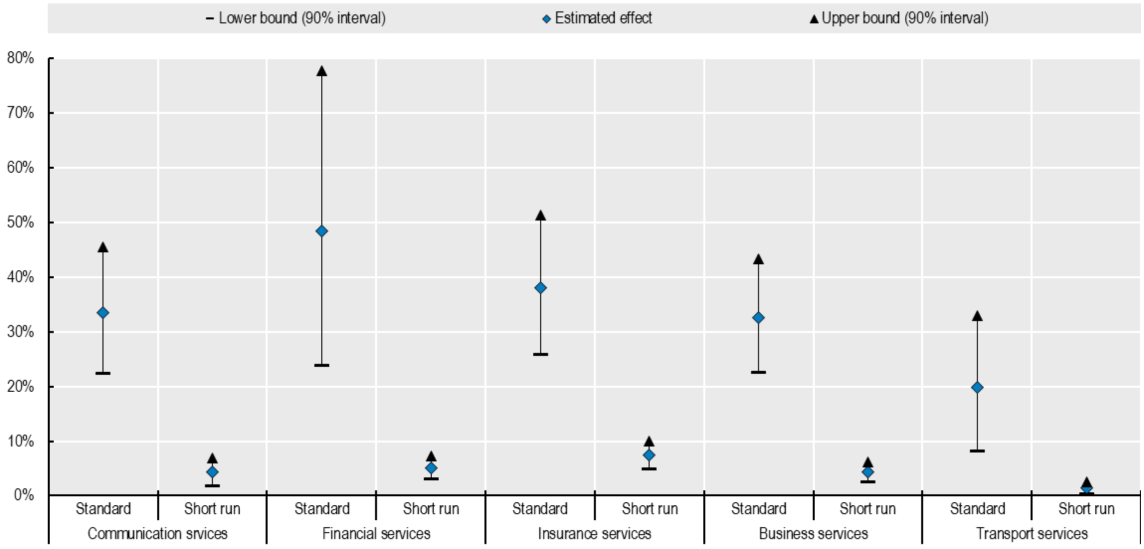
[18] A downside of this specification is the potential for omitted variable bias due to the existence of unobservable bilateral determinants of services trade that are not being accounted for by the standard gravity variables. A specification with symmetric pair fixed effects to control for time-invariant bilateral determinants of services trade flows is reported in Annex A as robustness check.

[19] Assuming an STRI change of 0.05 is one of the approaches proposed by the literature that uses the OECD STRI to model the trade effects of services trade reforms (Nordås and Rouzet, 2015[9]). It represents a relatively sizable policy change without being an outlier in the distribution of STRI changes at the country-sector-year level. Considering the actual policy content of a hypothesised reform, a decrease in the STRI by 0.05 can only be achieved through a comprehensive liberalisation covering different policy areas. For example, in the telecommunications sector, this reduction would require the elimination of a residency requirement for at least one board member; cancelling the consideration of economic interest in investment screening for either fixed or mobile telecommunication; eliminating labour market tests for intra-corporate transferees, contractual services suppliers and independent services suppliers; introduction of a policy that assigns universal services obligation on a competitive basis for either fixed or mobile telecommunication; and a reduction of the processing time for business visa to less than or equal to ten days.

[20] Regression tables are reported in Annex A.

Figure 2.1. Trade expansion from multilateral services reform

Expected growth in bilateral trade associated with a reduction of the STRI by 0.05



Note: All regressions include exporter-year and importer-year fixed effects and standard gravity variables. Standard errors are clustered by country-pair. Full regression results are reported in Annex A. STRI variable is bounded between 0 and 1 with an average of 0.18 in the countries and sectors covered.
Source: Authors' calculations

Patterns of results are relatively similar across sectors. In all sectors, this specification suggests that the short-run association of regulatory barriers and cross-border services trade is only between 7% and 20% of the corresponding association in the long run.[21] This implies that a major share of the trade-creating impact of services liberalisation only materialises in the long run, after the necessary adjustment of bilateral trade capacities.

The coefficient of the lagged dependant variable in the short-run gravity model lies between 0.8 and 1 and it is highly statistically significant in all sectors. This value indicates the partial correlation between past and contemporaneous services trade flows.[22] The remainder of this section describes results for individual sectors in more detail.

For communication services, the coefficient of -5.77 in the standard gravity model indicates that a reduction in a country's STRI by 0.05 points is associated with an increase of bilateral trade by around 33%. By contrast, the coefficient is -0.849 in the short-run gravity analysis. Based on this coefficient, a 0.05-point reduction in the STRI is associated with a mere 4% growth of trade in communication services in the short run.

The difference between the results from the standard gravity model and the short-run model are even more pronounced for financial services. The regression coefficient of -7.888 in the standard model falls to -1.002 when considering the lagged dependent variable as additional explanatory factor of services trade. This

[21] Identification of long-run adjustments is limited by the relatively short time span covered in this analysis. E.g., by construction, the impact of services reforms taking place in 2018 is only captured through changes in services trade in the contemporaneous and subsequent year. This implies that the standard gravity coefficient identified here might be a lower bound for the true long-run impact. In this case, the range between 7% and 20% would be an upper bound for the share of services trade growth occurring in the short term.

[22] More precisely, it can be considered an upper bound for the relationship between past trade flows and contemporaneous trade. The use of a lagged dependent variable may imply dynamic panel bias, the so-called Nickell (1981[80]) bias. The estimate is likely to be biased upward due to a positive correlation between the lagged dependent variable and unobservable bilateral determinants of trade not captured by the standard gravity variable. This correlation attributes additional predictive power to the lagged dependant variable.

implies that in the short run, a reduction in a country's STRI by 0.05 points is associated with an increase of bilateral trade by around 5%, while the long-run impact is around 48%.

Also, cross-border trade in insurance services is significantly impacted by the multilateral services trade barriers recorded in the STRI. Reducing a country's STRI by 0.05 is associated with an increase in insurance services exports of 38% in the standard gravity model, whereas the smaller coefficient in the short-run model implies an expansion of trade in insurance services by 7%.

For business services, a regression coefficient of -5.636 in the standard model suggests that a 0.05 reduction in the STRI can expand bilateral exports by around 33%, whereas short-run growth is only around 4%. Regulatory barriers to services trade have a relatively smaller impact for cross-border trade in transport services. A 0.1-point reduction in a country's STRI is associated with a 20% increase in transport services trade in the standard model and a 1% increase in the short run.

2.5. Discussion

The evidence summarised in this section shows that there is a time lag before the impact of services liberalisation fully materialises in an expansion of cross-border services trade. The short-run impact accounts for between 5% and 50% of the overall trade creation of liberalising services reform, depending on the specification. These results highlight the importance of structural adjustments as prerequisite for economic gains from services liberalisation. Economic policies that promote structural adjustment, including investment in business relationships between two countries, can be important components in a whole-of-government approach to services trade policy aiming to minimise transitory adjustment costs.

Policy should address obstacles that prevent or decelerate these adjustments on all stages of global value chains. Multilateral, plurilateral and bilateral agreement that include chapters on investment, competition, intellectual property and the temporary movement of workers can be important components of such a policy agenda, potentially amplifying the impact of services liberalisation on investment, growth and job creation.

Other building blocks include an education and skills strategy, allowing workers to develop the qualifications to compete in an increasingly knowledge-based economy. Additional policies for inclusive employment and income growth can ensure that all types of workers gain from the opportunities of services liberalisation. Facilitating the adjustment process, including reallocation of workers across sectors is crucial. It should be supported by well-designed social policies and a functioning labour market, including re-employment services and training programmes (Grundke and Arnold, 2022[33]).

3. Identifying heterogeneous reform impacts with synthetic controls

3.1. Motivation

The results from the gravity analysis provide insight into the overall short- and long-term effect of reforms in services trade policy. However, these findings are by construction averages across policies and countries. The precise impact of a particular instance of reform is likely to vary considerably depending on the circumstances.

To begin to unpack the variation in the effect of services trade policy, this section looks at four episodes of reform: communication services in Indonesia and in Israel, and the professional services and transport sectors in Hungary. Examining and comparing the results of this set of cases can offer some initial insights into the patterns of heterogeneity in policy impact.

Empirical studies aimed at identifying the effect of a specific event, such as a policy change, face a fundamental problem. The ideal would be to compare the situation after the reform to what would have happened if that reform had not taken place. But only one of these two versions of the world is observable.

To approximate the counterfactual situation without the reform, this analysis uses the synthetic control method developed by Abadie and Gardeazabal (2003[22]).[23]

The idea behind the synthetic control method is to borrow information from all the sectors and countries that did not experience any policy changes in order to approximate a counterfactual scenario. A weighted average of other country-sector pairs is constructed such that it resembles the country and sector of interest as closely as possible prior to the reform. Comparing how this weighted average evolves after the reform to the path followed by the case of interest gives an indication of the effect of the policy change.

The synthetic control method has several important advantages. It allows for an estimation of the effect of a particular event, such as a policy reform, without needing a direct comparison to be available. Global trends that affect all countries and sectors are to some extent accounted for, as are changes in the effect of different covariates on the outcome. The results produced by the synthetic control method are transparent, making it easy to assess the performance of the model and interpret the results.

3.2. Methodology

The starting point for an application of the synthetic control method is the selection of a particular case where a change took place (in this case, a policy reform). The goal is to identify the impact of the changes that took place in that specific instance.[24] The challenge in doing this is finding a way of estimating what would have happened in absence of the reform.

In order to do this, a sizeable set of observations is needed where no changes took place over the same period. Here, that means countries and sectors where no reform was implemented and which therefore continued as before. The assumption is that the case of interest would have continued to follow a broadly similar path, had the reform not taken place.

No one particular foreign sector is going to be identical to another, so a direct comparison between different sectors is not possible. Instead, a weighted average of the outcome variable of interest of several different sectors is used to approximate the relevant case. The combination of these other sectors in other countries together creates a synthetic control, a sort of "artificial twin", for the reform case.

To construct such a synthetic control case, a comparison case made up of several other sectors, the weights to be used need to be determined. These weights are calculated to make sure that the combined average resembles the real sector as closely as possible.[25] This resemblance can be based on any number of variables and features to obtain a realistic approximation.

The crucial assumption is that the weighted average outcome of this combination of other sectors after the reform is implemented, is identical to what would have happened to the sector of interest in absence of such a reform. This means that calculating the synthetic outcome based on these other sectors after the policy change gives an estimate of that same outcome for the reformed sector. Comparing this potential-but-unrealised outcome to the real observed outcome then shows the effect that the reform has had.

Instances of reform that are suitable to investigate using the synthetic control method need to meet several criteria. First, the change in policy should be large in order to be able to detect an effect. Countries' scores on the Services Trade Restrictiveness Index (STRI) were used to identify instances where large reforms had taken place. Second, to ensure that any detected effect can indeed be attributed to the particular policy change, there should be no other major changes taking place at the same time and no other major services trade policy reform throughout the period under investigation. Lastly, the algorithm needs to be able to

[23] An extensive body of work has used the synthetic control approach to assess the impact of a wide range of policy reforms, e.g. Abadie, Diamond and Hainmueller (2010[98]), Bohn, Lofstrom and Raphael (2014[99]), Kleven, Landais and Saez (2013[100]), Mitze et al. (2020[101]), Billmeier and Nannicini (2013[102]).

[24] This means, by necessity, that various changes happening simultaneously cannot be disentangled from one another. Any effect found is the cumulative impact of all reforms and shocks happening in the relevant time period.

[25] To be more specific, the weights assigned to other sectors as well as the weight given to the different variables on which the comparison is made are constructed to minimize mean squared prediction error (the sum of the squared difference between observed and synthetic variable across the pre-reform period). For an overview of the methodology, see Abadie (2021[97]).

construct a sufficiently fitting synthetic unit, the difficulty of which can vary depending on how far the country and sector lie from the average of the other units.

Based on these criteria, four episodes of services trade reform – concerning three sectors in three countries – were selected: the communication sector in Indonesia in 2016 and in Israel[26] in 2015, and the professional services and transport sectors in Hungary in 2015. These two pairs of cases have the additional benefit of allowing for a comparative perspective: the first is a set of different but comparable reforms in different countries, and the second is the same policy reform but affecting two different sectors.

In addition to instances of reform, the method requires data on countries and sectors in which no reform has taken place, which can be used to construct the synthetic control. The STRI enables identification of the presence or absence of such policy changes. For each country-sector pair, all the changes in STRI score between 2014 and 2020 were added up. If this combined change over this entire period was smaller than the change in STRI score of the particular reform episode under examination, the country-sector was included in the group of non-reform cases that could be used to construct the synthetic control unit.[27][28]

The structure of the data used for this analysis is broadly similar to the data used in the structural gravity model described above. The 22 STRI sectors have been aggregated into five broader sector groups for each country, as shown in Annex A. The outcome variable of interest is the logged import of services for the relevant country and sector.

In the construction of the synthetic control, the algorithm can take into account a variety of dimensions on which the comparison unit ought to be similar to the real case. The primary variable targeted is past services imports in the sector for each year, as replicating this closely is what gives confidence that the post-reform values of the outcome variable resemble the true counterfactual. To ensure a broader similarity between the two units, other aspects of the services trade profile were included as targets as well, namely the sector's exports, total services imports and exports for the country, and the sectoral STRI score.

3.3. Results

The first pair of cases are both instances of reform in the communications sector. In both instances market access to a part of this sector was liberalised significantly. However, the precise policy changes in question differ based on the relevant barriers in their context. Given these similarities and differences, these episodes can usefully be analysed side-by-side.

In an effort to encourage foreign investment Indonesia saw an extensive change in policy in its communications sector in 2016. This meant that rules were relaxed for the motion pictures and sound recording sectors. The precise restrictions removed included limitations on foreign ownership and the nationality of board members, as were minimum capital requirements.

The results of applying the synthetic control method for Indonesia's communications sector are shown in Figure 3.1.[29] As intended, the graph shows the observed and synthetic average moving together prior to the implementation of the reform. After 2016, Indonesia's imports of communication services continue to rise steadily. While imports of the synthetic unit also increase, they do so at a slower pace. A gap

[26] The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

[27] This approach only verifies that no major services trade-specific policy changes were implemented in other countries and sectors over the period. It remains possible that other reforms or political shifts took place that could also impact services trade flows. Since the synthetic control units are constructed as a combination of different other countries and sectors, the effect of such unaccounted-for changes on the results is considerably softened.

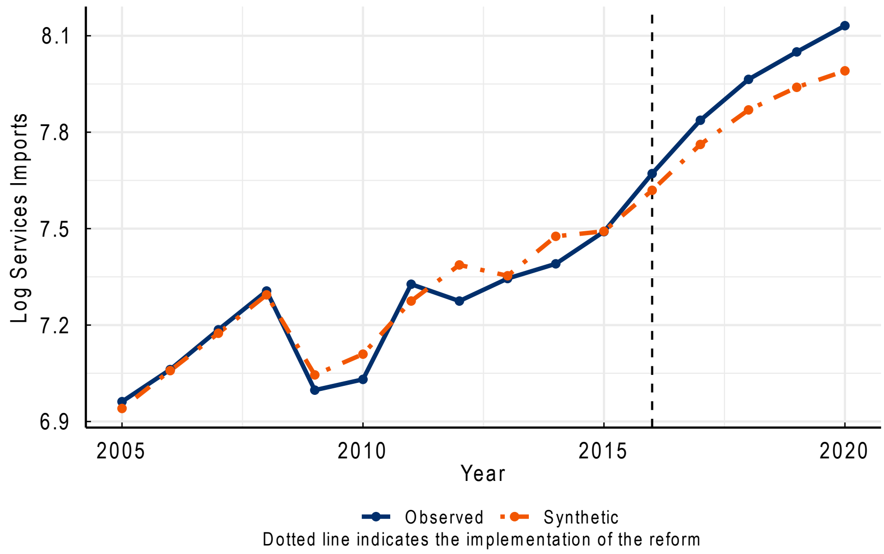
[28] Regardless of the specific sector under investigation, all services sectors in other countries are included in this set of potential comparison cases. However, when including only the same sector in the set of potential control cases instead, the results look almost identical (Annex B).

[29] The full set of countries and industries considered for inclusion in the comparison units as well as the actual weights generated to construct the synthetic controls can be found, for all four cases, in Table A B.1 and Figure A B.1 of Annex B.

increasingly opens between the two lines, suggesting a positive impact of the reforms on services trade imports.

Reform in the communication sector also took place in Israel in 2015. There, the policy changes focused on the telecommunications sector. Several technical measures, regulating reference offers and access to networks, were reformed. In practice, this meant that foreign telecommunications firms had far easier access into the Israeli market.

Figure 3.1. The effect of reform in Indonesia's communication sector



Note: This graph shows the outcome of a synthetic control analysis of the changes in Indonesia's communications sector. The blue solid line shows real services imports, while the dotted orange line shows the weighted average imports of the synthetic control. The difference between the two after 2016, when the policy change took place, can be interpreted as the effect of the reform on imports.
Source: Authors' calculations.

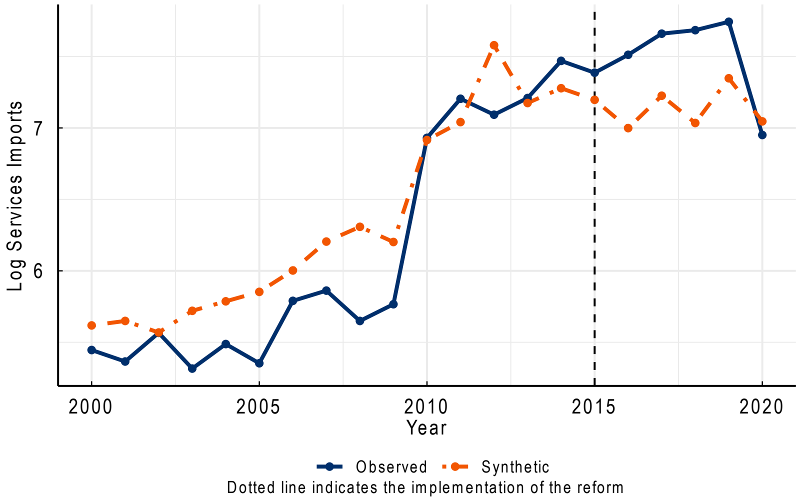
Figure 3.2 shows the results of the synthetic control analysis for Israel's communications sector. Prior to the reform in 2015, the weighted average line follows the same trend as the real observed imports, although the rather sharp jump in 2009 makes it more challenging for the model to fit precisely. After the policy change, the synthetic control stays roughly constant, while in fact communication services imports in Israel increased notably. This suggests that the opening of the telecom market indeed led to more export by foreign firms.

Comparing the results for Indonesia and Israel, both cases show that liberalisation resulted in increased imports of telecommunication services. Tentatively, it appears as if the effects on services imports in Israel happened more rapidly whereas those in Indonesia took several years to fully take effect. This could be the result of the nature of the policy changes, the competitiveness and lucrativeness of the particular market, or how quickly the relevant firms were able to adjust their investment and sales decisions.

For example, Israel's reform allowed the use of pre-existing networks by foreign firms which may have required less investment in physical capital. The fast-moving nature of the telecommunications sector may also have encouraged rapid response by businesses. In Indonesia, the policy changes concerned the film and sound recording industry, a sector where project planning cycles can take several years to complete and execute.

The second pair of cases is a comparison within a single country. Hungary implemented a series of changes to its migration law in 2015, which impacted services trade regulation by limiting the quotas of available visas needed for work permits for service providers coming from outside the EU/EEA and shortening the length of such visas from three to one years. These changes affected all services sectors, captured by so-called "horizontal" measures in the STRI. Yet while the substance of the reforms was consistent across the economy, this need not mean that their impact was also equally felt in all sectors.

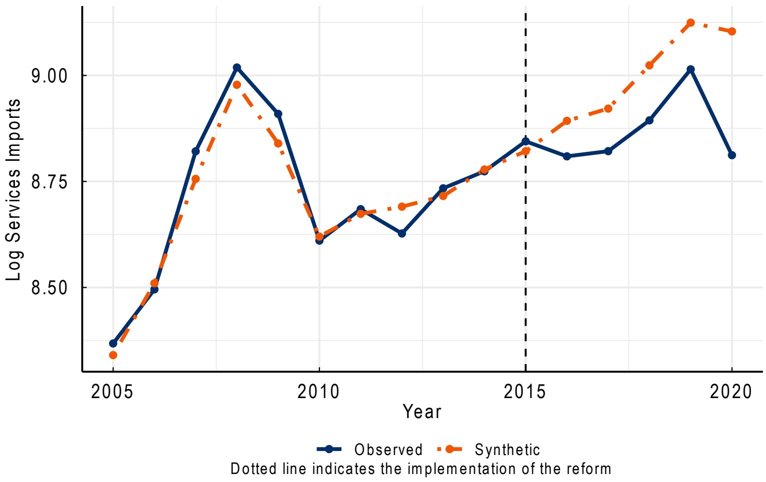
Figure 3.2. The effect of reform in Israel's communication sector



Note: This graph shows the outcome of a synthetic control analysis of the changes in Israel's communications sector. The blue solid line shows real services imports, while the dotted orange line shows the weighted average imports of the synthetic control. The difference between the two after 2015, when the policy change took place, can be interpreted as the effect of the reform on imports.
Source: Authors' calculations.

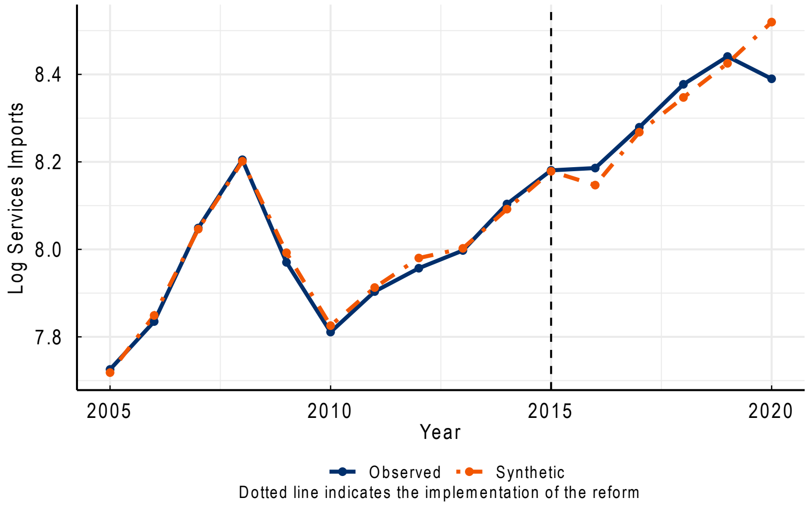
Figures 3.3 and 3.4 show the results of separate synthetic control analyses for Hungary's professional services and transport sectors. In the professional services sector, services imports grew at a slower pace after 2015 than for the weighted average of country-sector pairs that did not experience similar policy change, suggesting a negative impact of the restrictive regulatory change. But for transport services the pattern looks different: the observed and synthetic trade flows continue to follow each other very closely even after the reform has been implemented, suggesting no significant impact on imports.

Figure 3.3. The effect of reform in Hungary's Professional Services sector



Note: This graph shows the outcome of a synthetic control analysis of the changes in Hungary's professional services sector. The blue solid line shows real services imports, while the dotted orange line shows the weighted average imports of the synthetic control. The difference between the two after 2016, when the policy change took place, can be interpreted as the effect of the reform on imports.
Source: Authors' calculations.

Figure 3.4. The effect of reform in Hungary's transport sector



Note: This graph shows the outcome of a synthetic control analysis of the changes in Hungary's transport sector. The blue solid line shows real services imports, while the dotted orange line shows the weighted average imports of the synthetic control. The difference between the two after 2016, when the policy change took place, can be interpreted as the effect of the reform on imports.
Source: Authors' calculations.

One potential explanation for the difference in results for these two sectors is the nature of the policy change. Visas are an important component for the provision of services in Mode 4, i.e. the supply of a service through the temporary movement of suppliers across borders. But the relative importance of such travel in the trade of services depends considerably on the nature of the precise service in question.

