Working Paper DTC-2020-3

Quantifying Trade Law: New Perspectives on the Services Trade Restrictiveness Index

Ben Shepherd, Principal.[1]

This Draft Dated: March 12th, 2020.

Abstract: Measuring the restrictiveness of applied services trade policies is far from straightforward. In addition to identifying policy measures of interest, there is also the problem of weighting and aggregating them into Services Trade Restrictiveness Indices (STRIs). This paper tackles that problem, which has traditionally been solved by using weights determined by analyst or expert judgment. The approach here is novel: a machine learning algorithm is used to determine the weights that have the best predictive power for bilateral trade costs. This alternative approach produces an index with significantly greater explanatory power for bilateral trade than the OECD STRI, using the banking sector as an example. A quantitative simulation shows that the alternative methodology makes a major difference in policy terms: the global impact of a 10% reduction in the restrictiveness of applied services policies is about 10 times larger than the estimated impact using the OECD's STRI.

JEL Codes: F13; F15.

Keywords: Trade in services; Trade policy; Machine learning; Artificial neural network.

[1] The author is grateful to the OECD Secretariat for supplying data, and to seminar participants in Washington, D.C., for helpful comments on initial results.

developing trade consultants
policy • research • capacity building

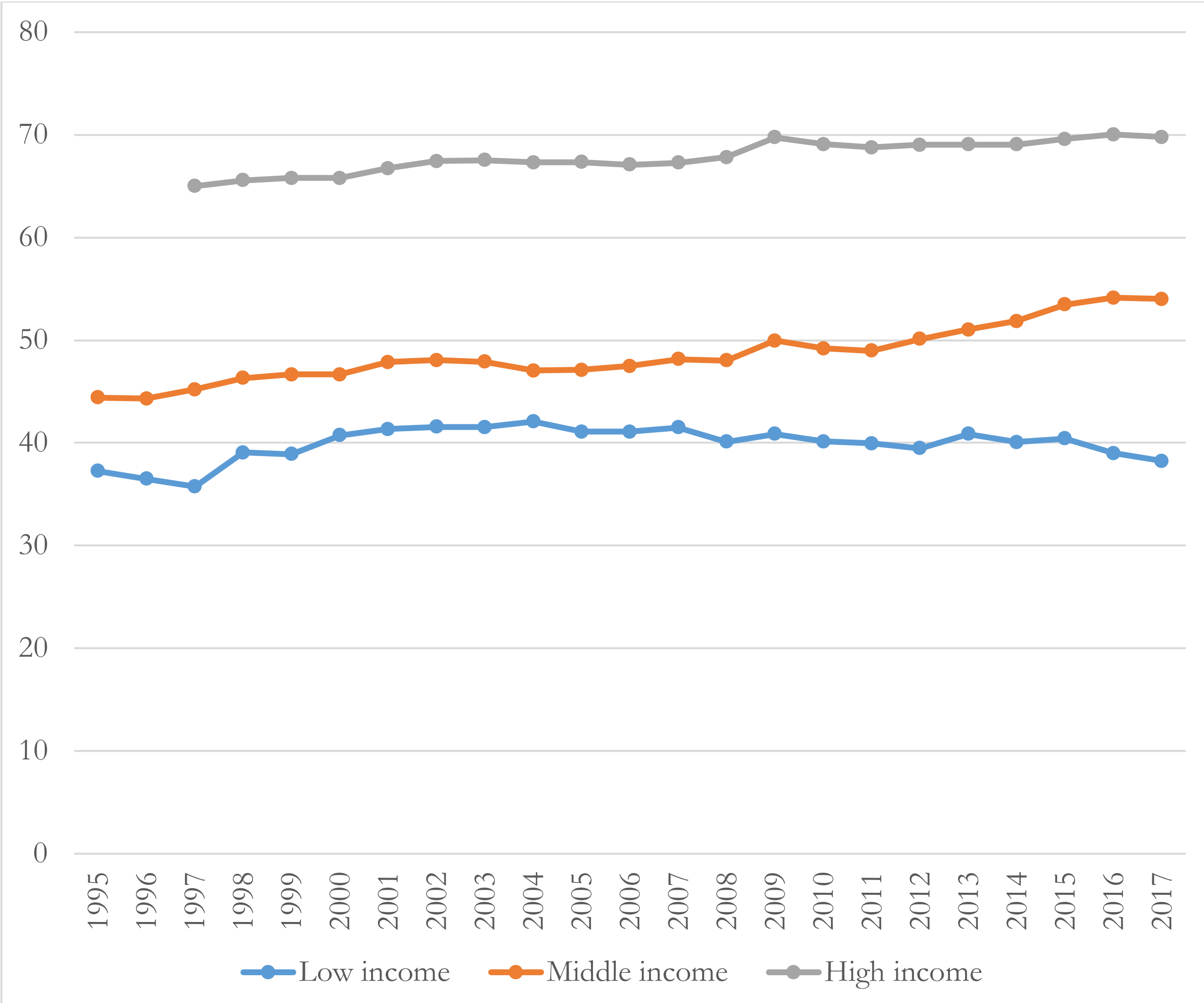
315 W 99th St. #7C
New York, NY 10025

1 INTRODUCTION

Over the medium- to long-term, the world economy is inexorably shifting towards services. Figure 1 clearly shows a trend in countries at all income levels to move towards a services economy. It is more pronounced in middle- and high-income countries, but is in evidence everywhere. Even in low-income countries, services account for an average of 40% of all economic activity, and typically a bit more than that in terms of employment.

A number of structural forces lie behind this process, which is often referred to as "servicification" or "servitization" (see Helble and Shepherd, 2019, for an overview). On the one hand, it has long been recognized that demand shifts towards services as income increases, so the servicification of the world economy in part reflects the remarkable increases in per capita incomes that have been observed in many developing countries in recent years, led by, but not limited to, China and India. But this is not the only factor at play. Also relevant is a technological shift towards the transformation of physical objects into disembodied information flows ("digitalization"). For instance, Shepherd (2020) shows that the rollout of Amazon's Kindle led to substantial decreases in world trade in physical books, but they were more than offset by increased trade in information services: historical trade in physical objects (books) was to some extent replaced by movements of information flows (eBooks).