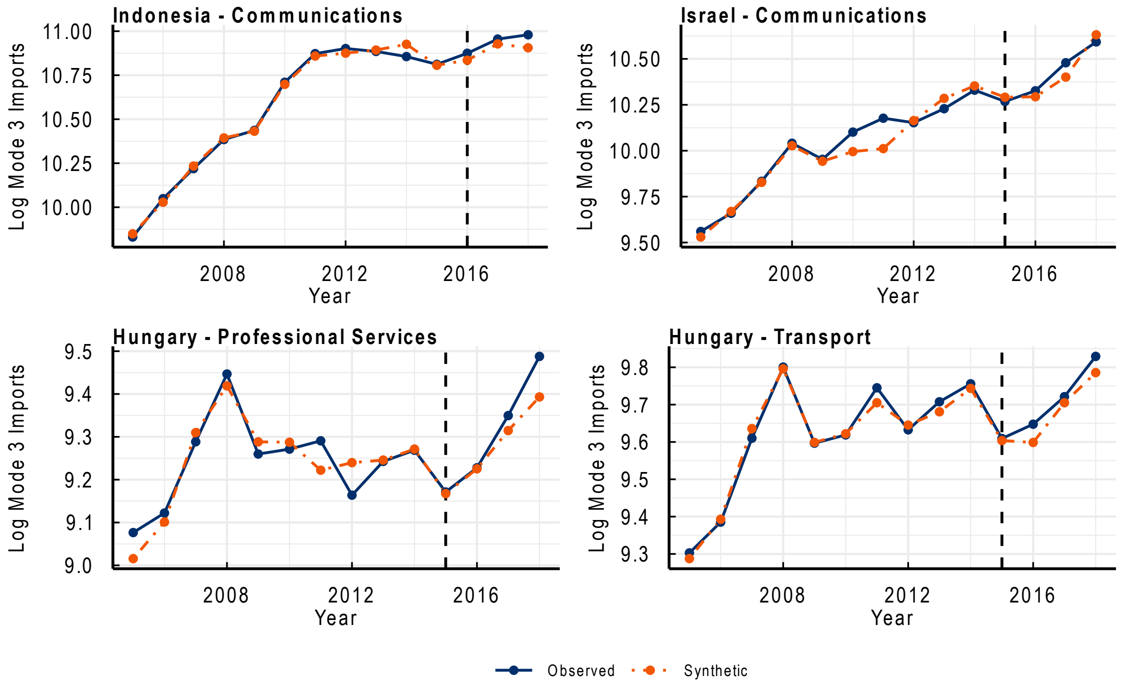
If Hungary's import of professional services often takes the form of in-person presence requiring visas while the same does not apply to transport services, this set of findings would be unsurprising. While the transport sector does rely on visas extensively, the reforms under investigation here applied to inter-corporate transferees specifically. Maritime and air crews are generally granted exemptions or fall under specific visa categories which would not be affected by the same quotas and limitations.[30]

The variation in impact of reforms on different modes of services supply can also be explored empirically. This is challenging in part, as disaggregated trade data by mode of supply is less available. The OECD's Analytical AMNE dataset provides insight on trade in mode 3 (commercial presence) in particular through the activities of multinational enterprises (Cadestin et al., 2018[34]). The time span is more limited, up to 2018, which gives only two or three years of post-reform data for these cases where the policy changes occurred in 2015 and 2016. Yet even this shorter period can be used to provide tentative insights.

Figure 3.5 shows the results of synthetic control models for all four cases under investigation, with as the outcome variable the (log of) Mode 3 services inflows in the relevant sector. Across all cases, no clear impact of reforms can be identified. There are several possible reasons for this. Other restrictions may remain in place that make Mode 3 trade costly relative to Mode 1, even with the removal of some salient barriers. Investment decisions also tend to be longer-term, meaning that firm behaviour may take more years to adjust. Therefore, these results should be revisited when a longer time range of Mode 3 data is available.

[30] A standard way to probe the robustness of results using synthetic control method is using a placebo test, whereby the cases of no policy reform are analysed as if they saw a change and their result is compared to the main finding. Graphs showing this analysis for all four cases can be found in Annex B (Figure AB.3). While these results show some of the limitations of this case-study based analysis, overall they are in line with the interpretation of results presented here.

Figure 3.5. Effect of reforms on Mode 3 imports



Note: This graph shows the outcome of four synthetic control analyses for the four cases of policy reform described above. In these graphs, the outcome variable under investigation is imports in mode 3 (commercial presence). The blue solid line shows real investment inflows, while the dotted orange line shows the weighted average imports of the synthetic control.
Source: Authors' calculations.

3.4. Discussion

The results of the synthetic control analyses above suggest two ways in which the overall effect of services reform on trade flows might vary considerably depending on circumstances. First, a liberalisation (or restriction) of roughly the same magnitude or ambition may have effects of different sizes and speeds depending on the precise policy measures in question and the local context. Second, the same reform might differ in its impact depending on the sector and the extent to which the relevant restriction presents a major barrier to the trade of that particular service.

However, the evidence presented here is far from conclusive on these conjectures. The synthetic control method cannot be used to test whether these patterns found here apply systematically across the reforms of different countries and sectors. By its design, it is limited to a case-study based approach.[31] Nevertheless, these results provide an intuitive illustration of the impact of services policy reform and can form the basis for further work that unpacks these findings further. Box 3.1 provides a brief discussion on the potential role of Machine-Learning based methods in this space.

[31] Methods that generalise the synthetic control approach to a more systematic analysis do exist (Yiqing, 2017[103]; Davide and Bradic, 2022[104]), but the data and cases relevant to this analysis are not suited to their application precisely because of the considerable heterogeneity in the content and impact of reforms.

Box 3.1. Machine learning-based methods

Machine Learning (ML) tools appear to be well-suited to further advance the investigation of the effects of services trade policy along the path traced by the Synthetic Control analyses presented above. Indeed, a new strand of methods have been developed to apply the strengths of ML algorithms to causal inference problems. These techniques can be potentially applied to study the impact of services trade reforms defined in terms of specific policy measures, also distinguishing between a restrictive and liberalising direction of the policy change. Moreover, being largely data-driven and requiring fewer assumptions on functional forms, they might seem also well-equipped to further explore heterogeneity in the effect of services trade policy. Robust applications of these techniques could therefore potentially contribute to shedding new light on the features of the economic environment that shape the effects of services trade reforms and, as a result, to higher precision in policy assessments.

A few recent applications of ML-based causal inference methods in observational settings (i.e. not relying on randomised control trials) reinforce expectations of the growing effectiveness of these methods in tackling economic and policy questions like the ones addressed in this study. These works include, among others: O'Malley (2018[35]), investigating the role of household leverage in shaping the causal effect of repossession on mortgage default probability; Tiffin (2019[36]) an empirical study of heterogeneity in the causal effect of financial crisis on GDP growth; Hoffman and Mast (2019[37]), offering new evidence on the causal impact of place-based fiscal policy on crime; Daoud and Johansson (2019[38]) exploring the factors behind of heterogeneous causal effects of IMF-led austerity programs in emerging economies on child poverty; Burauel and Schroeder (2019[39]), a study of the heterogeneous effects of the introduction of a minimum wage in Germany in 2015; Breinlich et al. (2021[40]) , assessing the trade effects of specific provisions in trade agreements.

Notwithstanding these and other successful studies and despite the continued data collection effort for the OECD STRI, an empirical framework constructed on the current panel database of STRI individual policy measures does not yet feature a sufficient degree of variation and statistical power for a robust application of ML-based causal inference methods. While these limitations are critical for exercises designed to exploit the granularity of policy information embedded in the STRI policy measures, they do not apply to the other methods presented in this synthesis paper which are based on the continuous STRI sectoral indicators.

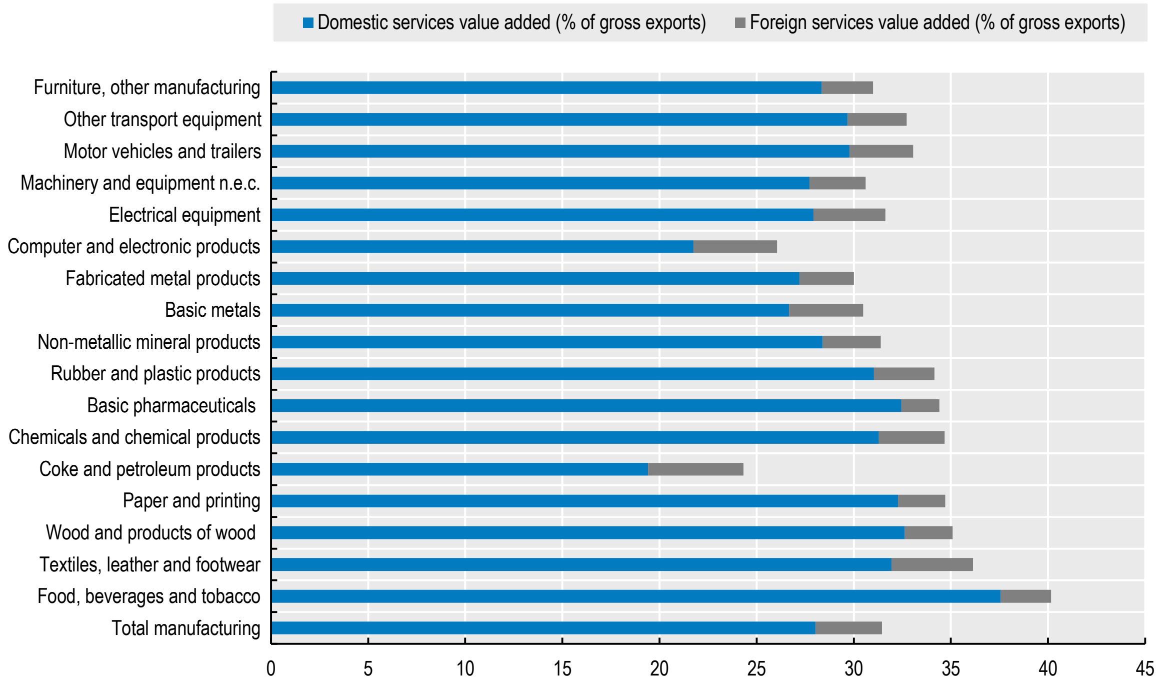
Annex C reports on exercises that deploy selected ML-based techniques to estimate the impact of services trade reforms using STRI individual policy measures and offers a rigorous assessment of the data and methodological limitations which prevent interpretable and robust estimation of causal effects in this application. By relying on an extended and continuously updated STRI database, similar applications might be successful in the future. Annex C (Section 4) also discusses potential efforts in data collection related to the OECD STRI that could be pursued to increase coverage and variation of relevant services trade policy measures.

4. Downstream effects of services trade reforms

4.1. Motivation

While the design of services trade policies is of obvious relevance to the economic performance of services sectors, there is a growing recognition that services and manufacturing are closely intertwined in global value chains (Miroudot and Cadestin, 2017[2]; Ariu et al., 2019[41]; Liu et al., 2020[42]). Services act as crucial inputs to manufacturing activities and play a central role in the coordination of the flows of goods, capital, and knowledge between different locations (Low, 2013[1]). Often referred to as business services or producer services, they allow the company buying this service to concentrate on its core activities and reduce costs of functions previously performed internally. These services may "make possible innovations either in organisation or in types of output" (Greenfield, 1966[42], p.128) and act as crucial agents in the realisation of economic benefits from economic integration. By facilitating the fine slicing of production processes across functions and locations, services inputs – such as transport services, logistics, and telecommunication services – enable firms to reap returns to specialisation in global value chains (Francois, 1990[43]; Deardorff, 2001[44]; Francois and Hoekman, 2010[45]).

Figure 4.1. Services value added embodied in manufacturing exports of OECD countries



Source: OECD TiVA database 2021.

Services value added accounts for nearly a third of gross manufacturing exports of the OECD countries. On average across 17 2-digit ISIC manufacturing industries, domestic services value added explains 29% of gross exports and foreign services value added a further 3% (Figure 4.1). Given the importance of services as intermediate inputs, reforms of the regulatory framework for services trade are likely to affect the economic performance of manufacturing industries. Trade in services can help downstream manufacturing firms to achieve a reduction in the quality-adjusted cost of services required for final goods production (Markusen, Rutherford and Tarr, 2005[46]). Especially in small economies and developing countries, the sourcing of foreign services may enable firms to overcome bottlenecks caused by the limited local availability of essential inputs and move up the quality ladder (Nordås, 2011[47]).

By facilitating access to a larger variety of services inputs at lower quality-adjusted cost, a reduction of barriers to services trade is likely to boost the productivity of downstream manufacturing activities.[32] Existing firm-level evidence shows that regulatory reforms lowering barriers to foreign entry into services sectors positively affected manufacturing productivity in the Czech Republic (Arnold, Javorcik and Mattoo, 2011[48]) and Chile (Fernandes and Paunov, 2012[49]).[33] Drawing on sector-level data for 57 countries,

[32] For a detailed discussion of the micro channels between services inputs and manufacturing performance, see Hoekman and Mattoo (2008[87]).

[33] Bas and Causa (2013[85]) also found that regulatory reforms in upstream sectors boosted the productivity of downstream manufacturing firms in China. Similarly, Bas (2014[88]) showed that regulatory liberalisation in three services sectors in India in the mid-1990s improved the export performance of downstream manufacturing firms. In a further study dedicated to the case of India, Arnold et al. (2015[86]) identified positive effects on manufacturing firms' productivity from regulatory reforms regarding banking, telecommunication, insurance, and transport services implemented in India after 1991.

Beverelli, Fiorini, and Hoekman (2017[50]) identified a positive effect of regulatory openness to services trade on downstream manufacturing productivity.[34]

The empirical analysis presented in this section contributes to this literature by linking the OECD STRI for five services sectors (air transport, telecommunications, financial and insurance services, computer services, logistics) to the labour productivity of 17 manufacturing industries in 44 economies[35] during 2014-2018.[36] By exploiting information on services trade policy reform over this five-year period, the analysis complements related cross-country studies. Earlier contributions often rely on broader measures of product market regulation (e.g. Bourlès et al., 2013[48]) rather than indicators of services trade regulation or capture barriers to services trade in a cross-sectional setup (e.g. Beverelli, Fiorini and Hoekman, 2017[50]), which are therefore incapable of analysing regulatory changes over time.

4.2. Methodology and data

As manufacturing industries are heterogeneous with respect to their sourcing relationships with upstream services sectors, they can be expected to differ in terms of their exposure to potential spillovers from regulatory reforms affecting a specific services sector. A downstream manufacturing industry heavily relying on the sourcing of air transport services seems likely to be more affected by regulatory changes for air transport services than a manufacturing industry that barely sources any inputs from air transport services. These sourcing relationships are captured in input-output tables.

This variation in sourcing patterns allows for the calculation of weighted measures of exposure to regulatory reforms, with the weights reflecting the intensity of a sector's input-output linkages to the economic activity for which a change in regulatory barriers to services trade is observed. The analysis relies upon variations of the following regression specification:

$$Y_{i,c,t} = \beta STRI_{s_{i,c,t-1}} + \boldsymbol{\delta}'\boldsymbol{Z}_{i,c,t-1} + \lambda_{it} + \lambda_{ct} + \varepsilon_{i,c,t}$$

where the dependent variable is apparent labour productivity – calculated as value added divided by the number of employees based on the OECD Trade in Value Added (TiVA) dataset and the OECD Trade in employment (TiM) database – of manufacturing industry i in country c in year t. The first right-hand side variable represents the weighted STRI score of upstream services sector s with respect to downstream manufacturing industry i in country c in year t-1. Taken from the 2010 input-output table of the United States, the weight reflects the share of inputs provided by upstream services sector s in the total expenditure on inputs of downstream manufacturing industry i.[37] The weighted STRI variable is calculated

[34] Moreover, Barone and Cingano (2011[89]), Bourlès et al. (2013[73]), and Cette, Lopez, and Mairesse (2016[90]) used data for OECD countries and found that anticompetitive upstream regulation of services sectors negatively affects downstream manufacturing productivity. A related set of empirical contributions directly links measures of services performance to measures of manufacturing performance. Focusing on the overall level of financial market development, a seminal study by Rajan and Zingales (1998[91]) found that industries that are more dependent on external finance develop faster in countries with more developed financial markets. In a firm-level analysis covering 119 countries, Hoekman and Shepherd (2015[92]) identified a positive link between services productivity and the productivity and export performance of manufacturing firms. Bilir, Chor, and Manova (2019[93]) exploited data on United States (US) multinational enterprises and found that higher levels of financial development in a country are associated with more entry by multinational affiliates.

[35] The sample covers all OECD countries, as well as Brazil, China, India, Indonesia, Russia, and South Africa.

[36] To ensure STRI data can be mapped to relevant information at the sector level for the dependent variable as well as control variables, the STRI scores of the aggregate "financial and insurance services" sector were calculated as the average of the STRI score of financial services and the STRI score of insurance services.

[37] Note that weights vary across industries, but do not vary across countries or years. This is due to the use of US weights for 2010 for all observations in the database. The use of time-invariant weights based on input–output data from a reference country and a pre-sample reference year is customary in the literature to address endogeneity concerns (Bas and Causa, 2013[85]; Arnold et al., 2015[86]; Beverelli, Fiorini and Hoekman, 2017[50]).

by multiplying this weight with the STRI score of services sector s in country c in year t-1.[38] The coefficient β indicates the association between upstream services regulation and downstream manufacturing productivity, with a negative coefficient indicating that a higher STRI score (i.e. a more restrictive regulatory framework) is associated with lower labour productivity.

$\boldsymbol{Z}$ is a vector of control variables encompassing input and output tariffs for manufacturing industry *i* in country *c* in year *t-1*. Given the relevance of GVC integration to productivity (Gal and Witheridge, 2019[51]), the controls also include two industry-level variables computed using the TiVA dataset: Foreign value added from all supplying industries as share of gross exports as well as the domestic value added embodied in foreign countries' exports expressed as a share of gross exports. Industry-year fixed effects (λ_{it}) absorb all shocks concerning a specific industry each year, such as a sudden drop in demand for the products of the automotive industry. In addition, country-year fixed effects (λ_{ct}) address all unobservable factors affecting all industries in a given country in a specific year, such as disruptions caused by a natural disaster.

4.3. Results

Table 1 displays the results of the cross-country panel analysis. While the same dependent variable (logarithm of added value per employee) was used in all regressions, the weighted measure of upstream services regulation refers to a different sector in each of the columns of Table 1. The first column reports results of a regression where the weighted STRI of the air transport sector was entered as the variable of interest. The coefficient of this variable is strongly significant and negative, indicating that a more restrictive regulatory framework for air transport services is associated with lower labour productivity of downstream manufacturing industries. This finding resonates with contributions highlighting the importance of air transport services to the flow of components and finished products around the globe (Hummels, 2007[52]; Feyrer, 2019[53]) and the facilitation of face-to-face encounters and knowledge diffusion (Hovhannisyan and Keller, 2014[54]; Tanaka, 2019[55]; Coscia, Neffke and Hausmann, 2020[56]).

In the case of logistics services (column 2), the coefficient similarly points to a link between regulatory barriers to trade in this sector and the productivity of downstream manufacturing industries. Yet, this result is not statistically significant. Considering the pivotal role of logistics services in modern supply chains (Blyde and Molina, 2015[57]), this is unexpected.[39] Conversely, the strongly significant negative coefficient of the weighted telecoms STRI (column 3) is in line with this sector's critical contribution to the exchange of information and the coordination of geographically dispersed production activities (Fink, Mattoo and Neagu, 2005[58]; Robert-Nicoud, 2008[59]; Nordås and Kim, 2013[24]; Fort, 2016[60]). While the coefficient in the case of computer services (column 4) is also negative, it is, surprisingly, not statistically significant.[40] As modern computer services rely upon the availability of advanced telecommunications infrastructure, it appears likely that the telecoms STRI (column 3) partly captures manufacturing firms' capacity to use modern data-intensive computer services.

[38] Formally, the calculation of the weighted STRI variable referring to the exposure of downstream manufacturing industry *i* in country *c* and year *t* to regulations of upstream services sector *s* can be expressed as $STRI\_weighted^{s}_{i,c,t} = STRI_{i,c,t} \times weight^{s}_{i}$. The weight is calculated as the share of services sector *s* in the total expenditure on inputs sourced by manufacturing industry *i* as recorded in the 2010 input-output table for the United States.

[39] The difficulty of measuring accurately the diverse set of activities and corresponding regulations that should be attributed to this sector might partly explain this pattern. The OECD STRI for logistics encompasses 19 subsectors. However, a robustness check with an alternative data source for the dependent variable (UNIDO INDSTAT 4 database) provides a negative and statistically significant coefficient estimate for logistics services (see Table A D.8 in the Annex).

[40] One potential explanation relates to the fundamental role of telecommunications infrastructure in defining the scope for the use of modern data-intensive computer services. Unlike the computer services STRI, the telecoms STRI covers details of the regulation of infrastructure access. For this analysis the telecoms STRI might therefore be a more relevant measure of manufacturing firm's capacity to draw on cutting-edge computer services. Moreover, measurement error – for example because of a divergence between the *de jure* regulation captured in the STRI and the de facto regulatory constraints experienced by business in practice – might explain this result.

As shown in the fifth column, there is a strongly significant, negative association between regulatory barriers to trade in financial and insurance services and manufacturing productivity. This finding chimes with a large body of empirical studies shedding light on the relevance of well-functioning financial services to the economic performance of manufacturing industries (Amiti and Weinstein, 2011[61]; Manova, Wei and Zhang, 2015[62]; Liu et al., 2020[42]).

Overall, these results underscore the importance of adopting a holistic perspective of services trade policy: The effects of reforms lowering regulatory barriers to services trade do not necessarily manifest themselves only "right here", i.e. in the services sector experiencing the policy change. When assessing the economic benefits of services trade liberalisation as well as the costs of restrictions on services trade, significant spillover effects on downstream sectors arising from the essential role of business services as inputs must be taken into account.[41]

Table 4.1. Regression results

	(1)	(2)	(3)	(4)	(5)
	Air transport	Logistics	Telecoms.	Computer services	Financial and insurance services
Weighted upstream STRI, t-1	-467.443***	-28.463	-134.692***	-38.051	-123.179***
	(114.616)	(34.022)	(42.544)	(104.452)	(47.352)
Mean output tariffs, t-1	-0.022	-0.027	-0.024	-0.029	-0.015
	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)
Mean input tariffs, t-1	0.020	0.026	0.019	0.030	0.009
	(0.048)	(0.049)	(0.049)	(0.049)	(0.048)
Foreign value added share of gross exports, t-1	-0.010***	-0.010***	-0.009***	-0.010***	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Domestic value added embodied in foreign countries' exports, t-1	0.049	0.051	0.056	0.050	0.053
	(0.039)	(0.038)	(0.038)	(0.038)	(0.039)
Observations	2,911	2,911	2,911	2,911	2,911
R-squared	0.850	0.845	0.847	0.845	0.847
Country-Year F.E.	YES	YES	YES	YES	YES
Industry-Year F.E.	YES	YES	YES	YES	YES

Note: Dependent variable: logarithm of value added per employee. Note that the weighted upstream STRI refers to a different upstream services sector in each of the regressions whose results are shown in this table. The column titles indicate the corresponding upstream services sector whose STRI score was used to calculate the weighted upstream STRI variable. Robust standard errors (clustered at country-industry level) in parentheses.
Source: Authors' calculations.

Whereas the results presented in Table 4.1 demonstrate that the regulation of upstream services sectors has spillover effects on downstream manufacturing industries, they do not directly indicate the economic magnitude of these effects. Focusing on the three services sectors (air transport, telecommunications, financial and insurance services) for which a statistically significant coefficient of the STRI variable was observed, the economic significance of the findings can be illustrated based on a reform scenario in which

[41] An ancillary step suggests the negative link between upstream restrictions on services trade and downstream manufacturing productivity is particularly pronounced at lower levels of economic development. In an exploratory analysis (not reported), the main specification is complemented by an interaction of the weighted STRI score with a dummy equalling one if a given country was in the bottom 25% of the GDP per capita distribution among all countries included in the analysis in 2014. The coefficient of this interaction is always negative, and it is statistically significant in four of the five sectors (air transport, logistics, computer services, financial services). Yet, the limited variation in levels of economic development across the countries included in the analysis as well as the short period covered by the sample impede an in-depth analysis of this aspect.

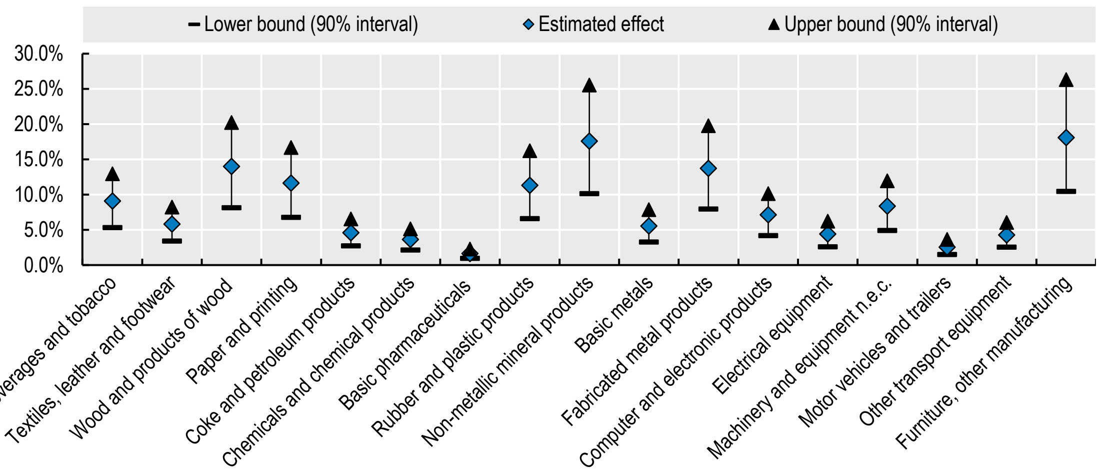
the dismantling of barriers of services trade is assumed to lower the STRI score in a given services sector by 0.05.

This is a relatively ambitious reform scenario that can only be achieved through the combination of several policy changes. In the case of telecommunications services, a reduction of the STRI score by 0.05 could be achieved, for example, through the implementation of all of the following regulatory changes: Privatisation of a state-owned company in this sector, a reform ensuring that the government can no longer overrule decisions taken by the telecommunications regulator, and the abolition of labour market tests for intra-corporate transferees as well as independent service suppliers.[42]

Figures 4.2 to 4.4 display the estimated downstream manufacturing productivity effects of these simulated reductions of barriers to services trade for each of these three services sectors. The blue marker indicates the potential percentage change in labour productivity that could be achieved in this hypothetical scenario. Further markers refer to the upper and lower bounds of the 90% confidence interval.

Estimated productivity effects are large for several manufacturing industries. For example, the simulated reductions in barriers to air transport trade are expected to increase productivity by 18% for the set of economic activities subsumed under ISIC (rev.4) categories 31 to 33 (Figure 4.2). Examples of manufacturing industries included in this division include “manufacture of medical and dental instruments” as well as “installation of industrial machinery and equipment”. The transport of goods with a high value-to-weight ratio and business travel of engineers and managers in these manufacturing industries is likely to benefit from greater openness to trade in air transport services.

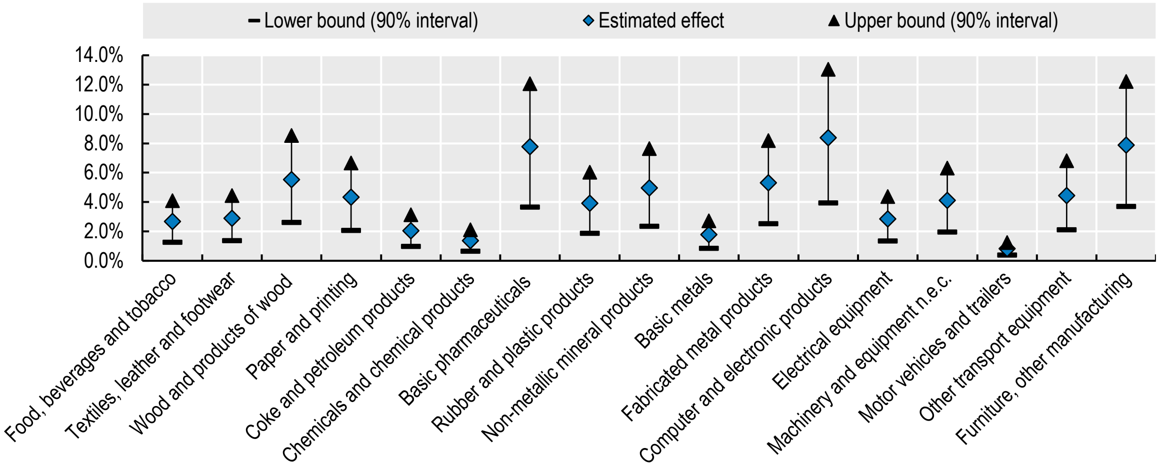
Figure 4.2. Productivity gains in liberalisation scenario: Air transport



Note: This graph illustrates estimated gains in manufacturing labour productivity in the case of a reform scenario concerning trade in air transport services. The liberalising reforms in this scenario are assumed to be equivalent to a 0.05 reduction in the STRI score for air transport services. The blue markers indicate the expected increase in labour productivity for each of the 17 manufacturing sectors included in the analysis. Based on the standard errors of the estimated coefficients reported in Table 4.1, this graph also displays the lower bound as well as the upper bound of the 90% confidence interval.
Source: Author’s calculations.

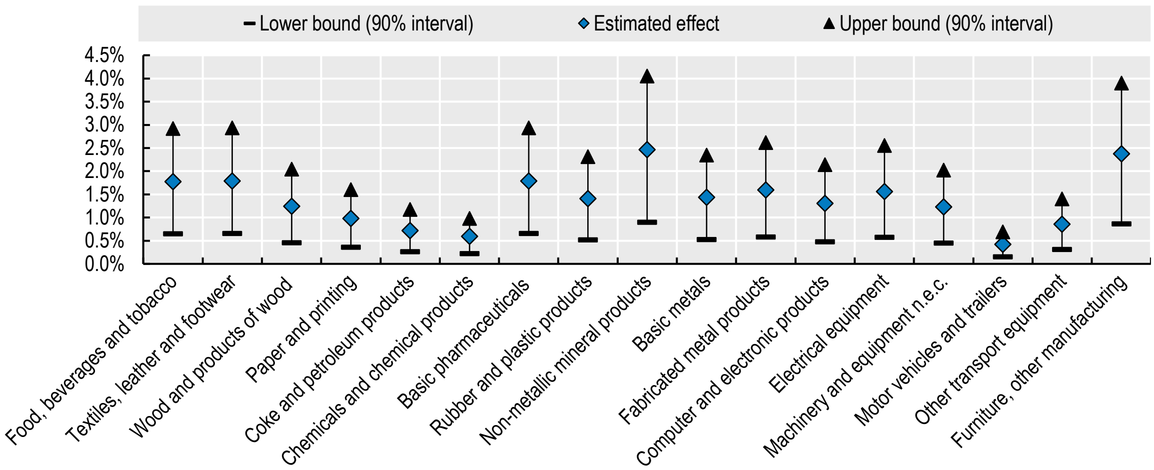
[42] With respect to financial and insurance services, the following set of reforms would roughly correspond to a 0.05 reduction of the STRI score: The introduction of a maximum time limit for the regulator to decide upon applications, the removal of discretionary control of the government regarding the regulatory agency’s funding, the removal of explicit local bias in public procurement, a reform specifying that the government can no longer overrule decisions of the regulator, and the introduction of an adequate open comment procedure open to interested persons, including foreign suppliers.

Figure 4.3. Productivity gains in liberalisation scenario: Telecommunications



Note: This graph illustrates estimated gains in manufacturing labour productivity in the case of a reform scenario concerning trade in telecommunications services. The liberalising reforms in this scenario are assumed to be equivalent to a 0.05 reduction in the STRI score for telecommunications services. The blue markers indicate the expected increase in labour productivity for each of the 17 manufacturing sectors included in the analysis. Based on the standard errors of the estimated coefficients reported in Table 4.1, this graph also displays the lower bound as well as the upper bound of the 90% confidence interval.

Figure 4.4. Productivity gains in liberalisation scenario: Financial and insurance services



Note: This graph illustrates estimated gains in manufacturing labour productivity in the case of a reform scenario concerning trade in financial and insurance services. The liberalising reforms in this scenario are assumed to be equivalent to a 0.05 reduction in the STRI score for finance and insurance services. The blue markers indicate the expected increase in labour productivity for each of the 17 manufacturing sectors included in the analysis. Based on the standard errors of the estimated coefficients reported in Table 4.1, this graph also displays the lower bound as well as the upper bound of the 90% confidence interval

In the scenario concerning the lowering of barriers to trade in telecommunications services (Figure 4.3), the simulation predicts an 8.4% increase in the labour productivity of the manufacturing industry “Computers and electronic products”. This effect is likely to reflect the key role of telecommunications services as providers of the infrastructure that forms the backbone of modern information and communications technology. Sizeable productivity increases are similarly predicted for several manufacturing industries in the scenario concerning financial services, e.g. 2.5% in the case of “non-metallic mineral products”. This category encompasses, for example, the manufacture of cement – a capital-intensive activity that can be expected to benefit from access to high-quality financial services.

There is considerable heterogeneity across manufacturing industries in the estimated productivity effect of the simulated reforms. These differences reflect diverse sourcing patterns. Sectors that are predicted to experience relatively small productivity gains, such as “basic pharmaceuticals” in the air transport reform scenario (Figure 4.2, +1.6%), display a low input sourcing intensity with respect to the corresponding upstream services sector.[43] On average across the 17 manufacturing industries included in the analysis, the simulated reforms of services trade policy – equivalent to a 0.05 reduction in the STRI score – are estimated to increase downstream manufacturing productivity by 8.4% (air transport reform scenario), 6.5% (telecommunications), and 2.3% (financial services).[44] The full set of estimated productivity gains by sector is provided in Table A.D1, Tables A.D2 and A.D3 in Annex D.

The econometric results presented in this section were confirmed by a set of robustness checks, such as the use of an alternative set of controls.[45] Yet, several shortcomings ad caveats should be taken into account when interpreting these findings. The regression analysis relies on sectoral data at a high level of aggregation, leaving no room for the exploration of dynamics at the level of more fine-grained industries or firms. While the discussion has pointed to several potentially relevant channels, detailed information on firm-level sourcing decisions, expenditure and trading patterns would be needed to investigate the mechanisms underlying spillover effects from services trade policy to manufacturing performance. Similarly, the relatively short period for which STRI data can be combined with industry-level productivity data limits the sample size.[46]

Moreover, the simulated reforms should be considered as hypothetical scenarios for illustrative purposes. As policy makers carefully balance different priorities and objectives, the scale of the simulated regulatory changes might appear relatively ambitious. In addition, it is important to bear in mind that the STRI captures *de jure* policy changes. For these regulatory changes to result in *de facto* changes of trade barriers perceived by businesses and generate the potential for productivity gains, efforts to strengthen rule of law and limit corruption should not be neglected: The contribution of Beverelli, Fiorini, and Hoekman (2017[50]) underscores the role of domestic institutional quality in shaping a country’s capacity to generate productivity gains form lower services trade restrictions.

[43] The productivity effect PE for a given manufacturing industry was calculated as PE=(exp(policy_change * coefficient) -1) *100, with policy_change being the simulated change in the STRI score weighted by the corresponding input share and coefficient referring to the regression coefficient of the corresponding STRI variable in Table 4.1.

[44] Note that considerably higher effects are obtained in the case of a significantly more ambitious reform scenario (Table A.D4, Table A.D5, and Table A.D6). In this alternative setting, the most restrictive country is assumed to implement regulatory changes that halve the gap between its own STRI score in the corresponding sector and the average STRI score across all countries covered by the STRI database. Drawing on STRI data for 2020, this would mean lowering the STRI score by 0.09 in air transport. In telecommunications services, the illustrative scenario assumes a reduction in the STRI score by 0.24. In financial and insurance services, the simulated halving of the gap between the most restrictive country and the average across all countries in the STRI database would be equivalent to a reduction of the most restrictive country’s score by 0.14. On average across the 17 manufacturing industries included in the analysis, these very ambitious reforms of services trade policy are estimated to increase downstream manufacturing productivity by 16% (air transport reform scenario), 22% (telecommunications), and 11% (financial services). These effects are similar to the magnitude identified in a related study by Beverelli, Fiorini, and Hoekman (2017[50]). In a broadly similar quantification exercise, these authors report an average productivity increase of 22%.

[45] The results remain similar when using different set of control variables (Table A.D7). The results are also robust to the omission of the United States from the sample.

[46] Efforts to analyse the distinction between short run versus long run effects of upstream services reforms on downstream manufacturing productivity would benefit from the availability of a longer time series. Exploratory regressions with different lag structures for the weighted STRI variables based on the sample used for this analysis did not produce conclusive results.

5. Conclusions

An open and ambitious research agenda on the effects of services trade reforms is the travelling companion of any evidence-based policy effort for promotion of services market access. This synthesis report lays out three strands of new empirical analysis on the effects of services trade reforms on economic outcomes using the OECD STRI. The study leverages on the strengths and complementarities of different methods, including gravity modelling, synthetic control method and country-sector-level panel econometrics. The findings that have been presented here shed new light on important dimensions of the impact of services trade policy reforms.

First, reducing policy barriers as captured by the STRI can increase services trade flows already in the short run, while additional effects continue to accrue over the longer run. The results derived from a state-of-art extension of the structural gravity model designed to sharpen the empirical assessment of trade response dynamics, offer new evidence in this direction. These estimates suggest that, while a share of the trade-creating effect of services liberalisation can indeed materialise during the first years immediately after a reform, this effect is likely to be relatively small and sector specific, with trade in communication services showing the strongest and more robust response in the short run. Capturing the full effects of services reform requires adopting a long-run perspective, with additional impact accruing as production and investment decisions catch up to the new policies.

A potential trade effect of reforms already in the short run is also seen in applications of the synthetic control method. Moreover, the case-study nature of these exercises highlights other relevant dimensions of heterogeneity in the effect of services trade policy. Not only reforms tend to have different trade effects over time and across sectors, but their impact is likely to vary across different economic environments. Furthermore, by distinguishing between liberalising and tightening reforms, the findings from the synthetic control applications suggest that removing services trade restrictions might not necessarily trigger trade responses with opposite symmetric magnitudes and dynamics with respect to those that follow a protectionist policy change. Overall, heterogeneity emerges as an important feature characterising trade-responses to services trade reforms.

While this synthesis report shows how ML-based methods for causal inference have desirable properties to deepen the investigation of such heterogeneity, it also highlights the limitations of the empirical setting of the current OECD STRI database that prevent successful applications of these techniques.

Finally, the impact of services trade reforms can be investigated beyond their effects on trade flows. By exploiting variation across manufacturing sectors in the use of services inputs, panel econometric analysis shows that industries downstream the supply chain can benefit from policy reforms targeting upstream services sectors. In particular, removing barriers to services trade imports has economically sizable spillover effects on the productivity of domestic manufacturing sectors. These findings reflect the essential role of services as intermediate inputs in production processes and provide a strong motive for a holistic approach when assessing the impact of services trade reforms.

While this paper is a significant step forward, further empirical work can be done in the future to sharpen and extend the understanding of the economic responses to services trade policy. The analyses presented in this synthesis report suggest several open challenges for the research agenda on causal effects of services trade reforms.

First, services trade policy is complex and multidimensional. It features elements that pertain to market access but also to regulations related to movement of people, competition policy, public procurement, and transparency. For this reason, services trade reforms can be designed to target specific elements within a wide set of relevant policy measures.

The OECD STRI regulatory database keeps track of a rich set of individual policy measures, and these can be used to build indicators of reforms with very specific policy content. However, existing methods are limited in their ability to leverage these detailed indicators, particularly given the finite number of reforms observed. Future research should continue to experiment with methods that can potentially use individual STRI policy measures and estimate their effect on economic outcomes.

Second, services trade reforms can potentially impact on a variety of economic outcomes. These include not only services trade and downstream manufacturing performance, but also competition, innovation, and employment in domestic services sectors; access to services; and economic development. Combining different methods will help deepening the existing body of evidence in this space. In particular, well established cross-country panel exercises should be complemented by case-studies approaches, including micro-data based empirical studies but also more qualitative investigations, to address research questions where cross-country panel dataset are too small or not available.

Finally, heterogeneity in the effect of services trade policy should be further explored. Services trade reforms do not happen in a vacuum. On the contrary, they interact in complex ways with relevant features of the policy and economic environment and their impact can be affected by those interactions. The analysis presented in this synthesis paper confirms that heterogeneity does exist and that it shapes the effects of services trade policy across sectors, in different economic environments, and over time. Investigating this heterogeneity can provide useful insights for the design of services trade reforms that can contribute to inclusive and sustainable economic growth.

References

Abadie, A. (2021), "Using synthetic controls: Feasibility, data requirements, and methodological aspects", *Journal of Economic Literature*, Vol. 59/2, pp. 391-425. [97]

Abadie, A., A. Diamond and J. Hainmueller (2010), "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program", *Journal of the American Statistical Association*, Vol. 105/490, pp. 493-505. [98]

Abadie, A. and J. Gardeazabal (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country", *American Economic Review*, Vol. 93/1, pp. 113-132, https://doi.org/10.1257/000282803321455188. [22]

Ahmad, S. et al. (2020), *Brexit Uncertainty and its (Dis)Service Effects*, National Bureau of Economic Research, Cambridge, MA, https://doi.org/10.3386/w28053. [14]

Amiti, M. and D. Weinstein (2011), "Exports and Financial Shocks", *The Quarterly Journal of Economics*, Vol. 126/4, pp. 1841-1877, https://doi.org/10.1093/qje/qjr033. [61]

Anderson, J. et al. (2018), "Dark costs, missing data: Shedding some light on services trade", *European Economic Review*, Vol. 105, pp. 193-214, https://doi.org/10.1016/j.euroecorev.2018.03.015. [25]

Anderson, J., M. Larch and Y. Yotov (2020), "Transitional Growth and Trade with Frictions: A Structural Estimation Framework", *The Economic Journal*, Vol. 130/630, pp. 1583-1607, https://doi.org/10.1093/ej/ueaa020. [17]

Anderson, J. and Y. Yotov (2020), "Short run gravity", *Journal of International Economics*, Vol. 126, p. 103341, https://doi.org/10.1016/j.jinteco.2020.103341. [21]

Ariu, A. et al. (2019), "The interconnections between services and goods trade at the firm-level", *Journal of International Economics*, Vol. 116, pp. 173-188, https://doi.org/10.1016/j.jinteco.2018.10.005. [41]

Arnold, J. et al. (2015), "Services Reform and Manufacturing Performance: Evidence from India", *The Economic Journal*, Vol. 126/590, pp. 1-39, https://doi.org/10.1111/ecoj.12206. [86]

Arnold, J., B. Javorcik and A. Mattoo (2011), "Does services liberalization benefit manufacturing firms?", *Journal of International Economics*, Vol. 85/1, pp. 136-146, https://doi.org/10.1016/j.jinteco.2011.05.002. [48]

Athey, S. and G. Imbens (2016), "Recursive partitioning for heterogeneous causal effects", *Proceedings of the National Academy of Sciences*, Vol. 113/27, pp. 7353-7360, https://doi.org/10.1073/pnas.1510489113. [65]

Baier, S. and J. Bergstrand (2007), "Do free trade agreements actually increase members' international trade?", *Journal of International Economics*, Vol. 71/1, pp. 72-95, https://doi.org/10.1016/j.jinteco.2006.02.005. [27]

Barone, G. and F. Cingano (2011), "Service Regulation and Growth: Evidence from OECD Countries", *The Economic Journal*, Vol. 121/555, pp. 931-957, https://doi.org/10.1111/j.1468-0297.2011.02433.x. [89]

Bas, M. (2014), “Does services liberalization affect manufacturing firms’ export performance? Evidence from India”, *Journal of Comparative Economics*, Vol. 42/3, pp. 569-589, https://doi.org/10.1016/j.jce.2013.06.005. [88]

Bas, M. and O. Causa (2013), “Trade and product market policies in upstream sectors and productivity in downstream sectors: Firm-level evidence from China”, *Journal of Comparative Economics*, Vol. 41/3, pp. 843-862, https://doi.org/10.1016/j.jce.2013.01.010. [85]

Benz, S. (2017), “Services trade costs: Tariff equivalents of services trade restrictions using gravity estimation”, *OECD Trade Policy Papers*, No. 200, OECD Publishing, Paris, https://doi.org/10.1787/dc607ce6-en. [11]

Benz, S. and F. Gonzales (2019), “Intra-EEA STRI Database: Methodology and Results”, *OECD Trade Policy Papers*, No. 223, OECD Publishing, Paris, https://doi.org/10.1787/2aac6d21-en. [32]

Benz, S. and A. Jaax (2022), “The costs of regulatory barriers to trade in services: New estimates of ad valorem tariff equivalents”, *Economics Letters*, Vol. 212, p. 110057, https://doi.org/10.1016/j.econlet.2021.110057. [19]

Benz, S. and A. Jaax (2020), “The costs of regulatory barriers to trade in services: New estimates of ad valorem tariff equivalents”, *OECD Trade Policy Papers*, No. 238, OECD Publishing, Paris, https://doi.org/10.1787/bae97f98-en. [5]

Benz, S., A. Jaax and Y. Yotov (2022), “Shedding light on the drivers of services tradability over two decades”, *OECD Trade Policy Papers*, No. 264, OECD Publishing, Paris, https://doi.org/10.1787/d5f3c149-en. [4]

Benz, S., D. Rouzet and F. Spinelli (2020), “Firm heterogeneity in services trade: Micro-level evidence from eight OECD countries”, *The World Economy*, Vol. 43/11, pp. 2905-2931, https://doi.org/10.1111/twec.13031. [18]

Beverelli, C., M. Fiorini and B. Hoekman (2017), “Services trade policy and manufacturing productivity: The role of institutions”, *Journal of International Economics*, Vol. 104, pp. 166-182, https://doi.org/10.1016/j.jinteco.2016.11.001. [50]

Bilir, L., D. Chor and K. Manova (2019), “Host-country financial development and multinational activity”, *European Economic Review*, Vol. 115, pp. 192-220, https://doi.org/10.1016/j.euroecorev.2019.02.008. [93]

Billmeier, A. and T. Nannicini (2013), “Assessing economic liberalization episodes: A synthetic control approach”, *Review of Economics and Statistics*, Vol. 95/3, pp. 983-1001. [102]

Blyde, J. and D. Molina (2015), “Logistic infrastructure and the international location of fragmented production”, *Journal of International Economics*, Vol. 95/2, pp. 319-332, https://doi.org/10.1016/j.jinteco.2014.11.010. [57]

Bohn, S., M. Lofstrom and S. Raphael (2014), “Did the 2007 Legal Arizona Workers Act reduce the state’s unauthorized immigrant population?”, *Review of Economics and Statistics*, Vol. 96/2, pp. 258-269. [99]

Borchert, I. and M. Di Ubaldo (2021), *Deep Services Trade Agreements and their Effect on Trade and Value Added*, World Bank, Washington, DC. [26]

Borchert, I. et al. (2019), *Applied Services Trade Policy. A Guide to the Services Trade Policy Database and the Services Trade Restrictions Index*, https://www.wto.org/english/res_e/reser_e/ersd201914_e.pdf. [71]

Borchert, I., B. Gootiiz and A. Mattoo (2013), "Policy Barriers to International Trade in Services: Evidence from a New Database", *The World Bank Economic Review*, Vol. 28/1, pp. 162-188, https://doi.org/10.1093/wber/lht017. [6]

Bourlès, R. et al. (2013), "Do Product Market Regulations in Upstream Sectors Curb Productivity Growth? Panel Data Evidence For OECD Countries", *The Review of Economics and Statistics*, Vol. 95/5, pp. 1750-1768, https://doi.org/10.1162/rest_a_00338. [73]

Breiman, L. (2001), "Random Forests", *Machine Learning*, Vol. 45/1, pp. 5-32, https://doi.org/10.1023/a:1010933404324. [67]

Breinlich, H. et al. (2021), *Machine Learning in International Trade Research : Evaluating the Impact of Trade Agreements*, http://hdl.handle.net/10986/35451. [40]

Burauel, P. and C. Schroeder (2019), "The German Minimum Wage and Wage Growth: Heterogeneous Treatment Effects Using Causal Forests", *SOEP papers on Multidisciplinary Panel Data Research*, No. 1018, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin, https://doi.org/10.2139/ssrn.3415479. [39]

Cadestin, C. et al. (2018), "Multinational enterprises and global value chains: New Insights on the trade-investment nexus", *OECD Science, Technology and Industry Working Papers* No. 2018/05, https://doi.org/10.1787/194ddb63-en. [34]

Cadestin, C. et al. (2018), "Multinational enterprises and global value chains: New Insights on the trade-investment nexus", *OECD Science, Technology and Industry Working Papers*, No. 2018/05, OECD Publishing, Paris, https://doi.org/10.1787/194ddb63-en. [23]

Cadestin, C. and S. Miroudot (2020), "Services exported together with goods", *OECD Trade Policy Papers*, No. 236, OECD Publishing, Paris, https://doi.org/10.1787/275e520a-en. [20]

Cette, G., J. Lopez and J. Mairesse (2016), "Market Regulations, Prices, and Productivity", *American Economic Review*, Vol. 106/5, pp. 104-108, https://doi.org/10.1257/aer.p20161025. [90]

Chaney, T. (2014), "The Network Structure of International Trade", *American Economic Review*, Vol. 104/11, pp. 3600-3634, https://doi.org/10.1257/aer.104.11.3600. [77]

Chernozhukov, V. et al. (2018), "Double/debiased machine learning for treatment and structural parameters", *The Econometrics Journal*, Vol. 21/1, pp. C1-C68, https://doi.org/10.1111/ectj.12097. [64]

Coscia, M., F. Neffke and R. Hausmann (2020), "Knowledge diffusion in the network of international business travel", *Nature Human Behaviour*, Vol. 4/10, pp. 1011-1020, https://doi.org/10.1038/s41562-020-0922-x. [56]

Daoud, A. and F. Johansson (2019), *Estimating Treatment Heterogeneity of International Monetary Fund Programs on Child Poverty with Generalized Random Forest*, Center for Open Science, https://doi.org/10.31235/osf.io/awfjt. [38]

Davide, V. and J. Bradic (2022), "Synthetic learner: model-free inference on treatments over time", *Journal of Econometrics*. [104]

Deardorff, A. (2001), "International Provision of Trade Services, Trade, and Fragmentation", *Review of International Economics*, Vol. 9/2, pp. 233-248, https://doi.org/10.1111/1467-9396.00276. [44]

Dix-Carneiro, R. and B. Kovak (2017), "Trade Liberalization and Regional Dynamics", *American Economic Review*, Vol. 107/10, pp. 2908-2946, https://doi.org/10.1257/aer.20161214. [16]

Egger, P., M. Larch and Y. Yotov (2021), "Gravity Estimations with Interval Data: Revisiting the Impact of Free Trade Agreements", *Economica*, Vol. 89/353, pp. 44-61, https://doi.org/10.1111/ecca.12394. [28]

Egger, P. and A. Shingal (2020), "Determinants of services trade agreement membership", *Review of World Economics*, Vol. 157/1, pp. 21-64, https://doi.org/10.1007/s10290-020-00394-y. [70]

Evenett, S. and J. Fritz (2021), *Digital Policy Alert*, https://digitalpolicyalert.org/?url=digital_policy. [94]

Fernandes, A. and C. Paunov (2012), "Foreign direct investment in services and manufacturing productivity: Evidence for Chile", *Journal of Development Economics*, Vol. 97/2, pp. 305-321, https://doi.org/10.1016/j.jdeveco.2011.02.004. [49]

Ferracane, M. (2022), *Digital Trade Integration Database*, https://dti.eui.eu/. [95]

Ferracane, M., H. Lee-Makiyama and E. van der Marel (2018), *Digital Trade Restrictiveness Index*, https://ecipe.org/wp-content/uploads/2018/05/DTRI_FINAL.pdf. [96]

Feyrer, J. (2019), "Trade and Income—Exploiting Time Series in Geography", *American Economic Journal: Applied Economics*, Vol. 11/4, pp. 1-35, https://doi.org/10.1257/app.20170616. [53]

Findlay, C. and T. Warren (eds.) (1990), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge. [7]

Fink, C., A. Mattoo and I. Neagu (2005), "Assessing the impact of communication costs on international trade", *Journal of International Economics*, Vol. 67/2, pp. 428-445, https://doi.org/10.1016/j.jinteco.2004.09.006. [58]

Fort, T. (2016), "Technology and Production Fragmentation: Domestic versus Foreign Sourcing", *The Review of Economic Studies*, p. rdw057, https://doi.org/10.1093/restud/rdw057. [60]

Francois, J. (1990), "PRODUCER SERVICES, SCALE, AND THE DIVISION OF LABOR", *Oxford Economic Papers*, Vol. 42/4, pp. 715-729, https://doi.org/10.1093/oxfordjournals.oep.a041973. [43]

Francois, J. and B. Hoekman (2010), "Services Trade and Policy", *Journal of Economic Literature*, Vol. 48/3, pp. 642-692, https://doi.org/10.1257/jel.48.3.642. [45]

Gal, P. and W. Witheridge (2019), "Productivity and innovation at the industry level: What role for integration in global value chains?", *OECD Productivity Working Papers*, No. 19, OECD Publishing, Paris, https://doi.org/10.1787/a5cec52c-en. [51]

Geloso Grosso, M. et al. (2015), "Services Trade Restrictiveness Index (STRI): Scoring and Weighting Methodology", *OECD Trade Policy Papers*, No. 177, OECD Publishing, Paris, https://doi.org/10.1787/5js7n8wbtk9r-en. [8]

Grundke, R. and J. Arnold (2022), "Mastering the transition: A synthetic literature review of trade adaptation policies", *OECD Economics Department Working Papers*, No. 1719, OECD Publishing, Paris, https://doi.org/10.1787/5fad3487-en. [33]

Hannan, S. (2017), *The Impact of Trade Agreements in Latin America using the Synthetic Control Method*, https://www.imf.org/-/media/Files/Publications/WP/2017/wp1745.ashx. [84]

Hidalgo, C. and R. Hausmann (2009), "The building blocks of economic complexity", *Proceedings of the National Academy of Sciences*, Vol. 106/26, pp. 10570-10575, https://doi.org/10.1073/pnas.0900943106. [72]

Hoekman, B. and A. Mattoo (2008), "Services trade and growth", *World Bank Policy Research Working Paper*, Vol. WPS 4461. [87]

Hoekman, B. and B. Shepherd (2015), "Services Productivity, Trade Policy and Manufacturing Exports", *The World Economy*, Vol. 40/3, pp. 499-516, https://doi.org/10.1111/twec.12333. [92]

Hoffman, I. and E. Mast (2019), "Heterogeneity in the effect of federal spending on local crime: Evidence from causal forests", *Regional Science and Urban Economics*, https://www.sciencedirect.com/science/article/pii/S0166046219300122 (accessed on 3 December 2019). [37]

Holland, P. (1986), "Statistics and Causal Inference", *Journal of the American Statistical Association*, Vol. 81/396, pp. 945-960, https://doi.org/10.2307/2289064. [69]

Hovhannisyan, N. and W. Keller (2014), "International business travel: an engine of innovation?", *Journal of Economic Growth*, Vol. 20/1, pp. 75-104, https://doi.org/10.1007/s10887-014-9107-7. [54]

Hummels, D. (2007), "Transportation Costs and International Trade in the Second Era of Globalization", *Journal of Economic Perspectives*, Vol. 21/3, pp. 131-154, https://doi.org/10.1257/jep.21.3.131. [52]

International Monetary Fund (2009), *Balance of Payments and International Investment Position Manual - 6th edition*. [79]

Johannemann, J. et al. (2021), *Sufficient Representations for Categorical Variables*. [83]

Khachaturian, T. and S. Oliver (2021), *The Role of "Mode Switching" in Services Trade*. [12]

Kleven, H., C. Landais and E. Saez (2013), "Taxation and international migration of superstars: Evidence from the European football market", *American Economic Review*, Vol. 103/5, pp. 1892-1924. [100]

Krautheim, S. (2012), "Heterogeneous firms, exporter networks and the effect of distance on international trade", *Journal of International Economics*, Vol. 87/1, pp. 27-35, https://doi.org/10.1016/j.jinteco.2011.11.004. [75]

Liu, X. et al. (2020), "Services development and comparative advantage in manufacturing", *Journal of Development Economics*, Vol. 144, p. 102438, https://doi.org/10.1016/j.jdeveco.2019.102438. [42]

Low, P. (2013), "The role of services in global value chains", in *Global Value Chains in a Changing World*, World Trade Organization, Geneva/Temasek Foundation Centre for Trade & Negotiations (TFCTN), https://doi.org/10.30875/8d564dca-en. [1]

Lucas, R. and E. Prescott (1971), "Investment Under Uncertainty", *Econometrica*, Vol. 39/5, p. 659, https://doi.org/10.2307/1909571. [30]

Manova, K., S. Wei and Z. Zhang (2015), "Firm Exports and Multinational Activity Under Credit Constraints", *Review of Economics and Statistics*, Vol. 97/3, pp. 574-588, https://doi.org/10.1162/rest_a_00480. [62]

Markusen, J., T. Rutherford and D. Tarr (2005), "Trade and direct investment in producer services and the domestic market for expertise", *Canadian Journal of Economics/Revue canadienne d'<html_ent glyph="@eacute;" ascii="e"/>conomique*, Vol. 38/3, pp. 758-777, https://doi.org/10.1111/j.0008-4085.2005.00301.x. [46]

Mattoo, A., N. Rocha and M. Ruta (2020), *Handbook of Deep Trade Agreements*, World Bank. [31]

Mion, G. and L. Opromolla (2014), "Managers' mobility, trade performance, and wages", *Journal of International Economics*, Vol. 94/1, pp. 85-101, https://doi.org/10.1016/j.jinteco.2014.06.001. [78]

Mion, G., L. Opromolla and A. Sforza (2022), "The Value of Managers' Export Experience: Lessons from the Angolan Civil War", *Accepted at the Review of Economics and Statistics*, https://drive.google.com/open?id=1GI97fPtQpv5Np4td3LfWtPXBKngxgf07. [76]

Miroudot, S. and C. Cadestin (2017), "Services In Global Value Chains: From Inputs to Value-Creating Activities", *OECD Trade Policy Papers*, No. 197, OECD Publishing, Paris, https://doi.org/10.1787/465f0d8b-en. [2]

Mitze, T. et al. (2020), "Face masks considerably reduce COVID-19 cases in Germany", *Proceedings of the National Academy of Sciences*, Vol. 117/51, pp. 32293-32301. [101]

Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, Vol. 49/6, p. 1417, https://doi.org/10.2307/1911408. [80]

Nordås, H. (2011), "Opening the Markets for Business Services: Industrial Perspective for Developing Countries", *Journal of Economic Integration*, Vol. 26/2, pp. 306-328. [47]

Nordås, H. et al. (2014), "Services Trade Restrictiveness Index (STRI): Computer and Related Services", *OECD Trade Policy Papers*, No. 169, OECD Publishing, Paris, https://doi.org/10.1787/5jxt4np1pjzt-en. [82]

Nordås, H. and Y. Kim (2013), "The Role of Services for Competitiveness in Manufacturing", *OECD Trade Policy Papers*, No. 148, OECD Publishing, Paris, https://doi.org/10.1787/5k484xb7cx6b-en. [24]

Nordås, H. and D. Rouzet (2015), "The Impact of Services Trade Restrictiveness on Trade Flows: First Estimates", *OECD Trade Policy Papers*, No. 178, OECD Publishing, Paris, https://doi.org/10.1787/5js6ds9b6kjb-en. [9]

OECD (2021), *OECD Services Trade Restrictiveness Index: Policy Trends up to 2021*. [3]

O'Malley, T. (2018), "The Impact of Repossession Risk on Mortgage Default", Central Bank of Ireland, https://www.centralbank.ie/news/article/adverse-selection-and-moral-hazard-in-forecasting-and-limiting-arrears-and- (accessed on 3 December 2019). [35]

Rajan, R. and L. Zingales (1998), "Financial Dependence and Growth", *American Economic Review*, Vol. 88/3, pp. 559-586. [91]

Reverdy, C. (2022), "Estimating the general equilibrium effects of services trade liberalization", *Review of International Economics*, https://doi.org/10.1111/roie.12635. [13]

Robert-Nicoud, F. (2008), “Offshoring of routine tasks and (de)industrialisation: Threat or opportunity—And for whom?”, *Journal of Urban Economics*, Vol. 63/2, pp. 517-535, https://doi.org/10.1016/j.jue.2007.03.002. [59]

Roberts, M. and J. Tybout (1997), *What Markets Exports Boom?*, The International Bank for Reconstruction/The World Bank, Washington D.C. [29]

Robins, J., A. Rotnitzky and L. Zhao (1994), “Estimation of Regression Coefficients When Some Regressors are not Always Observed”, *Journal of the American Statistical Association*, Vol. 89/427, pp. 846-866, https://doi.org/10.1080/01621459.1994.10476818. [63]

Roodman, D. (2009), “How to do Xtabond2: An Introduction to Difference and System GMM in Stata”, *The Stata Journal: Promoting communications on statistics and Stata*, Vol. 9/1, pp. 86-136, https://doi.org/10.1177/1536867x0900900106. [81]

Rouzet, D., S. Benz and F. Spinelli (2017), “Trading firms and trading costs in services: Firm-level analysis”*, OECD Trade Policy Papers*, No. 210, OECD Publishing, Paris, https://doi.org/10.1787/b1c1a0e9-en. [10]

Rouzet, D. and F. Spinelli (2016), “Services Trade Restrictiveness, Mark-Ups and Competition”, *OECD Trade Policy Papers*, No. 194, OECD Publishing, Paris, https://doi.org/10.1787/5jln7dlm3931-en. [15]

Rubin, D. (1974), “Estimating causal effects of treatments in randomized and nonrandomized studies.”, *Journal of Educational Psychology*, Vol. 66/5, pp. 688-701, https://doi.org/10.1037/h0037350. [68]

Sala, D. and E. Yalcin (2014), “Export Experience of Managers and the Internationalisation of Firms”, *The World Economy*, Vol. 38/7, pp. 1064-1089, https://doi.org/10.1111/twec.12222. [74]

Tanaka, K. (2019), “Do international flights promote FDI? The role of face-to-face communication”, *Review of International Economics*, Vol. 27/5, pp. 1609-1632, https://doi.org/10.1111/roie.12437. [55]

Tiffin, A. (2019), “Machine Learning and Causality: The Impact of Financial Crises on Growth”, https://www.imf.org/en/Publications/WP/Issues/2019/11/01/Machine-Learning-and-Causality-The-Impact-of-Financial-Crises-on-Growth-48722 (accessed on 22 November 2019). [36]

Wager, S. and S. Athey (2018), “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests”, *Journal of the American Statistical Association*, Vol. 113/523, pp. 1228-1242, https://doi.org/10.1080/01621459.2017.1319839. [66]

Yiqing, X. (2017), “Generalized synthetic control method: Causal inference with interactive fixed effects models”, *Political Analysis*, Vol. 25/1, pp. 57-76. [103]

Annex A. Short-run gravity modelling

Empirical specification

The standard gravity model without pair fixed effects is specified as

$$X_{ijt,k} = \exp\left(\beta_1 BRDR_{ij} + \beta_2 G_{ij} BRDR_{ij} + \beta_3 STRI_{ij} BRDR_{ij} + \beta_4 RTA_{S_{ij}} + \beta_5 RTA_{G_{ij}} + \eta_{it,k} + \mu_{jt,k}\right) + \varepsilon_{ijt,k}$$

where $X_{ijt,k}$ indicates bilateral cross-border services trade from exporter *i* to importer *j* in year *t* and sector *k*. BRDR is a cross-border dummy that is equal to one when *i* and *j* refer to different countries and equal to zero when *i* and *j* refer to the same country. G_{ij} is a set of gravity control variables, including bilateral distance, contiguity, common language, common religion, common legal system, previous colonial ties and EEA membership. $STRI_{ij}$ indicates the importer's STRI score, while it is the importer's intra-EEA STRI score, if both *i* and *j* are members of the European Economic Area. $RTA_{S_{ij}}$ and $RTA_{G_{ij}}$ are dummy variables indicating whether two trading partners are party to a regional trade agreement covering services or goods, respectively.

Exporter-year fixed effects $\eta_{it,k}$ and importer-year fixed effects $\mu_{jt,k}$ control for multilateral resistance to cross-border services trade. The error term $\varepsilon_{ijt,k}$ might be heteroskedastic, which is taken into account using the PPML estimator.

A lagged dependent variable is added as additional regressor in the gravity specification for the estimation of the short-run model. The estimation equation can be written as

$$X_{ijt,k} = \exp\left(\beta_0 X_{ij,t-1,k} + \beta_1 BRDR_{ij} + \beta_2 G_{ij} BRDR_{ij} + \beta_3 STRI_{ij} BRDR_{ij} + \beta_4 RTA_{S_{ij}} + \beta_5 RTA_{G_{ij}} + \eta_{it,k} + \mu_{jt,k}\right) + \varepsilon_{ijt,k}$$

The gravity model with pair fixed effects does not include a border dummy or standard gravity variables, because both are perfectly multicollinear with the symmetric pair fixed effects. This specification can be written as

$$X_{ijt,k} = \exp\left(\beta_3 STRI_{ij} BRDR_{ij} + \beta_4 RTA_{S_{ij}} + \beta_5 RTA_{G_{ij}} + \vartheta_{ij,k} + \eta_{it,k} + \mu_{jt,k}\right) + \varepsilon_{ijt,k}$$

where $\vartheta_{ij,k}$ is the new symmetric pair fixed effect indicating trade between exporter *i* and importer *j* and between exporter *j* and importer *i*, respectively.

The estimation equation for the short-run gravity model with pair fixed effects can be written as

$$X_{ijt,k} = \exp\left(\beta_0 X_{ij,t-1,k} + \beta_3 STRI_{ij} BRDR_{ij} + \beta_4 RTA_{S_{ij}} + \beta_5 RTA_{G_{ij}} + \vartheta_{ij,k} + \eta_{it,k} + \mu_{jt,k}\right) + \varepsilon_{ijt,k}$$

Sector coverage and definition

The analysis covers five services sectors. Communications services refers to item SI of the EBOPS (Extended Balance of Payments Services) 2010 classification. The sector includes telecommunications services, computer services (including software) and information services, such as news agency services. Financial services (item SG) are recorded separately from insurance and pension services (item SF). Insurance covers direct insurance, reinsurance, auxiliary insurance services, as well as pension and standardised guarantee services.

Business services refer to item SJ of the EBOPS 2010 classification. This item includes research and development services, professional and management consulting services, as well as technical and trade related services. It is the largest category of services trade included in this analysis, with global flows of more than USD 1.4 trillion in 2019. Transport services are recorded under item SC of the EBOPS 2010 classification. This item includes sea transport, air transport, other modes of transport, as well as postal and courier services.

A correspondence of EBOPS 2010 sectors to sectoral production data is established for the construction of domestic trade flows (Tables A A.1 and A A.2) contains a detailed list of activities included in each sector, based on the ISIC rev.4 classification.

Table A A.1. Sector correspondence

Sector	EBOPS 2010	ISIC rev. 4	STRI
Business services	SJ	69-82	PSacc, PSarc, PSeng, PSleg°
Communication	SI	58-63	CS + TC°
Financial services	SG	64 + 66*	FSbnk
Insurance	SF	65 + 66**	FSins
Transport	SC	49-53	TRair, TRmar, TRrai, TRrof°°

Note: Transport includes water transport, air transport and transport not elsewhere classified. Not covered are electricity; gas manufacture, distribution; water; construction; trade; recreational and other services; public administration, defence, health, education; dwellings. * Two-thirds of ISIC 66 (Activities auxiliary to financial service and insurance activities) is attributed to the financial services. ** One-third of ISIC 66 (Activities auxiliary to financial service and insurance activities) is attributed to the insurance services. ° Simple average. °° Simple average of all transport STRIs available for a country (TRmar is not available for landlocked countries).
Source: Authors' compilation.

Table A A.2. Sector classification by reference to ISIC rev. 4

Sector	ISIC rev. 4	Description
Business services	69-82	Professional, scientific and technical activities and Administrative and support service activities
Communication	58	Publishing activities
	59	Motion picture, video and television programme production, sound recording and music publishing activities
	60	Programming and broadcasting activities
	61	Telecommunications
	62	Computer programming, consultancy and related activities
	63	Information service activities
Financial services	64	Financial service activities, except insurance and pension funding
	661	Activities auxiliary to financial service activities, except insurance and pension funding
	663	Fund management activities
Insurance	65	Insurance, reinsurance and pension funding, except compulsory social security
	662	Activities auxiliary to insurance and pension funding
Transport	49	Land transport and transport via pipelines
	50	Water transport
	51	Air transport
	52	Warehousing and support activities for transportation
	53	Postal and courier activities

Source: Authors' compilation based on ISIC rev. 4 classification.

Main regression results

Table A A.3. Standard gravity regression results for all sectors

	Communication	Finance	Insurance	Business	Transport
Ln distance	-0.673***	-0.883***	-0.309**	-0.709***	-0.606***
	(0.073)	(0.118)	(0.137)	(0.073)	(0.081)
Contiguity	-0.144	-0.520	0.301	0.062	0.437***
	(0.164)	(0.346)	(0.442)	(0.177)	(0.151)
Common language	0.881***	1.431***	1.075***	0.719***	0.350**
	(0.152)	(0.355)	(0.329)	(0.156)	(0.161)
Common religion	-0.011	0.096	0.735	0.166	0.122
	(0.235)	(0.517)	(0.490)	(0.230)	(0.215)
Common legal system	0.164	0.162	0.289	0.228**	0.095
	(0.106)	(0.301)	(0.219)	(0.109)	(0.116)
Previous colonial ties	-1.047***	-2.370***	-1.468***	-0.494	0.067
	(0.360)	(0.607)	(0.522)	(0.369)	(0.233)
STRI	-5.776***	-7.884***	-6.442***	-4.701***	-3.634***
	(1.050)	(2.198)	(1.123)	(0.874)	(1.250)
Services RTA	-0.262*	-0.828***	-0.444	-0.518***	-0.206
	(0.139)	(0.282)	(0.312)	(0.191)	(0.210)
Goods RTA	-1.546***	-1.673***	-1.823***	-2.062***	-0.990***
	(0.241)	(0.568)	(0.592)	(0.176)	(0.301)
Observations	9,569	8,909	8,666	9,737	9,897
Exporter time F.E.	YES	YES	YES	YES	YES
Importer time F.E.	YES	YES	YES	YES	YES
Time-varying border	YES	YES	YES	YES	YES

Note: Standard errors clustered by country-pair in parentheses *** p<0.01, ** p<0.05, * p<0.1. Additional dummy for EEA membership included but not reported in columns (1) and (2).

Table A A.4. Short-run gravity regression results for all sectors

	Communication	Finance	Insurance	Business	Transport
Ln distance	-0.036*	-0.021	-0.030	-0.055*	-0.045***
	(0.021)	(0.025)	(0.024)	(0.030)	(0.012)
Contiguity	-0.056*	-0.059**	-0.031	-0.026	0.020
	(0.032)	(0.026)	(0.065)	(0.029)	(0.015)
Common language	0.100***	0.092**	0.096*	0.085***	-0.002
	(0.038)	(0.038)	(0.057)	(0.029)	(0.017)
Common religion	0.027	-0.018	-0.024	0.046	0.021
	(0.050)	(0.049)	(0.072)	(0.040)	(0.023)
Common legal system	0.010	0.012	0.033	0.039*	-0.002
	(0.022)	(0.026)	(0.044)	(0.021)	(0.014)
Previous colonial ties	-0.146***	-0.090	-0.225***	-0.075	-0.066***
	(0.050)	(0.068)	(0.071)	(0.119)	(0.026)
STRI	-0.852***	-0.999***	-1.441***	-0.595***	-0.288**
	(0.306)	(0.245)	(0.289)	(0.173)	(0.132)
Services RTA	0.020	-0.029	-0.126***	-0.041*	-0.019
	(0.037)	(0.032)	(0.041)	(0.025)	(0.018)
Goods RTA	-0.129*	-0.020	-0.202**	-0.288***	0.034
	(0.069)	(0.064)	(0.092)	(0.070)	(0.052)
Lag ln exports	0.911***	0.943***	0.911***	0.901***	0.937***
	(0.026)	(0.022)	(0.019)	(0.026)	(0.012)
Observations	9,569	8,909	8,666	9,737	9,897
Exporter time F.E.	YES	YES	YES	YES	YES
Importer time F.E.	YES	YES	YES	YES	YES
Time-varying border	YES	YES	YES	YES	YES

Note: Standard errors clustered by country-pair in parentheses *** p<0.01, ** p<0.05, * p<0.1. Additional dummy for EEA membership included but not reported in columns (1) and (2).

Robustness checks and further analysis

Specification with pair fixed effects

Symmetric pair fixed effects control for all time-invariant determinants of bilateral trade, including those that are unobservable, such as consumer-preferences for specific services.[47] Standard dyadic gravity variables cannot be included in this specification. Their only source of variation is on the exporter-importer dimension, implying multi-collinearity with pair fixed effects. This approach also addresses potential endogeneity of RTAs, strengthening the interpretation of the coefficient as causal effect of trade agreements on bilateral trade flows (Baier and Bergstrand, 2007[27]).

However, this specification only relies on variation over time for the identification of regression coefficients. Permanent characteristics of the relationship between two specific trading partners are captured by the pair fixed effects. This complicates the identification of regression coefficients pertaining to variables with relatively little variation over time, such as the STRI. In general, regression coefficients are subject to large uncertainty, indicated by large standard errors and correspondingly wide confidence intervals, so that most regression coefficients are not statistically different from zero.

[47] In addition, this approach mitigates the dynamic panel bias (Roodman, 2009[81]; Anderson and Yotov, 2020[21]).

Table A A.5. Gravity regression results with pair fixed effects for all sectors

	Communication	Finance	Insurance	Business	Transport
STRI	-4.925**	-1.899	2.834	-0.960	0.013
	(2.095)	(3.019)	(2.971)	(0.755)	(1.357)
Services RTA	0.497**	0.128*	0.336	0.038	-0.168***
	(0.228)	(0.070)	(0.237)	(0.065)	(0.050)
Goods RTA	0.332	0.631***	-0.669	0.121*	0.199
	(0.213)	(0.104)	(1.492)	(0.065)	(0.297)
Observations	9,569	8,909	8,666	9,737	9,897
Exporter time F.E.	YES	YES	YES	YES	YES
Importer time F.E.	YES	YES	YES	YES	YES
Time-varying border	YES	YES	YES	YES	YES
Symmetric pair F.E.	YES	YES	YES	YES	YES

Note: Standard errors clustered by country-pair in parentheses *** p<0.01, ** p<0.05, * p<0.1. Additional dummy for EEA membership included but not reported in columns (1) and (2).

Table A A.6. Short-run regression results with pair fixed effects for all sectors

	Communication	Finance	Insurance	Business	Transport
STRI	-2.565***	-1.567	1.653	-0.521	0.083
	(0.988)	(1.824)	(1.333)	(0.446)	(0.546)
Services RTA	0.172***	0.046	-0.172	0.052	-0.063**
	(0.061)	(0.053)	(0.146)	(0.049)	(0.032)
Goods RTA	-0.036	0.369***	-0.584	0.016	0.114
	(0.092)	(0.123)	(0.826)	(0.056)	(0.126)
Lag ln exports	0.551***	0.393**	0.575***	0.397***	0.632***
	(0.083)	(0.159)	(0.046)	(0.116)	(0.059)
Observations	9,299	8,202	7,780	9,524	9,722
Exporter time F.E.	YES	YES	YES	YES	YES
Importer time F.E.	YES	YES	YES	YES	YES
Time-varying border	YES	YES	YES	YES	YES
Symmetric pair F.E.	YES	YES	YES	YES	YES

Note: Standard errors clustered by country-pair in parentheses *** p<0.01, ** p<0.05, * p<0.1. Additional dummy for EEA membership included but not reported in columns (1) and (2).

Results show that regression coefficients of the STRI variable are statistically significant for communications services. In this sector, a reduction of services trade barriers measured in the STRI by 0.05 points can lead to an increase in bilateral trade of 28% in the standard model and around 14% in the short-run model, so that around half of the overall effect would already materialise in the short term. Regression coefficients in all other sectors are statistically insignificant.

Results from the specification with pair fixed effects can also identify the impact of RTAs on trade flows in the short run relative to the standard model. Not all resulting regression coefficients are statistically different from zero. A services RTA can boost cross-border trade in communication services by around 65% in the standard model, whereas the effect in the short run is only around 19%. The estimated regression coefficient for financial services implies an increase of bilateral services trade after the entry into force of a services RTA of around 14%. The corresponding short-run impact is somewhat smaller and not statistically significant.

No significant coefficients can be identified for business services and insurance services. In the transport sector, services RTAs even seem to reduce cross-border services trade. Relying on variation over time for the identification of the services RTA effect implies that only agreements having entered into force between 2015 and 2019 are employed for the identification of the regression coefficient. For the sample of countries used in this study, only seven RTAs enter into force during this period, primarily towards the end of the period. A short observation period after the policy change means that the results from the standard model might be biased because the trade-creating effect of an RTA only materialises with a certain time lag.

Moreover, these agreements often focus on digitally deliverable services such as finance, professional services, or telecommunication rather than transport services. In addition, there could be an expansion of Mode 3 services trade that is not captured in the balance-of-payment flow used for this analysis. These results are at odds with existing evidence, showing a positive relationship between RTAs and cross-border trade in transport services (Benz and Jaax, 2020[5])

The relative low number of RTAs used for identification also implies that it is not possible to take into account potential heterogeneity of RTAs, e.g. relating to the extent of commitments in an RTA. Overall, however, coefficients support the understanding that only a share of the overall gains from services liberalisation materialises in the short term.

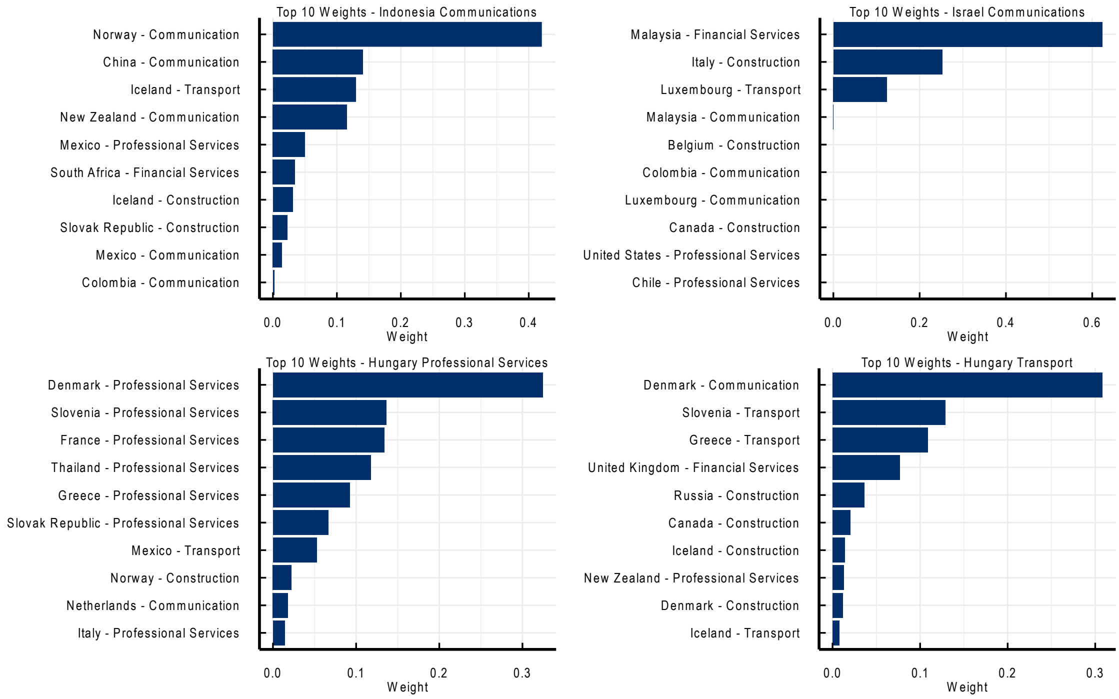
Heterogeneity between more liberal and more restrictive economies

Heterogeneity between more liberal and more restrictive economies can be analysed in various way. In a first attempt, we split the sample depending on a country's services restrictiveness in 2014. Imports by countries more liberal than the median are analysed separately from imports of countries more restrictive than the median. In a second attempt, we use a square term of the STRI in addition to the standard linear term. A significant regression coefficient on the square term can indicate whether the trade-creating impact of services liberalisation is more or less pronounced for more restrictive economies.

Neither of the two strategies can provide clear results regarding the heterogeneous impact of services liberalisation between more liberal and more restrictive economies. In strategy 1, A comparison of regression coefficients from the two samples usually shows that these coefficients are unlikely to be statistically significant from each other. Regression coefficients on the square STRI term in strategy 2 are sometimes significant but exhibit different signs across sectors, so that no unambiguous conclusions can be drawn.

Annex B. Synthetic control method

Figure A B.1. Weights used in synthetic control construction



Note: These graphs show, for each of the four case studies using the synthetic control method, the weights assigned to other countries and sectors to construct the synthetic comparison case. Weights are constrained to be non-negative and sum to one, and are calculated to minimise mean squared predictive error. For each case study, the ten largest weights are displayed.
Source: Authors' calculations.

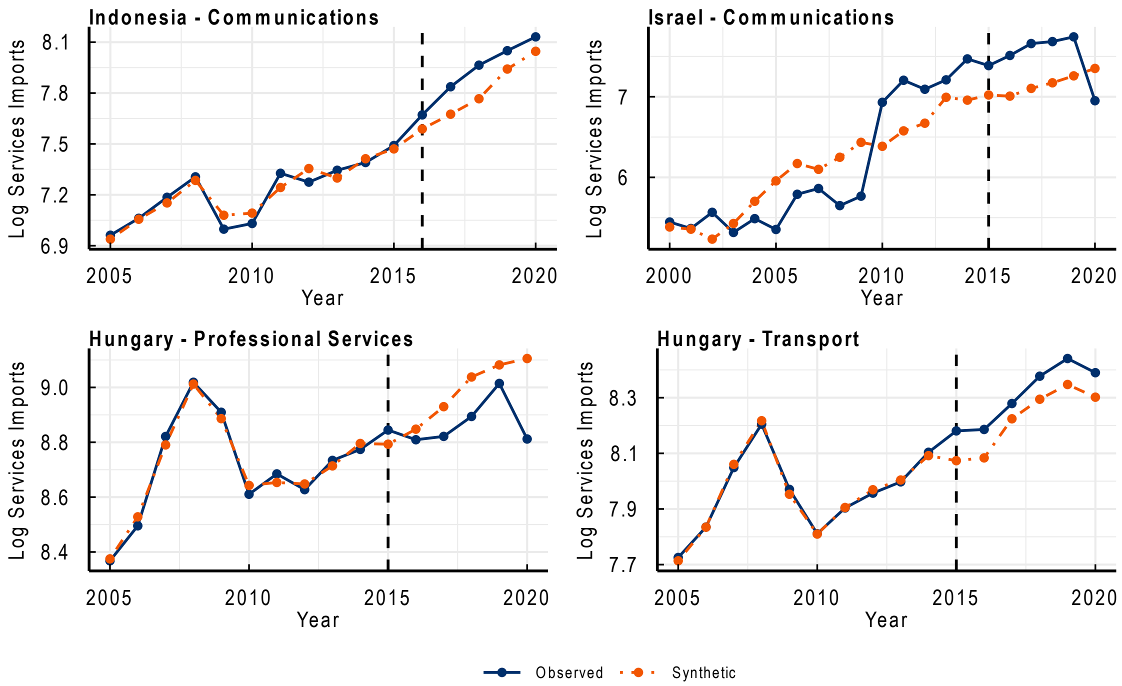
Table A B.1. Inclusion of countries and sectors in weight calculations

Country	Communication	Construction	Financial Services	Professional Services	Retail	Transport
Australia	(a)(c)	(a)(c)(d)	(a)(c)	(a)(c)(d)	(a)(c)(d)	(a)(c)
Austria	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Belgium	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Brazil	(a)(c)	(a)(c)	(a)	(a)(c)(d)	(a)(c)	(a)(c)
Canada	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Chile	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
China	(a)			(a)		(a)
Colombia	(a)(b)(c)(d)	(a)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(c)(d)	(a)(c)(d)
Costa Rica	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Czech Republic	(a)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(c)(d)
Denmark	(a)(c)	(a)(c)(d)	(a)(b)(c)(d)	(a)(c)	(a)(b)(c)(d)	(a)(b)(c)(d)
Estonia	(a)(c)	(a)(c)	(a)(c)	(a)(c)	(a)(c)(d)	(a)(c)
Finland	(a)(c)	(a)	(a)(c)	(a)(c)	(a)(c)	(a)(c)
France	(a)(c)	(a)	(a)(c)	(a)(c)	(a)(c)(d)	(a)(c)
Germany	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Greece	(a)(c)(d)	(a)(c)(d)	(a)(c)	(a)(c)	(a)(b)(c)(d)	(a)(c)(d)
Iceland	(a)(b)(c)(d)	(a)(c)(d)	(a)	(a)(c)(d)	(a)(c)(d)	(a)(c)(d)
India	(a)	(a)	(a)	(a)(b)(c)(d)	(a)	(a)
Ireland	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Italy	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(c)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Japan	(a)(c)(d)	(a)(c)	(a)(c)	(a)(b)(c)(d)	(a)(c)	(a)(c)(d)
Kazakhstan	(a)	(a)		(a)	(a)	(a)
Korea	(a)(c)(d)	(a)(c)(d)	(a)(c)(d)	(a)(b)(c)(d)	(a)(c)	(a)(c)
Latvia	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Lithuania	(a)(c)	(a)(c)	(a)(c)	(a)(c)(d)	(a)(b)(c)(d)	(a)(c)(d)
Luxembourg	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Malaysia	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Mexico	(a)(c)	(a)	(a)(c)	(a)(b)(c)(d)	(a)(c)(d)	(a)(c)(d)
Netherlands	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
New Zealand	(a)	(a)	(a)(c)	(a)(c)	(a)(c)	(a)
Norway	(a)	(a)(c)	(a)(c)(d)	(a)(c)	(a)	(a)
Peru	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Poland	(a)	(a)	(a)(c)(d)	(a)(b)(c)(d)	(a)	(a)(c)
Portugal	(a)(c)		(a)(c)(d)	(a)	(a)(c)	(a)(c)
Russia	(a)	(a)(b)(c)(d)	(a)	(a)(c)(d)	(a)(c)	(a)(c)(d)
Singapore	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)	(a)(b)(c)(d)	(a)(b)(c)(d)

Slovak Republic	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Slovenia	(a)(c)(d)	(a)(c)	(a)(c)(d)	(a)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
South Africa	(a)(c)	(a)	(a)(c)	(a)(c)	(a)(c)	(a)(c)
Spain	(a)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(c)(d)
Sweden	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Switzerland	(a)(b)(c)(d)	(a)(c)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Thailand	(a)(c)	(a)	(a)(c)	(a)(c)(d)	(a)(c)	(a)(c)
Türkiye	(a)	(a)	(a)	(a)	(a)	(a)
United Kingdom	(a)(c)	(a)(c)(d)	(a)(c)	(a)(c)	(a)(c)(d)	(a)
United States	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)
Viet Nam	(a)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)(b)(c)(d)	(a)	(a)

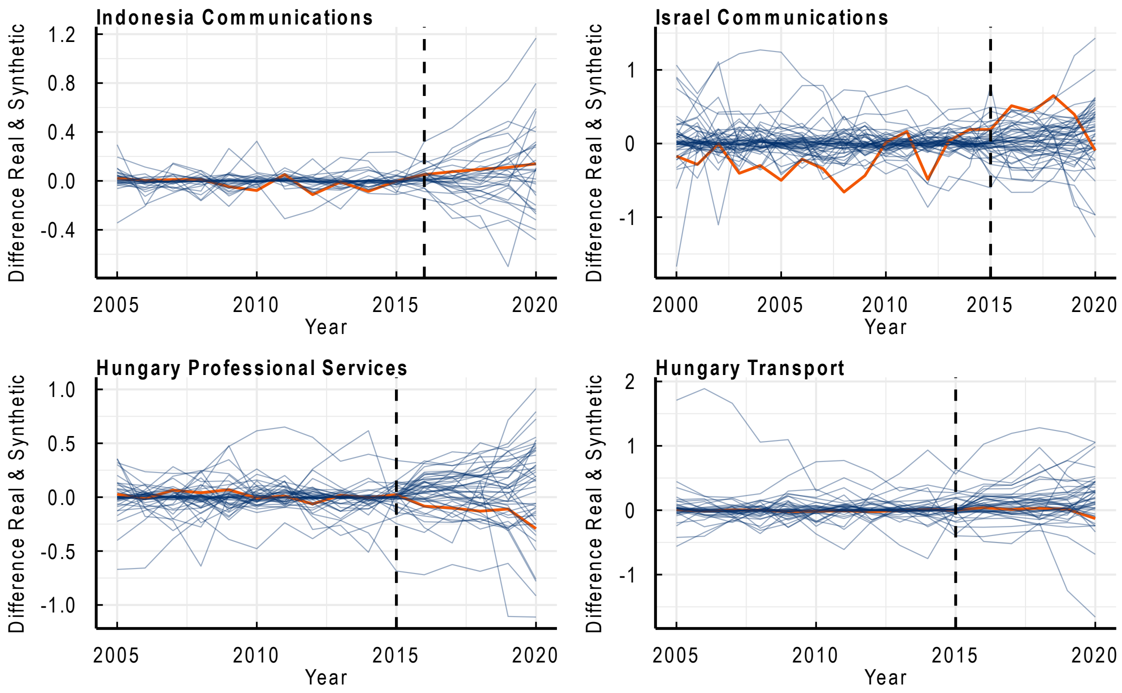
Note: Each cell indicates whether a given country and sector was included as a potential control case in the construction of the synthetic observations. Letters refer to the four cases: (a) Indonesia Communications, (b) Israel Communications, (c) Hungary Professional Services, (d) Hungary Transport. Inclusion does not mean the sector received a non-zero weight; see Figure A B.1 for the top 10 largest weights for each case.
Source: Authors' calculations.

Figure A B.2. Results restricting sectors



Note: These graphs show the results when applying the synthetic control method, bur restricting the control sectors used for comparisons to be the same economic sector as the case under investigation.
Source: Authors' calculations.

Figure A B.3. Placebo tests for synthetic control method



Note: These graphs show the outcome of a placebo-based robustness check for the synthetic control analyses in the paper. The idea of the placebo check is to see if an effect can be observed in instances where there was no real "treatment", no policy reform. For each of the sectors and countries included as potential contributors to the synthetic control the same analysis as the main case is conducted as if a change in policy had taken place. The difference between the observed trend and the synthetic control for these alternate cases is plotted here, showing the "placebo effect". In each plot, the real case is displayed in orange. Comparing this real treatment effect to the blue lines showing the outcome of the placebo cases shows the extent to which the found effect is exceptional and can be taken to be meaningful. For the three cases where an effect was identified, the results show that while larger treatment effects can be found in a few other instances, on the whole the cases of interest do stand out. For the one case of null results (Hungary's transport services sector), it is clear that most other cases show a bigger effect, confirming the conclusion that an effect could have been but was not observed.
Source: Authors' calculations.

Annex C. Estimating the causal effect of services trade reforms using machine learning based methods

This annex provides a technical report on the application of Machine Learning (ML)-based methods for causal inference to the problem of estimating the effect of services trade reforms using the STRI database.[48]

In general, ML algorithms cannot be directly applied to solve a causal inference problem. Indeed, the primary goal of many ML tools is to solve a prediction problem, i.e. to correctly predict the value of a target variable given observable features (for instance predicting whether an email is to be classified as spam or not, based on its subject, sending address and main text). To do so, ML relies on flexible functional forms for prediction models, accompanied by data-driven rules to limit their expressiveness and avoid overfitting. Validation procedures are at the core of ML solutions to prediction problems: they allow algorithms to learn how to make better predictions by repeatedly comparing their output with the actual realisation of the data.

In a causal estimation problem instead, the underlying goal is to make inference on the true relationship between a treatment and an outcome variable. In applications tackling the research questions raised in this paper, the treatment variable would be an indicator capturing relevant dimensions of services trade policy reforms and the outcome variable would be a specific economic indicator, such as sectoral trade flows or GDP per capita.

A standard ML algorithm for predicting the value of services trade does not necessarily offer valuable insights on the true relationship between trade flows and the variables used for prediction, including possibly indicators of services trade policy. In fact, the very features that make an ML model successful at solving prediction problems, are the same that make it particularly hard to apply the model for causal estimation. The use of complex functional forms, regularization rules and data-driven tuning usually does not allow to understand the relationship between variables and causes non6negligible biases in the estimates.

A successful approach to use ML tools to help solving a causal inference problem is to break down the latter into different steps and to use ML to solve those that look like prediction tasks and that allow validation procedures to be applied.

Examples of this approach are average treatment effect estimation strategies based on the Augmented Inverse Propensity Weighted (AIPW) estimator (Robins, Rotnitzky and Zhao, 1994[63]). Within the AIPW estimation pipeline, off-the-shelf ML-algorithms can be deployed to estimate the ingredients of the final estimator. Under considerable generality, these methods have optimal statistical properties that preserve all the guarantees required by causal inference, including unbiasedness, consistency, and the possibility to build confidence intervals around the estimates of the causal parameters (Chernozhukov et al., 2018[64]).

Another effective approach consists in modifying existing ML algorithms to make them directly applicable to causal inference problems. Remarkable examples in this space are the Causal Tree and the Causal Forest algorithms (Athey and Imbens, 2016[65]; Wager and Athey, 2018[66]). Specifically designed to assess and characterise treatment effect heterogeneity, these models adapt the standard regression tree and the random forest algorithm (Breiman, 2001[67]) to estimate average treatment effects in subsets of the empirical population defined in terms of observable features.

This annex contributes to the main synthesis paper in two ways. First, it offers a suggested routine for ML-based causal inference applied to the problem of estimating the effect of services trade reforms using the

[48] The work presented in this Annex benefitted from the activities of the OECD Study Group on Causal Inference and Machine Learning coordinated by the OECD Trade and Agriculture Directorate and the OECD New Approaches to Economic Challenges (NAEC) Innovation Lab and supervised by Stefan Wager. The main resources used by the Study Group consisted in the 2021 tutorial on ML-based casual inference prepared by the Golub Capital Social Impact Lab of the Stanford Graduate School of Business available at https://bookdown.org/halflearned/ml-ci-tutorial/; and videos prepared by Susan Athey, Stefan Wager and Jann Spiess, available at https://youtube.com/playlist?list=PLxq_lXOUlvQAoWZEqhRqHNezS30lI49G-.

STRI. In doing so, it highlights specific policy questions which the proposed methods are well placed to provide answers to, and shows that ML-based causal inference is potentially suited to address open challenges in the investigation of the effect of services trade reforms. The proposed approach can be adapted and applied to other empirical settings and used to tackle causal questions defined around different policy reforms.

Second, the exercise demonstrates that an empirical framework constructed around the current STRI database does not yet feature a sufficient degree of variation for a robust application of ML-based causal inference. In doing so, it also suggests specific directions for potential data collection efforts in the future.

The remainder of this annex is organised as follows. Section 1 presents the baseline empirical setting, including a descriptive assessment of the variation in the STRI database when used to construct detailed treatment variables on services trade reforms. Section 2 presents an ML-based method for estimating the average treatment effect of services trade reforms, while Section 3 proposes a few template exercises for the application of ML-based causal inference methods to the study of treatment effect heterogeneity of services trade policy.

Empirical setting

A formal framework to think about causality

The methods discussed in this annex rely on a causal framework based on potential outcomes, also referred to as the Rubin causal model or the counterfactual framework (Rubin, 1974[68]). The baseline model consists of two variables: an outcome and a treatment. Formally, for each empirical unit i, we observe the value of the outcome Y_i and the treatment $W_i \in \{0,1\}$, with $W_i = 1$ denoting a treated unit and $W_i = 0$ representing control.

For each empirical unit the model defines two potential outcomes: $Y_i(1)$ as the value of the outcome for the empirical unit i if i gets treated, and $Y_i(0)$ as i's outcome value if i is a control. For unit i, the causal effect of the treatment W_i on the outcome Y_i is given by $Y_i(1) - Y_i(0)$.

This notation is useful to assess the fundamental problem of causality, which can be stated as follows: a single empirical unit cannot be observed simultaneously as treated and control, or, in other words, counterfactuals do not exist for a single unit (Holland, 1986[69]). For this reason, the analyst cannot observe both $Y_i(1)$ and $Y_i(0)$ but only the realization of one or the other potential outcomes:

$$Y_i = \begin{cases} Y_i(1) & \text{if } W_i = 1 \;\; (i \text{ treated}) \\ Y_i(0) & \text{if } W_i = 0 \;\; (i \text{ control}) \end{cases} \tag{1}$$

As a consequence, it is impossible to observe the causal effect $Y_i(1) - Y_i(0)$ for a single unit i. Any statistical solution to a causality problem will therefore focus on estimating the causal effect for well-defined groups, building the empirical population (when possible) and/or making assumptions to overcome the fundamental problem of causality.

Modelling treatment and outcome variables for estimating the causal effect of services trade reforms on economic outcomes

This causality framework allows to model the problem of estimating the causal effect of services trade reforms. The exercise illustrated here relies on the OECD Services Trade Restrictiveness Index (STRI) regulatory database, which represents a unique source of information on services trade reforms.

The STRI regulatory database consists of approximately 1920 variables for each country, capturing *specific policy measures* identified for their potential restrictiveness to services trade imports. Policy measures can be horizontal, applying to all services sectors covered in the database, or sector specific.[49]

[49] The complete lists of policy measures that apply to each sector of the STRI database are presented in the sector papers available at https://www.oecd.org/trade/topics/services-trade/ (see, for instance, Nordas et al. (2014[82]) for computer services).

Information on policy measures is updated every year since 2014 against the legal framework in force in each country. Therefore, units of observation in the database are the triples formed by a country c, a services sector s and a year t. The version of the database used in this annex covers 50 countries, 22 services sectors and eight years, for a total of 8 800 empirical units.

Most of the variables on individual policy measures have a binary structure, taking value 1 for the empirical unit (c, s, t) if country c at time t applies the respective policy measure in the services sector s in a restrictive way, and 0 otherwise. As an illustration, consider the variable capturing the existence of explicit preferences for local suppliers in public procurement. If the laws on public procurement for services sector s which are in force in country c at time t allow the procuring authority to apply price preferences for domestic suppliers, the variable takes value 1 for the empirical unit (c, s, t).

Overall, these variables characterise the services trade relevant *policy regime* in each empirical unit. A value of 1 (0) represents a restrictive (non-restrictive) policy regime with respect to a certain policy measure. To fix notation, a dichotomous variable for individual policy measure m is denoted as Restrictive policy regime$_{c,s,t}^{m} \in \{0,1\}$.

Given this structure, it is possible to characterise many important dimension of services trade reforms. In particular, the STRI regulatory database allows to define policy measure-specific services trade reforms at the country-sector level. Moreover, it allows to distinguish between liberalising and tightening reforms.

Formally, for each unit (c, s, t), a dichotomous indicator for liberalising reforms on policy measure m is defined as the variable Liberalising reform$_{c,s,t}^{m} \in \{0,1\}$, which takes value 1 if the relevant policy regime in country c for services sector s changes from being restrictive at time $t-1$ (Restrictive policy regime$_{c,s,t-1}^{m} = 1$) to non-restrictive at time t (Restrictive policy regime$_{c,s,t}^{m} = 0$). This definition is captured in the following formula:

$$\text{Liberalizing reform}_{c,s,t}^{m} = \\ = \begin{cases} 1 & \text{if } \Delta\text{Restrictive policy regime}_{c,s,t}^{m} = \text{-}1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Similarly, an indicator for tightening reforms on policy measure m can be defined as

$$\text{Tightening reform}_{c,s,t}^{m} = \\ = \begin{cases} 1 & \text{if } \Delta\text{Restrictive policy regime}_{c,s,t}^{m} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where a value of 1 corresponds to a change in the relevant policy regime from non-restrictive to restrictive.

The two indicators in (2) and (3) can be used to construct a rich portfolio of treatment variables, varying in terms of both the direction of the policy reform $d \in$ {Liberalizing , Tightening} and the specific policy content of the reform given by policy measure m.

To this goal, it is now useful to reconnect to the counterfactual framework introduced above and establish a clear mapping between its elements and the variables defined on the STRI regulatory database. First, the empirical units of this application are given by the triples (c, s, t). Secondly, for any given direction d and content m of a reform, a treatment variable can be defined to take value 1 if the corresponding indicator defined in equations (2) and (3) is equal to 1.

Consider as an example the exercise of studying the causal effect of a liberalising reform concerning preferences for domestic suppliers in public procurement. Treated units will be those country-sector-year triples where the reform indicator Liberalizing reform$_{c,s,t}^{m=\text{procurement preferences for domestic suppliers}}$ takes value 1.

To complete the characterization of the treatment variable, attention should be given to the definition of a valid control group. Given that this example considers a liberalising reform, it is important that control units are those where the policy regime is a restrictive one, i.e. where the procurement authority is allowed to give preferences to local suppliers. In other words, we need to exclude from the control group those units where a non-restrictive regime is already in place. Formally, the treatment variable is equal to 0 when the variable Restrictive policy regime$^{m}_{c,s,t-1}$ is equal to 1.

In general, treated units are those characterised by a policy reform while controls are those units where the regime is equal to that in force prior to the reform. Formally, the treatment variables for a liberalizing or tightening reform can be defined as follows:

$$W^{d=\text{Liberalizing},m}_{c,s,t} == \begin{cases} 1 & \text{if Liberalizing reform}^{m}_{c,s,t}=1 \\ 0 & \text{if Restrictive policy regime}^{m}_{c,s,t}=1 \end{cases} \quad (4)$$

$$W^{d=\text{Tightening},m}_{c,s,t} == \begin{cases} 1 & \text{if Tightening reform}^{m}_{c,s,t}=1 \\ 0 & \text{if Restrictive policy regime}^{m}_{c,s,t}=0 \end{cases} \quad (5)$$

Finally, the outcome variable. This framework allows for high flexibility in the choice of the outcome variable. A first option is given by the set of economic indicators that feature the same dimensions of variability as the treatment variable: countries, services sectors and years. These include country-sector specific services trade flows, employment, and productivity, but also indicators that capture firms or consumers access to a specific service. More aggregate indicators, varying for instance only at the country-year level, can also be used to build outcome variables. In that case, the treatment variable will have to be aggregated accordingly, at the expenses of sample size. For instance, if we consider as empirical units country-year pairs, the maximum sample size allowed by the current version of the STRI database is 400.

A first look at the data: treated and control units across STRI policy measures and directions of reforms

Before turning to the estimation of the causal effect of services trade reforms, this section describes the variation and coverage of the treatment variables that can be defined using the OECD STRI regulatory database.

The data are presented after aggregating the sectoral dimension to match ISIC Rev.4 sections. This maximises the possibility to merge the treatment variables constructed using the OECD STRI database with as many other datasets as possible while keeping the sectoral dimension active. The aggregated data feature 6 services sectors: construction (ISIC Rev 4 Section F), distribution (G), transport (H), finance (K), information and communication (J), and professional services (M), reducing the number of empirical units to 2400 (50 countries times 6 sectors times 8 years). Table AC.1 describes the correspondence between ISIC Rev 4 sections and STRI sectors.

Table A C.1. Sectors correspondence between ISIC Rev 4 sections and STRI sectors

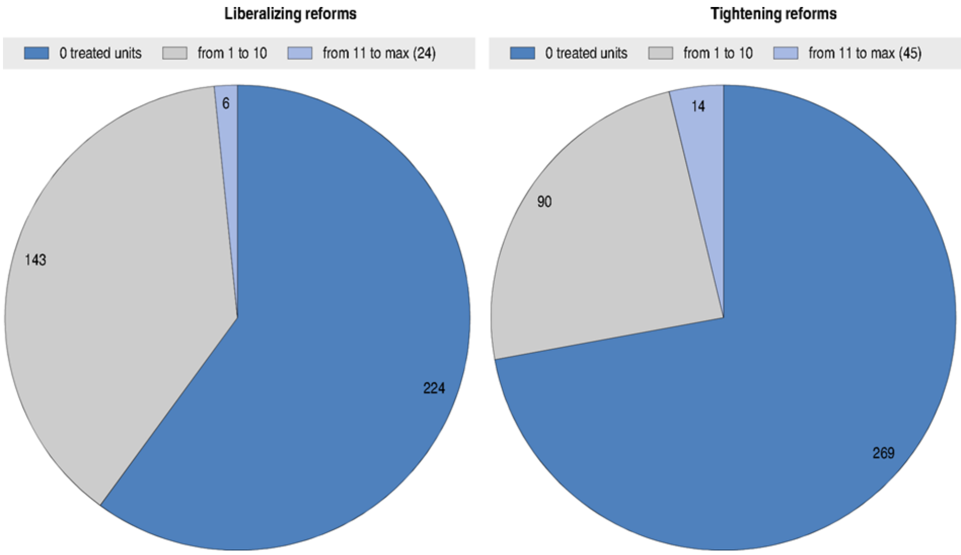
ISIC Rev 4 Sections	STRI sectors
F	Construction
G	Distribution services
H	Air transport; maritime transport; Rail freight transport ; Road freight transport; Logistics cargo-handling; Logistics customs brokerage; Logistics freight forwarding; Logistics storage and warehouse; Courier services
K	Commercial banking; Insurance
J	Broadcasting; Motion pictures; Sound recording; Telecommunication; Computer services
M	Legal services; Accounting services; Architecture services; Engineering services

Source: Authors' compilation based on ISIC rev. 4 classification.

Sectoral aggregation has implications for the definition of treated and control units in those cases where the aggregate sectoral dimension based on ISIC Rev 4 sections includes more than one STRI services sector. For those instances, treated units are defined as those where the relevant reform applies to at least one of the STRI services sectors included in the respective aggregate sectoral category. As for control units, they are defined as those where the policy regime is equal to the pre-reform status for all STRI services sectors included in the aggregate category.

The first important observation on the data is that, for all possible treatment variables defined through binary policy measures in the OECD STRI regulatory database, the number of treated units is very small if not 0. Figure AC.1 shows that most policy measures in the STRI database allow to generate treatment variables with 0 treated units. Indeed, for many STRI measures no policy change has been recorded during the sample period from 2014 to 2021. For only 6 (14) measures it is possible define treatment variables that identify more than 10 units treated by a liberalising (tightening) reform.

Figure A C.1. STRI policy measures by number of treated units identified by the respective treatment variable



Note: The pie charts in the figure categorise the STRI policy measures m by the number of treated units identified through the respective treatment variable $W_{c,s,t}^{d,m}$ as defined by equations (4) and (5). The figure distinguishes between the two possible directions of reform, Liberalising (left chart) and Tightening (right chart). The total number of binary policy measures considered is 373.
Source: OECD STRI Regulatory Database.

The second observation is on control units. For almost all treatment variables that can be defined based on STRI policy measures, the number of control units is smaller than the difference between the total number of empirical units covered in the STRI regulatory database (2400 after the sectoral aggregation) and the number of treated units. The conditions discussed above for a sensible definition of control units impose constraints that significantly reduce the size of the final estimation sample. On average across STRI policy measures, the number of control units identified by the respective treatment variables for liberalizing (tightening) reforms is 182 (459).[50]

Overall, these features of the data imply both a limited sample size as well as a strong unbalance between the numbers of treated and control units for almost all possible treatment variables. The applications discussed below will show how the lack of statistical power and the very small shares of treated units in estimation samples are critical issues that limit a feasible and robust estimation of any causal effect of services trade reforms using these empirical settings.

[50] The smaller average number of control units for treatment variables capturing a liberalising reform shows that in the it is more likely to find empirical units characterized by open policy regimes rather than restrictive ones.

By sharpening the assessment of the relevant data limitations, these observations allow to identify the STRI policy measures associated with the treatment variables featuring the highest number of treated units. Table AC.2 and table AC.3 list the OECD STRI measure code and title, the number of treated units and controls and the share of treated units in the empirical population for the ten treatment variables with the highest number of treated units. Table AC.2 focuses on treatment variables reflecting liberalising reforms while Table AC.3 characterises treatment variables for tightening reforms.

Table A C.2. Top 10 liberalising treatment variables for number of treated units

Measure code	Measure title	Treated units	Controls	% of treated units
1_9_1	Conditions on subsequent transfer of capital and investments	24	206	10.43
5_2_1	There is an adequate public comment procedure open to interested persons, including foreign suppliers	18	570	3.06
4_6_1	Minimum capital requirements	15	1108	1.34
4_3_1	National, state or provincial government control at least one major firm in the sector	13	281	4.42
3_2_2	Public procurement: Procurement regulation explicitly prohibits discrimination of foreign suppliers	12	2015	0.59
1_50_1	Other restrictions on foreign entry	11	292	3.63
1_5_1	Screening explicitly considers economic interests	8	252	3.08
1_4_4	Board of directors: at least one must be resident	8	720	1.10
2_1_3	Quotas: independent services suppliers	7	517	1.34
2_1_1	Quotas: intra-corporate transferees	7	235	2.89

Note: The table lists the OECD STRI measure code and title, the number of treated units and controls and the share of treated units in the empirical population for the 10 treatment variables capturing a liberalising reform with the highest number of treated units.
Source: OECD STRI Regulatory Database.

Table A C.3. Top 10 tightening treatment variables for number of treated units

Measure code	Measure title	Treated units	Controls	% of treated units
1_5_2	Screening exists without exclusion of economic interests	45	1397	3.12
2_2_1	Labour market tests: intra-corporate transferees	30	708	4.07
2_50_1	Other restrictions to movement of people	30	1883	1.57
1_20_5	Cross-border data flows: certain data must be stored locally	24	1775	1.33
1_3_141	Licences are subject to quotas or economic needs test	20	220	8.33
1_20_3	Cross-border data flows: cross-border transfer of personal data is possible to countries with substantially similar privacy protection laws	18	498	3.49
2_1_2	Quotas: contractual services suppliers	18	1952	0.91
2_1_1	Quotas: intra-corporate transferees	18	2160	0.83
2_1_3	Quotas: independent services suppliers	18	1878	0.95
2_2_2	Labour market tests: contractual services suppliers	12	684	1.72

Note: The table lists the OECD STRI measure code and title, the number of treated units and controls and the share of treated units in the empirical population for the 10 treatment variables capturing a tightening reform with the highest number of treated units.
Source: OECD STRI Regulatory Database.

The analysis illustrated in the reminder of this annex will use measures among those listed in Tables AC.2 and AC.3.

1. Average treatment effect: What is the average effect of a services trade reform?

Within the counterfactual framework introduced above, a first step in the study of the causal effect of services trade reforms on economic outcomes consists in the estimation of the average treatment effect (ATE). A specific pair of outcome and treatment variables are identified as an illustration.

The outcome variable $Y_{(c,s,t)}$ is given by services import penetration, defined as the ratio of imports over gross output in country c, in services sector s at time t. Data on services imports are built by combining EBOPS 2010 data from Eurostat, the OECD, the IMF, the WTO and UN Comtrade, while services gross output is sourced from the OECD TiVA database. Given the EBOPS framework for the services trade variable, we interpret this as Mode 1, Mode 2 and Mode 4 services import penetration.

The treatment variable instead is fixed to capture liberalising reforms in terms of the national, state or provincial government in country c ceasing to control any major firm in services sector s at time t. Using the definition introduced by equations (4) and (5), this corresponds to the treatment variable $W_{c,s,t}^{d=\text{Liberalizing},m=\text{National, state or provincial government control at least one major firm in the sector}}$. The reminder of the section will refer to this treatment variable simply by $W_{(c,s,t)}$ unless differently specified.

The average treatment effect of this liberalising reform on services import penetration is given by:

$$\text{ATE} = E[Y_{(c,s,t)}(1) - Y_{(c,s,t)}(0)] \quad (6)$$

The exercises presented here focus on ATE estimation under the assumption of unconfoundedness. This assumption requires that, conditioning on some observable covariates $\boldsymbol{X}_{(c,s,t)}$, the treatment is assigned to empirical units independently on potential outcomes. Formally, unconfoundedness can be stated as follows:

$$Y_{(c,s,t)}(1), Y_{(c,s,t)}(0) \perp W_{(c,s,t)} | \boldsymbol{X}_{(c,s,t)} \quad (7)$$

In the specific economic application analysed here, unconfoundedness requires that conditioning on observable features of the economic environment characterising country-sector-year empirical units, the fact that a liberalising trade reform is observed in a country-sector pair in a certain year is not telling anything about the import flows of that country, in that particular sector by the end of that year.

It might be the case for instance that governments in larger countries are more prone to implement a liberalising reform as the one captured by the treatment variable $W_{(c,s,t)}$, because of higher demand and economies of scale in the country. These features could also be driving forces triggering an increase in services import penetration and this would imply a positive correlation between $W_{(c,s,t)}$ and $Y_{(c,s,t)}(1) - Y_{(c,s,t)}(0)$. However, by comparing countries with similar levels of GDP and population, the occurrence of the reform would be uninformative of the level of services imports.

Unfortunately, the economic and demographic size of a country are not the only confounding factors in this relationship. Indeed, there are many observable features that are likely to make the treatment assignment correlated with the potential outcomes (see Egger and Shingal (2020[70]) for a recent discussion of the determinants of services trade policy). Moreover, observables might shape the relationship between the treatment and potential outcomes in non-linear ways, including through complex interactions between themselves which it would be impossible for the researcher to specify correctly *ex ante*.

This is one of the main entry points for deploying off-the-shelf ML algorithms to support inference about the ATE. Indeed, given a satisfactory set of observable features, ML tools allow for a non-parametric and data-driven identification of the relevant confounding variation. This entails significant advantages with respect to other approaches based on panel econometrics, especially in this empirical setting where demanding batteries of fixed effects would absorb important identifying variation (between countries, between sectors and between country-sector pairs) in the STRI-based treatment variables.

For this application, a comprehensive set of observable covariates is proposed. To capture any confounding variation originating from functions of sector, country, and year-specific shocks, the list of covariates includes: a categorical variable with 6 values corresponding to the ISIC Rev 4 sections covered in our data; a time variable; latitude and longitude of each country's centroid.[51] Deflated services export intensity (built in the same way as services import intensity), deflated Mode 3 services imports and exports flows (in millions of US dollars, sourced from the OECD Analytical AMNE database), are added to the set of covariates, all varying across the three dimensions of the empirical units.

Finally, the list of covariates includes variables varying at the country-year level that are likely to affect both national governments' incentives to implement services trade reforms as well as services import performance. These are: population, population density, applied tariff rate (weighted mean in percentage points), GDP and GDP per capita (PPP, constant 2017 international dollars), Gini index, gross capital formation (% of GDP) from World Bank databases; consumer price index sourced from the IMF; deflated aggregate commercial services imports and exports (in millions US dollars) from the WTO/OECD Balanced International Trade in Services Database; deflated aggregate goods imports and exports (in millions US dollars) from UN Comtrade; and a remoteness indicator by to capture forces of multilateral resistance at the country-year level.[52]

This set of covariates represents a good baseline for claiming unconfoundedness in this application, especially given the flexible and data-driven approach in controlling for their confounding potential allowed by ML algorithms. However, by merging all data sources which do not necessarily match the country, sector and time coverage of the STRI regulatory database, additional constraints are imposed on the estimation sample and ultimately imply a reduction of the total number N of empirical units that can used for estimation. The major issue here is data availability for the most recent years. Indeed, for some crucial variables (e.g. all variables sourced from the OECD TiVA database and the related Analytical AMNE) as well as for many other variables which represents good candidate for covariates (e.g. services employment, wages), very little information is systematically available covering the period after 2018.

Estimating the ATE using ML

This section presents the estimation pipeline for the ATE that uses ML tools to reduce reliance on parametric specifications while keeping all statistical guarantees required for causal inference.

A successful method for applying ML to the problem of estimating the ATE relies on the Augmented Inverse-Propensity Weighted (AIPW) estimator (Robins, Rotnitzky and Zhao, 1994[63]). The AIPW combines two well-known estimation strategies. The first one is based on the representation of the ATE under unconfoundedness in terms of the conditional expectation function $\mu(\boldsymbol{x}, w) = E\left[Y_{(c,s,t)} \middle| \boldsymbol{X}_{(c,s,t)} = \boldsymbol{x}, W_{(c,s,t)} = w\right]$. The ATE can be written as:

$$ATE = \mu\left(\boldsymbol{X}_{(c,s,t)}, W_{(c,s,t)} = 1\right) - \mu\left(\boldsymbol{X}_{(c,s,t)}, W_{(c,s,t)} = 0\right) \tag{8}$$

This representation delivers the following estimator for the ATE

$$\hat{\tau} = \frac{1}{N} \sum_{(c,s,t)=1}^{N} \hat{\mu}\left(\boldsymbol{X}_{(c,s,t)}, 1\right) - \hat{\mu}\left(\boldsymbol{X}_{(c,s,t)}, 0\right) \tag{9}$$

[51]As in Johannemann et al. (2021[83]) relevant time-invariant country-level features are flexibly identified through the combination of continuous variables built on the latitude and longitude of the country's centroid. This is a useful alternative to country-level fixed effects since many learning algorithms, including the regression-tree-based ones used in this annex, do not work well with categorical variables with many values.

[52] Remoteness for country c at time t is defined as $\sum_j = \text{bilateral ditance}_{cj} * {GDP_j}/{GDP_{world}}$ (Hannan, 2017[84]). Data on bilateral distance are from CEPII and GDP data from the World Bank.

It is important to acknowledge that deploying an ML algorithm (instead of running a linear regression) to estimate $\mu(\boldsymbol{X}_{(c,s,t)}, W_{(c,s,t)})$ can easily allow $\hat{\tau}$ to be a consistent estimator for the ATE, but ML based methods will always be too prone to bias to allow $\hat{\tau}$ to converge to its estimand at the rate required to generate confidence intervals.

The second strategy used in the construction of the AIPW estimator relies instead on the characterization of the ATE in terms of the propensity score, i.e. the probability of treatment given the value of the observable covariates: $e(\boldsymbol{x}) = \Pr[W_{(c,s,t)} = 1 | \boldsymbol{X}_{(c,s,t)} = \boldsymbol{x}]$. The estimator for the ATE that follows is the Inverse Propensity Weighted (IPW) estimator:

$$\hat{\tau}^{IPW} = \frac{1}{N} \sum_{(c,s,t)=1}^{N} \left(\frac{W_{(c,s,t)} Y_{(c,s,t)}}{\hat{e}(\boldsymbol{X}_{(c,s,t)})} - \frac{(1 - W_{(c,s,t)}) Y_{(c,s,t)}}{1 - \hat{e}(\boldsymbol{X}_{(c,s,t)})} \right) \tag{10}$$

Given the presence of the propensity score in the denominators of (10), the strategy based on $\hat{\tau}^{IPW}$ requires that the propensity score is bounded away from 0 and 1. Again, using ML methods to estimate the propensity score and compute $\hat{\tau}^{IPW}$ will generate non-negligible biases that usually prevents a successful application of off-the-shelf ML algorithms to this estimation strategy.

Qualitatively, the AIPW combines the conditional expectation and the propensity score strategies in the following way: it starts by estimating the ATE using $\hat{\tau}$ and then tries to correct any bias of the estimated conditional mean functions by applying the IPW to the residuals $Y_{(c,s,t)} - \hat{\mu}(\boldsymbol{X}_{(c,s,t)}, W_{(c,s,t)})$. This idea can be seen directly in the formula for the AIPW estimator below, where the first line in the sum is equal to $\hat{\tau}$ and the rest provides the IPW correction of the residuals:

$$\begin{aligned} \hat{\tau}^{AIPW} = \frac{1}{N} \sum_{(c,s,t)=1}^{N} & \hat{\mu}(\boldsymbol{X}_{(c,s,t)}, 1) - \hat{\mu}(\boldsymbol{X}_{(c,s,t)}, 0) \\ & + \frac{W_{(c,s,t)}}{\hat{e}(\boldsymbol{X}_{(c,s,t)})} \left(Y_{(c,s,t)} - \hat{\mu}(\boldsymbol{X}_{(c,s,t)}, 1) \right) \\ & - \frac{1 - W_{(c,s,t)}}{1 - \hat{e}(\boldsymbol{X}_{(c,s,t)})} \left(Y_{(c,s,t)} - \hat{\mu}(\boldsymbol{X}_{(c,s,t)}, 0) \right) \end{aligned} \tag{11}$$

Under unconfoundedness and overlap the estimator $\hat{\tau}^{AIPW}$ is available with many good statistical properties. In particular, ML methods can be flexibly used to solve the intermediate steps of estimating $\hat{\mu}(\boldsymbol{X}_{(c,s,t)}, 1)$, $\hat{\mu}(\boldsymbol{X}_{(c,s,t)}, 1)$ and $\hat{e}(\boldsymbol{X}_{(c,s,t)})$, which all share the structure of prediction problems that ML tools are good at solving. Indeed, making sure that the chosen ML tool is cross-fitting to avoid overfitting biases, the AIPW estimated using ML has excellent asymptotic properties that allow the construction of confidence intervals and that make it a preferred choice with respect to any other non-parametric estimator (Chernozhukov et al., 2018[64]).

The application presented in this annex proposes a random forest model to estimate the ingredients of the AIPW. In particular, it uses the Causal Tree and Causal Forest algorithms introduced by Athey and Imbens (2016[65]) and Wager and Athey (2018[66]) and implemented in the R package grf (grf/REFERENCE.md at master · grf-labs/grf · GitHub).[53] These algorithms are based on sample splits, such that, each time a tree is built, the data used to build the tree are different from those used to compute predictions on the tree's leaves. This sample splits-based approach, also called honesty, guarantees cross-fitting required for inference on the AIPW.

[53] These algorithms are also available in the Python package EconML, an open source software developed by Microsoft Research (https://www.microsoft.com/en-us/research/project/econml/).

Results and diagnostics

The AIPW-based estimation pipeline with Causal Forest is applied to the exercises defined in terms of the 20 treatment variables listed in Tables A C.2 and A C.3. The outcome variable is fixed and equal to services import penetration. The vector of covariates is also the same across all 20 treatment variables (see the list of observable covariates discussed above). Table A C.4 offers a qualitative summary of these exercises. For each treatment variable identified through the direction and policy content of the services trade reform, the table reports the sign of the point estimate for the ATE and information on whether the respective 95% confidence interval includes 0. The table also reports, for each estimation exercise, the number of treated units and controls in the final estimation sample.

Table A C.4 reflects the strong limitations of the empirical setting. For six exercises, the number of treated units is so small (0 in 2 cases) that no estimation is possible. For 12 exercises the 95% confidence interval estimated around the ATE includes 0, leaving the analyst with strong uncertainty on whether the true ATE is positive, negative or 0. Only for two exercises the 95% CI does not include 0.[54] This is the case for the treatment variable capturing a liberalising reform in terms of government control of major services providers in a specific sector, and for the one reflecting a tightening policy change that makes licenses subject to quotas or economic needs tests. Given the weaknesses of the empirical setting and in particular the high rate of non-computability of ATE estimates, it is of foremost importance to test whether the data satisfy the requirements necessary for any interpretation of the estimated coefficients.

As an illustration, a set of diagnostics is conducted on the exercise that studies the causal effect on services import intensity of removing government control from major services providers.[55] The AIPW point estimate for that exercise is 0.11, with standard error equal to 0.04. If these estimates were interpretable, they would point to a very large effect of the services trade reform, amounting to an increase in Mode 1, Mode 2 and Mode 4 services import intensity between 0.05 and 0.17 units.[56]

Table A C.4. ATE results across 20 exercises

Direction of reform	Policy content of reform (title of STRI policy measure)	AIPW and causal forest-based estimate of the ATE	Treated units	Controls
Liberalising	Conditions on subsequent transfer of capital and investments	Positive, CI includes 0	10	74
	There is an adequate public comment procedure open to interested persons, including foreign suppliers	Negative, CI includes 0	10	220
	Minimum capital requirements	Positive, CI includes 0	5	504
	National, state or provincial government control at least one major firm in the sector	Positive, CI does not include 0	11	78
	Public procurement: Procurement regulation explicitly prohibits discrimination of foreign suppliers	Not estimated	0	823
	Other restrictions on foreign entry	Not estimated	0	81
	Screening explicitly considers economic interests	Not estimated	3	62
	Board of directors: At least one must be resident	Not estimated	1	210
	Quotas: Independent services suppliers	Not estimated	1	228
	Quotas: Intra-corporate transferees	Negative, CI includes 0	6	108

[54] The grf package allows for contemporaneous estimation of the effect of a reform for multiple outcomes. This feature can be used to estimate the reform's effect on the same outcome variable at different points in time. Therefore, it represents a great tool to assess the dynamics of the causal impact of the reform, from the short-term effect on the outcome observed at the end of the reform year, to a medium-term effect three years after the reform. Unfortunately, none of the 20 exercises listed in Table AC.4 survives the additional data requirement of observing, for each treatment unit, a time series of the outcome variables covering three or even two years.

[55] The title of the corresponding STRI measure is: National, state or provincial government control at least one major firm in the sector (measure code: 3_4_1).

[56] These are large increases relatively to the mean value of services import intensity which is equal to 0.17 for the whole empirical population and 0.06 for the specific estimation sample.

Direction of reform	Policy content of reform (title of STRI policy measure)	AIPW and causal forest-based estimate of the ATE	Treated units	Controls
Tightening	Screening exists without exclusion of economic interests	Positive, CI includes 0	6	589
	Labour market tests: Intra-corporate transferees	Positive, CI includes 0	25	307
	Other restrictions to movement of people	Negative, CI includes 0	22	728
	Cross-border data flows: Certain data must be stored locally	Negative, CI includes 0	16	720
	Licences are subject to quotas or economic needs test	Positive, CI does not include 0	102	20
	Cross-border data flows: Cross-border transfer of personal data is possible to countries with substantially similar privacy protection laws	Not estimated	5	159
	Quotas: Contractual services suppliers	Positive, CI includes 0	10	732
	Quotas: Intra-corporate transferees	Positive, CI includes 0	10	809
	Quotas: Independent services suppliers	Positive, CI includes 0	10	689
	Labour market tests: Contractual services suppliers	Positive, CI includes 0	10	270

Note: This table describes the AIPW and causal forest-based estimate of the ATE for 20 different causal inference exercises on the short term effect of services trade reforms. Each exercise is characterised by a different treatment variable, defined in terms of the direction (liberalising or tightening) and policy content of the services trade reform. The content of the reform is given by the specific STRI policy measure used to build the treatment variable. Direction and content of the reform are specified in the first two columns of the table. For each exercise, the third column of the table reports the sign of the point estimate for the ATE and information on whether the respective 95% confidence interval (CI) includes 0 or not. Finally, the last two columns of the table report, for each estimation exercise, the number of treated units and controls in the final estimation sample.
Source: Authors' calculations.

The first diagnostic test to be conducted verifies the overlap requirement: the propensity scores need to be bound away from 0 and 1. To assess overlap, Figure A C.2 plots the histogram of causal forest-based estimates of the propensity scores stratified by treatment level.

The figure shows that, especially for the population of controls, propensity scores tend to be very close to 0. This is an issue for the proposed estimation pipeline which raises serious concerns on the interpretability of the results. When, as in this case, overlap fails or appears very weak, one solution can be to shift the focus to a different estimand, the Average Treatment effect for the Treated (ATT). The point estimate (0.14) and standard error (0.03) for the ATT are pointing to an even stronger and more precise positive effect. However, given the very limited size of the estimation sample, this could still only reflect noise in the data.

Another problem that might undermine the capacity of these estimates to provide insights on the causal linkages between the services trade reform and services import intensity is the lack of balance. Balance captures the degree of similarity between the treated and control groups. Differences in the distribution of observable covariates across the two groups is a source of bias in ATE estimates. The inverse propensity weights used in the estimation pipeline are designed to correct the biases generated by these discrepancies and it is important to check to what extent these weights are successful in doing so. To this purpose, Figure AC.3 plots the distribution of selected covariates, stratified by treatment status, before and after weighting observations by the estimated propensity scores.

The distributions of covariates across the upper and lower panel of Figure A C.3 shows that, while the inverse propensity weights are making the treated and untreated populations more similar in terms of observable covariates, this is not enough to correct the differences between the two groups, also due to the strong unbalance in the size of the two populations.

The failure of propensity weights to correct discrepancies between the treated and control groups can be assessed also by focusing on the absolute standardised mean difference (ASMD) of covariates, before and after the propensity score-based adjustment. For any element $X_{(c,s,t)}$ in the vector of covariates $\boldsymbol{X}_{(c,s,t)}$ the ASMD is defined as

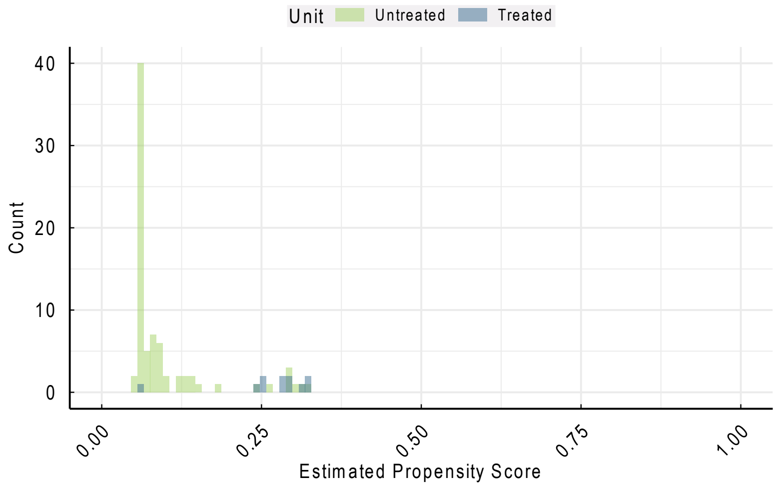
$$\frac{|\bar{X}_1 - \bar{X}_0|}{\sqrt{s_1^2 + s_0^2}} \quad (12)$$

with $\bar{X}_1$ ($\bar{X}_0$) and s_1^2 (s_0^2) being $X_{(c,s,t)}$. sample mean and standard deviation in the population of treated (control) units. Figure A C.4 compares the ASMD computed on actual observations of covariates with the ASMD computed on observations weighted by the inverse estimated propensity scores.

The unadjusted ASMDs, laying very far from the 0 vertical line, confirm that in this empirical setting the populations of treated and control units are very different from each other. Moreover, the adjusted ASDMs, still far away from the 0 line (and, for some covariates, even further away from the 0 line than their unadjusted counterpart), highlight how the estimated propensity scores are not well calibrated to fix this discrepancy.

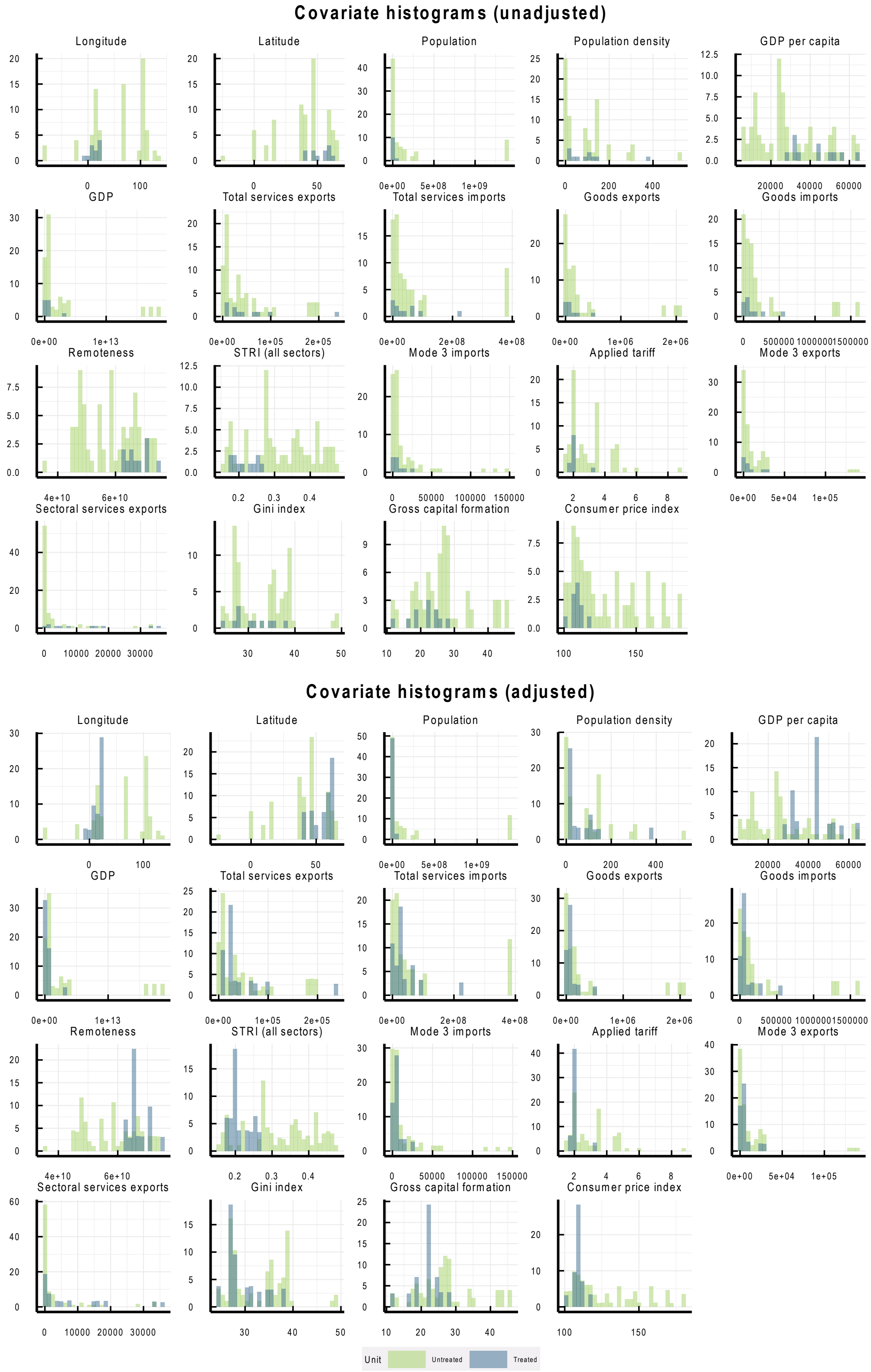
The results of these diagnostic tests cast serious doubts over the capacity of the proposed estimation pipeline to extract from our empirical setting any meaningful signal about the causal effect of the services trade reform on services import penetration. Similar diagnostics can be derived for the other exercises listed in Table A C.4 (in those cases where estimates are actually computed), which strongly suggests that ATE estimates in this empirical setting are likely to mostly reflect the noise in the data rather than an interpretable causal parameter.

Figure A C.2. Assessing overlap



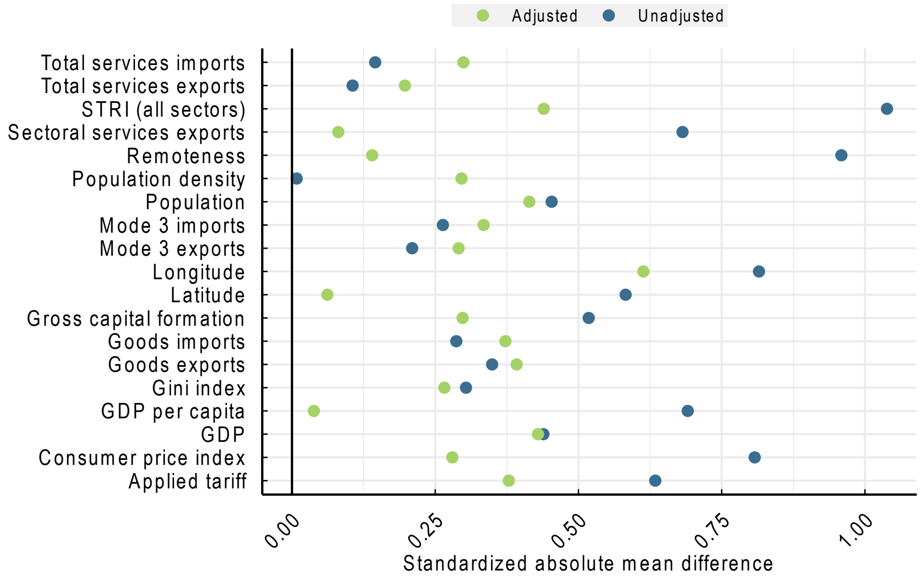
Note: This figure plots the histogram of causal forest-based estimates of the propensity scores ($\hat{e}$) for treated and untreated (controls) units.
Source: Authors' calculations.

Figure A C.3. Assessing balance



Note: This figure plots the histograms of selected covariates, stratified by treatment status of each observed empirical unit. The upper panel shows the histograms on unweighted observations while the lower panel replicates the exercise on observations weighted by the estimated propensity scores.

Figure A C.4. Assessing balance through ASMD



Note: The figure plots for selected covariates the ASMD computed on actual observations (Unadjusted) and the ASMD computed on observations weighted by the inverse estimated propensity scores (Adjusted).

2. Heterogeneous treatment effect

Being largely data-driven and requiring fewer assumptions on functional forms, ML-based methods for causal inference could be also useful to explore heterogeneity in the effect of services trade policy, potentially shedding new light on the features of the economic environment that shape the effects of reforms and allowing for individually tailored policy assessments. This section proposes selected exercises that apply ML-based methods to investigate treatment effect heterogeneity of services trade reforms. Given the weaknesses of ML-based ATE estimation on STRI data, the applications that follow can be interpreted only as templates and no robust economic interpretation nor policy implication can be derived from them. Indeed, ML-based methods for the study treatment effect heterogeneity require more data and variation than the methods for ATE estimation presented above.

ML for studying treatment effect heterogeneity: Selected template exercises

Services trade reforms interact in complex ways with relevant features of the economic environment and their impact on economic outcomes can be affected by those interactions. One challenge to the empirical test of this hypothesis lies in the limitations of linear regression frameworks when used to study heterogeneous causal effects. In those models all factors suspected to play a role in shaping the effect of services trade reforms would have to be identified, interacted with the treatment variable capturing services trade policy, and all those interactions included simultaneously in the regression. Moreover, for the factors that might shape the impact of services trade policy in a non-linear way, the exact functional form would have to be specified and accounted for in the model. It is often the case that both the determinants of heterogeneity and the way they shape the effect of services trade policy are *a priori* only partially determined.

ML could potentially allow to flexibly and simultaneously test the role of many factors in shaping the causal effect of services trade reforms on economic outcomes. In particular, the Causal Tree and Causal Forest algorithms introduced above are powerful tools to study heterogeneous causal effects. The basic principle of these models is to use validation procedures to identify the factors and their values (i.e. the underlying economic environments) that best predict similar effects of services trade reforms on the outcome of interest. Moreover, by using a dedicated subsample to estimate the treatment effect in each well-specified economic environment, they allow to estimate confidence intervals around point estimates of average treatment effects in that environment.

The main estimand to study treatment effect heterogeneity is the Conditional Average Treatment Effect (CATE), which is the ATE computed on empirical units characterised by a specific value $\boldsymbol{x}$ of the observable covariates.

$$\text{CATE} = E[Y_{(c,s,t)}(1) - Y_{(c,s,t)}(0) | \boldsymbol{X}_{(c,s,t)} = \boldsymbol{x}] \tag{13}$$

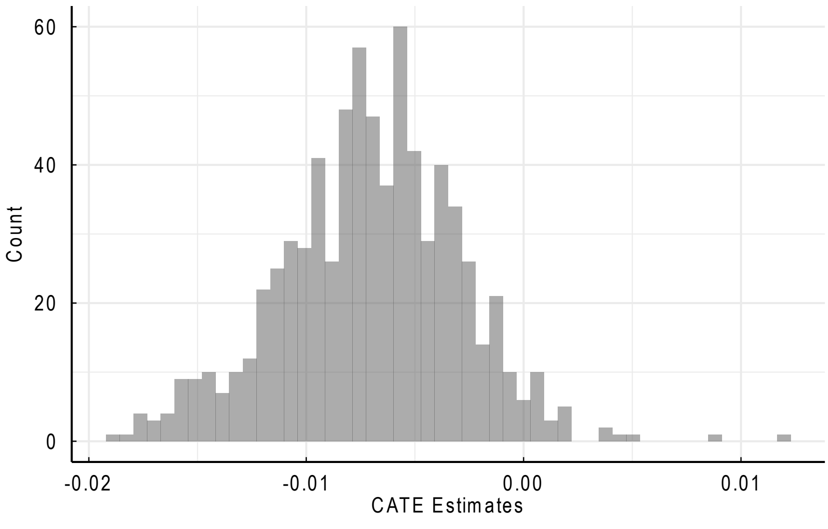
For each $\boldsymbol{x}$, the corresponding CATE prediction can be estimated using the Causal Forest algorithm and a first step in the assessment of treatment effect heterogeneity consists in the study of the distribution of those estimates.

As in the case of ATE analysis, the focus of the application will be on a specific exercise, defined by a unique outcome variable, a treatment, and a set of covariates. The outcome variable is again equal to services import intensity and the vector $\boldsymbol{X}_{(c,s,t)}$ also remains unchanged with respect to the one defined in the previous section. The treatment variable instead is chosen to capture the introduction of barriers to cross-border data flows in the form of local storage requirements.[57] The exercise is then to understand whether and how the effect of a tightening reform that introduces local storage requirements to cross-border data flows on services import intensity differs across economic environments.

The estimation pipeline for the ATE based on the AIPW and Causal Forest delivers an estimated average treatment effect equal to -0.01, with standard error equal to 0.01 and a 95% CI going from -0.04 to 0.01. If these numbers were interpretable from an economic point of view, they would signify that a reform introducing barriers to cross-border data flows would have a negative but not statistically significant effect on services import intensity. Similar diagnostics as the one discussed in the section above, suggest that the unbalance and lack of overlap in the empirical setting do not allow for any economic interpretation to be derived from these estimates.

Abstracting from this caveat, and for the purpose of offering templates for the investigation of treatment effect heterogeneity, Figure A C.5 plots the histogram of CATE estimates computed using the R package grf introduced above.

Figure A C.5. Distribution of estimated CATE predictions



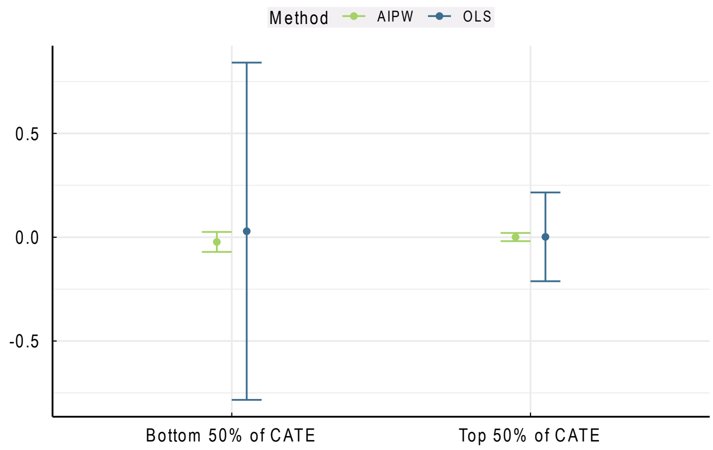
Note: This figure plots the histogram of CATE predictions estimated with Causal Forests.
Source: Authors' calculations.

57 The title of the corresponding STRI measure is: Cross-border data flows: certain data must be stored locally (measure code: 1_20_5). Choosing a different treatment variable than the one used to illustrate ATE estimation serves two purposes: (i) it allows to re-emphasise the flexibility of the method to work with treatment variables capturing services trade reforms with different policy content and direction; and (ii) it permits to base CATE estimation on a larger sample (736 observations rather than the 89 used in the ATE exercise).

Whether the variation in the estimated CATE can be interpreted as a robust insight on the true heterogeneity in the effect of introducing barriers to cross border data flows requires further investigation. A first rough test to assess the quality of the estimated CATE consists in ranking observations in terms of their estimated CATE prediction, dividing the sample in groups, from low to high CATE, and then estimating the ATE for each of these groups[58].

Figure A C.6 plots the results for this test, where observations are divided in two groups, below and above the median CATE prediction. The figure shows that there is no detectable treatment effect heterogeneity across these two sub-populations[59].

Figure A C.6. Assessing heterogeneity embedded in estimated CATE predictions



Note: This figure plots the OLS and AIPW estimates for the ATE of introducing barriers to cross border data flows on services import intensity by quantile of the estimated CATE distribution.
Source: Authors' calculations.

After assessing the existence of treatment effect heterogeneity, CATE estimates can be used to unpack such heterogeneity, or in other words, to learn about the features that characterise economic environments associated with a weaker or stronger effect of services trade reforms. A first exercise in this space consists in comparing average values of covariates of interest across rankings as defined by the CATE. Figure A C.7 shows a simple application based on the two quantiles of estimated CATE predictions.

[58] To make sure that observations are ranked based on a model that is not fitted using those same observations, the data can be divided into folds, and in turns rank the observations in one of them according to a model for the CATE built using the other folds. The very small number of treated units (both in absolute and relative terms) prevents from adopting this approach in the construction of the ranking. This would have implications on the reliability of any inference exercise based on the ranking, which in this case can be neglected given the illustrative nature of the exercise.

[59] More sophisticated calibration tests are available to assess the quality of CATE estimates in providing reliable signals of the true treatment effect heterogeneity (see the function test_calibration in the grf package also discussed in https://bookdown.org/halflearned/ml-ci-tutorial/hte-i-binary-treatment.html#eq:cate).

Figure A C.7. Unpacking heterogeneity

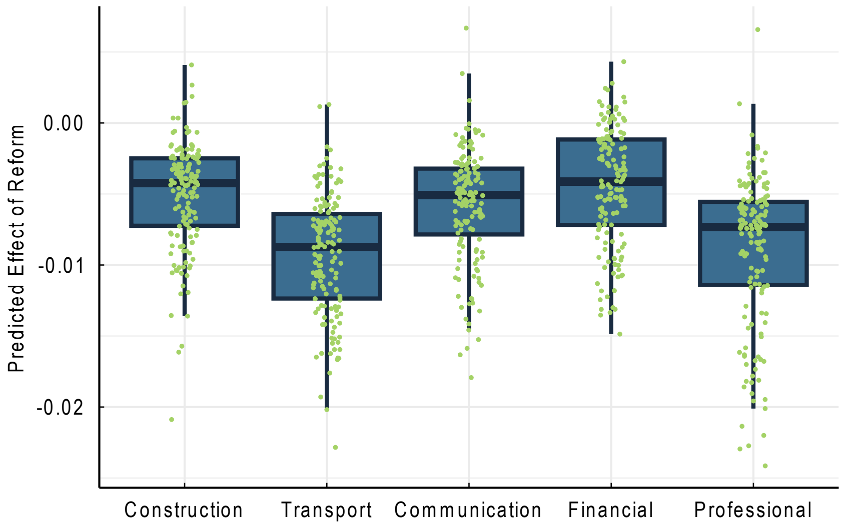
Note: The figure plots the mean and standard deviation for selected covariates by quantile of the estimated CATE distribution. The colors in the matrix highlight the quantile where a covariate has the highest (green) or lowest (blue) mean value.
Source: Authors' calculations.

According to the first four rows in the plotted matrix, country-sector-year observations with lower CATE predictions are those associated with higher barriers to services trade as captured by the overall STRI as well as by higher services trade performance. If these comparisons were meaningful, they would suggest that the analysed tightening reform could have a stronger negative effect on imports in environments characterised by higher services trade performance (across modes and directions) and higher barriers to services trade. With larger and more balanced datasets at hand, the same exercise can be replicated by dividing observations in more than just two groups. For instance, studying the distribution of covariates across rankings based on CATE quintiles would allow the analyst to detect potential non-linearities in the relationship between a specific feature of the economic environment and the magnitude of the estimated treatment effect.

Also with the aim of unpacking heterogeneity, the predicted effect of the services trade reforms can be compared in specific economic environments as defined for instance by countries or sectors. The main difference with the approach taken in the previous exercise is that here the analyst does not frame her investigation of heterogeneity on data-drive partitions of the data but rather on subgroups representing pre-specified economic environments. This approach can be used to derive individually tailored predictions, such has the effect of the reform for a specific country in a particular sector and year.

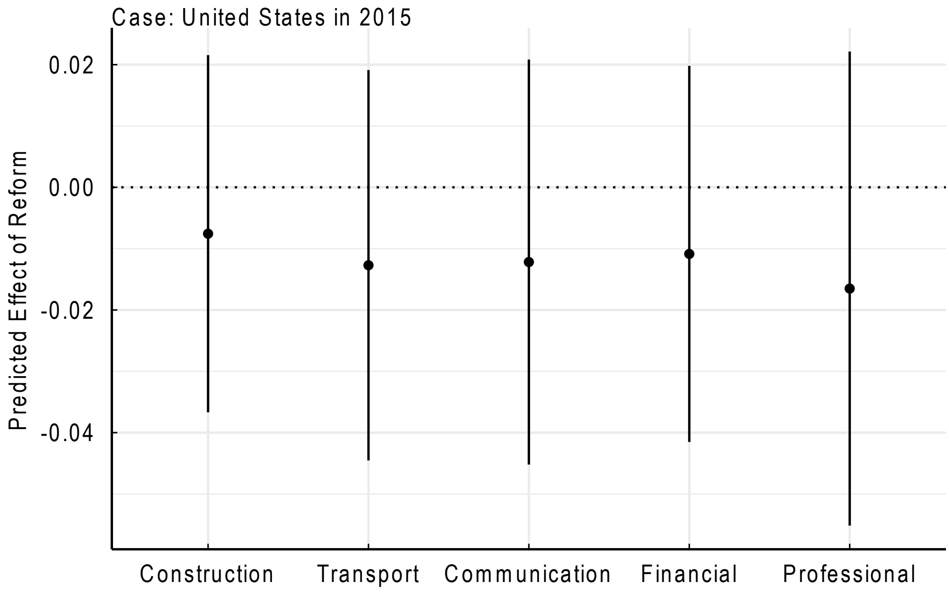
As examples of this, Figure AC.8 plots the distribution of predicted effects (across countries and years) grouped by sector, while Figure AC.9 shows the predicted effect of a hypothesised policy reform in the United States in 2015 for each sector covered in the data.

Figure A C.8. Predicted effects across sectors



Note: This figure displays for each sector the predicted effects of reform on services import intensity as captured by the estimated CATE predictions (green dots). It also shows the boxplot for the sector-specific distribution of the CATE predictions.
Source: Authors' calculations.

Figure A C.9. Predicted effects of a hypothesised reform



Note: The figure plots the estimated CATE prediction and the respective 95% confidence interval for the United States, in 2015 and for each sector covered in the estimation sample.
Source: Authors' calculations.

A third approach to unpack treatment effect heterogeneity consists in regressing the estimated CATE predictions on covariates of interest. The estimated coefficients of this linear model would be suggestive of the general trends in the relationship between relevant covariates and the treatment effect.

Caution should be exerted when using variable importance metrics to explore the role of covariates in shaping treatment effect heterogeneity. When applied to ensemble methods such as Causal Forests, variable importance metrics are informative about the relevance of covariates in contributing to the prediction of the model but not necessarily about their role in the data-generating process. In case of high correlation between two covariates, individual trees in the forest algorithm might use one of them more frequently than the other, resulting in attributing to that covariate a higher score in variable importance metrics. However, the two covariates might be equally important in determining the magnitude of the true treatment effect.

3. Implications for the STRI

ML-based methods for causal inference appear to be well-suited to contribute to the research agenda on the effects of services trade reforms on economic outcomes. However, the application presented in this annex has shown that an empirical setting based on the OECD STRI database is not suitable for a successful application of a causal estimation pipeline augmented with ML algorithms.

However, this exercise still provides a useful illustration of the routine required to apply (and evaluate the performance of) ML-based causal inference tools to empirical settings characterised by country-sector-year or country-year variation. This is a recurrent structure in many OECD policy databases, some of which might already feature the coverage and variation required for a better performance of these methods.

In the future, the OECD STRI database itself might allow successful applications of ML-based techniques for causal inference of the kind presented here. Assessing measure-level effects of services trade policies using other methodologies may also be fruitful. To enable such analysis, potential future efforts in data collection within and around the OECD STRI database could be pursued.

A first strategy to increase the relevant variation in services trade policy data could be that of expanding the country coverage of selected individual policy measures to a broader population of developing countries, where policy reforms have happened in more recent years than in OECD countries.

A second direction could be collecting information on selected STRI policy measures for a period before 2014, which is the beginning of the current time coverage in the STRI database. Many of the countries covered by the STRI have implemented relevant services trade policy reforms well before 2014 and a backward expansion of the time coverage for key measures of the database could introduce more policy changes and variation in the data.

Finally, among the few data collection projects that cover at least some dimensions of services trade policy, the joint WTO-World Bank Services Trade Policy Database (Borchert et al., 2019[71]) shares with the OECD STRI several methodological features, including the definition of some individual policy measures[60]. Future collaborations with the WTO and the World bank could be targeted to increase country and year coverage on those policy measures that are consistently defined across the OECD and the WTO-World Bank databases.

[60] Beyond the OECD STRI suite and the WTO-World Bank Services Trade Policy Database other data collection projects covering important dimensions of services trade policy include the ECIPE Digital Trade Estimates Project (Ferracane, Lee-Makiyama and van der Marel, 2018[96]), the Digital Trade Integration project (Ferracane, 2022[95]) developed and maintained by a consortium of academic institutions and international organisations, and the Digital Policy Alert (Evenett and Fritz, 2021[94]).

Annex D. Additional results of downstream effects analysis

Table A D.1. Productivity gains in liberalisation scenario: Air transport

Industry	Estimated effect	Lower bound (90% interval)	Upper bound (90% interval)
Food, Beverages and Tobacco	9.1%	5.3%	13.0%
Textiles, Leather and Footwear	5.8%	3.4%	8.3%
Wood and Products of Wood	14.0%	8.1%	20.2%
Paper and Printing	11.6%	6.8%	16.7%
Coke and petroleum products	4.6%	2.7%	6.5%
Chemicals and chemical products	3.6%	2.2%	5.1%
Basic pharmaceuticals	1.6%	1.0%	2.3%
Rubber and plastic products	11.3%	6.6%	16.2%
Non-metallic mineral products	17.6%	10.2%	25.5%
Basic metals	5.6%	3.3%	7.9%
Fabricated metal products	13.7%	8.0%	19.8%
Computer and electronic products	7.1%	4.2%	10.2%
Electrical equipment	4.4%	2.6%	6.2%
Machinery and equipment n.e.c.	8.4%	4.9%	12.0%
Motor vehicles and trailers	2.6%	1.5%	3.6%
Other transport equipment	4.3%	2.5%	6.1%
Furniture, other manufacturing	18.1%	10.4%	26.3%

Note: This table presents estimated gains in manufacturing labour productivity in the case of a reform scenario concerning trade in air transport services where the STRI score for this sector is assumed to be reduced by 0.05. It provides the detailed results underlying Figure 4.2.
Source: Authors' calculation.

Table A D.2. Productivity gains in liberalisation scenario: Telecommunications

Industry	Estimated effect	Lower bound (90% interval)	Upper bound (90% interval)
Food, beverages and tobacco	2.7%	1.3%	4.1%
Textiles, leather and footwear	2.9%	1.4%	4.4%
Wood and products of wood	5.5%	2.6%	8.5%
Paper and printing	4.3%	2.1%	6.7%
Coke and petroleum products	2.1%	1.0%	3.1%
Chemicals and chemical products	1.4%	0.7%	2.1%
Basic pharmaceuticals	7.8%	3.7%	12.1%
Rubber and plastic products	3.9%	1.9%	6.0%
Non-metallic mineral products	5.0%	2.4%	7.7%
Basic metals	1.8%	0.9%	2.7%
Fabricated metal products	5.3%	2.5%	8.2%
Computer and electronic products	8.4%	3.9%	13.0%
Electrical equipment	2.9%	1.4%	4.4%
Machinery and equipment n.e.c.	4.1%	2.0%	6.3%
Motor vehicles and trailers	0.8%	0.4%	1.3%
Other transport equipment	4.4%	2.1%	6.8%
Furniture, other manufacturing	7.9%	3.7%	12.2%

Note: This table presents estimated gains in manufacturing labour productivity in the case of a reform scenario concerning trade in telecommunications services where the STRI score for this sector is assumed to be reduced by 0.05. It provides the detailed results underlying Figure 4.3.
Source: Authors' calculation.

Table A D.3. Productivity gains in liberalisation scenario: Financial and insurance services

Industry	Estimated effect	Lower bound (90% interval)	Upper bound (90% interval)
Food, beverages and tobacco	1.8%	0.7%	2.9%
Textiles, leather and footwear	1.8%	0.7%	2.9%
Wood and products of wood	1.2%	0.5%	2.0%
Paper and printing	1.0%	0.4%	1.6%
Coke and petroleum products	0.7%	0.3%	1.2%
Chemicals and chemical products	0.6%	0.2%	1.0%
Basic pharmaceuticals	1.8%	0.7%	2.9%
Rubber and plastic products	1.4%	0.5%	2.3%
Non-metallic mineral products	2.5%	0.9%	4.1%
Basic metals	1.4%	0.5%	2.4%
Fabricated metal products	1.6%	0.6%	2.6%
Computer and electronic products	1.3%	0.5%	2.1%
Electrical equipment	1.6%	0.6%	2.6%
Machinery and equipment n.e.c.	1.2%	0.5%	2.0%
Motor vehicles and trailers	0.4%	0.2%	0.7%
Other transport equipment	0.9%	0.3%	1.4%
Furniture, other manufacturing	2.4%	0.9%	3.9%

Note: This table presents estimated gains in manufacturing labour productivity in the case of a reform scenario concerning trade in financial and insurance services where the STRI score for this sector is assumed to be reduced by 0.05. It provides the detailed results underlying Figure 4.4.
Source: Authors' calculation.

Table A D.4. Productivity gains in very ambitious liberalisation scenario: Air transport

Industry	Estimated effect	Lower bound (90% interval)	Upper bound (90% interval)
Food, beverages and tobacco	17.3%	10.0%	25.0%
Textiles, leather and footwear	10.9%	6.4%	15.6%
Wood and products of wood	27.2%	15.4%	40.1%
Paper and Printing	22.3%	12.8%	32.7%
Coke and petroleum products	8.6%	5.1%	12.3%
Chemicals and chemical products	6.7%	4.0%	9.6%
Basic pharmaceuticals	3.0%	1.8%	4.2%
Rubber and plastic products	21.7%	12.4%	31.7%
Non-metallic mineral products	34.6%	19.4%	51.7%
Basic metals	10.4%	6.1%	14.9%
Fabricated metal products	26.6%	15.1%	39.2%
Computer and electronic products	13.5%	7.8%	19.4%
Electrical equipment	8.2%	4.8%	11.7%
Machinery and equipment n.e.c.	15.9%	9.2%	23.0%
Motor vehicles and trailers	4.8%	2.8%	6.7%
Other transport equipment	8.0%	4.7%	11.4%
Furniture, other manufacturing	35.7%	20.0%	53.4%

Note: This table presents estimated gains in manufacturing labour productivity in the case of a very ambitious reform scenario concerning trade in air transport services. In this scenario, the most restrictive country is assumed to implement regulatory changes that halve the gap between its own STRI score in the corresponding sector and the average STRI score across all countries included in the STRI database. Drawing on STRI data for 2020, this would mean lowering the STRI score by 0.09 in air transport.
Source: Authors' calculation.

Table A D.5. Productivity gains in very ambitious liberalisation scenario: Telecommunications

Industry	Estimated effect	Lower bound (90% interval)	Upper bound (90% interval)
Food, beverages and tobacco	13.3%	6.2%	20.8%
Textiles, leather and footwear	14.5%	6.7%	22.8%
Wood and products of wood	29.0%	13.0%	47.2%
Paper and printing	22.2%	10.1%	35.6%
Coke and petroleum products	10.1%	4.7%	15.7%
Chemicals and chemical products	6.7%	3.2%	10.3%
Basic pharmaceuticals	42.4%	18.5%	71.1%
Rubber and plastic products	19.9%	9.1%	31.8%
Non-metallic mineral products	25.7%	11.6%	41.6%
Basic metals	8.7%	4.1%	13.5%
Fabricated metal products	27.7%	12.5%	45.0%
Computer and electronic products	46.3%	20.0%	78.2%
Electrical equipment	14.2%	6.6%	22.4%
Machinery and equipment n.e.c.	21.0%	9.6%	33.6%
Motor vehicles and trailers	3.9%	1.9%	6.0%
Other transport equipment	22.7%	10.3%	36.5%
Furniture, other manufacturing	43.0%	18.8%	72.3%

Note: This table presents estimated gains in manufacturing labour productivity in the case of a very ambitious reform scenario concerning trade in telecommunications services. In this scenario, the most restrictive country is assumed to implement regulatory changes that halve the gap between its own STRI score in the corresponding sector and the average STRI score across all countries included in the STRI database in 2020. In telecommunications services, this illustrative scenario assumes a reduction in the STRI score by 0.24.
Source: Authors' calculation.

Table A D.6. Productivity gains in very ambitious liberalisation scenario: Financial and insurance services

Industry	Estimated effect	Lower bound (90% interval)	Upper bound (90% interval)
Food, beverages and tobacco	14%	5%	23%
Textiles, leather and footwear	14%	5%	24%
Wood and products of wood	10%	3%	16%
Paper and printing	7%	3%	12%
Coke and petroleum products	5%	2%	9%
Chemicals and chemical products	4%	2%	7%
Basic pharmaceuticals	14%	5%	24%
Rubber and plastic products	11%	4%	18%
Non-metallic mineral products	20%	7%	34%
Basic metals	11%	4%	19%
Fabricated metal products	12%	4%	21%
Computer and electronic products	10%	4%	17%
Electrical equipment	12%	4%	20%
Machinery and equipment n.e.c.	9%	3%	16%
Motor vehicles and trailers	3%	1%	5%
Other transport equipment	6%	2%	11%
Furniture, other manufacturing	19%	7%	32%

Note: This table presents estimated gains in manufacturing labour productivity in the case of a very ambitious reform scenario concerning trade in financial and insurance services. In this scenario, the most restrictive country is assumed to implement regulatory changes that halve the gap between its own STRI score in the corresponding sector and the average STRI score across all countries included in the STRI database in 2020. In financial and insurance services, the simulated halving of the gap between the most restrictive country and the average across all countries in the STRI database would be equivalent to a reduction of the most restrictive country's score by 0.14.
Source: Authors' calculation.

Table A D.7. Robustness check with alternative set of controls

	(1)	(2)	(3)	(4)	(5)
	Air transport	Logistics	Telecoms.	Computer services	Financial services
Weighted upstream STRI, t-1	-459.071***	-29.893	-140.405***	-90.508	-136.243***
	(114.511)	(31.911)	(38.692)	(105.607)	(43.954)
Share of tariff lines with international	-0.007***	-0.006**	-0.007**	-0.006**	-0.007***
peaks, manufactured products (%), t-1	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
GDP per capita, t-1	0.232**	0.242**	0.239**	0.239**	0.218**
	(0.097)	(0.094)	(0.094)	(0.094)	(0.096)
Political stability and absence of	0.074***	0.080***	0.063**	0.077***	0.070***
violence, t-1	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Mean product complexity index, t-1	-0.022	-0.010	-0.002	-0.010	-0.040
	(0.059)	(0.060)	(0.059)	(0.059)	(0.059)
Constant	2.511**	1.775*	1.948**	1.845*	2.331**
	(1.012)	(0.947)	(0.940)	(0.958)	(0.982)
Observations	2,844	2,844	2,844	2,844	2,844
R-squared	0.842	0.838	0.840	0.838	0.841
Country F.E.	YES	YES	YES	YES	YES
Industry-Year F.E.	YES	YES	YES	YES	YES

Note: Dependent variable: logarithm of value added per employee. Note that the weighted upstream STRI refers to a different upstream services sector in each of the regressions whose results are shown in this table. The column titles indicate the corresponding upstream services sector whose STRI score was used to calculate the weighted upstream STRI variable. Robust standard errors (clustered at country-industry level) in parentheses. The share of tariff lines with international peaks as well as GDP per capita come from the World Bank World Development Indicators database. The variable on political stability and absence of violence was downloaded from the World Bank World Governance Indicators database. The mean product complexity index was computed based on product-level trade data from the BACI database of the Centre d'études prospectives et d'informations internationales (CEPII) and product complexity data (Hidalgo and Hausmann, 2009[72]) from the Observatory of Economic Complexity (OEC).
Source: Authors' calculations.

Table A D.8. Robustness check with alternative data source for dependent variable

	(3)	(4)	(5)	(6)	(7)
VARIABLES	Air transport	Logistics	Telecoms	Computer services	Financial and insurance services
Weighted upstream STRI, t-1	-252.212***	-84.922**	-44.538	212.003	-84.693*
	(90.701)	(38.051)	(49.311)	(144.318)	(45.656)
Mean output tariffs, t-1	-0.001	0.002	-0.001	-0.004	0.007
	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)
Mean input tariffs, t-1	-0.006	-0.012	-0.007	0.001	-0.017
	(0.032)	(0.031)	(0.031)	(0.030)	(0.032)
Foreign value added share of gross exports, t-1	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Domestic value added embodied in foreign countries' exports, t-1	0.044	0.042	0.039	0.044	0.042
	(0.032)	(0.031)	(0.031)	(0.031)	(0.032)
Observations	2,403	2,403	2,403	2,403	2,403
R-squared	0.776	0.776	0.775	0.775	0.775
Country-Year F.E.	YES	YES	YES	YES	YES
Industry-Year F.E.	YES	YES	YES	YES	YES

Note: Dependent variable: logarithm of value added per employee, based on the UNIDO INDSTAT 4 database. Note that the weighted upstream STRI refers to a different upstream services sector in each of the regressions whose results are shown in this table. The column titles indicate the corresponding upstream services sector whose STRI score was used to calculate the weighted upstream STRI variable. Robust standard errors (clustered at country-industry level) in parentheses.
Source: Authors' calculations.

OECD TRADE POLICY PAPERS

This report was declassified by the Working Party of the Trade Committee in March 2023 and was prepared for publication by the OECD Secretariat.

This report, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Comments are welcome and can be sent to tad.contact@oecd.org.

© OECD (2023)

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at http://www.oecd.org/termsandconditions.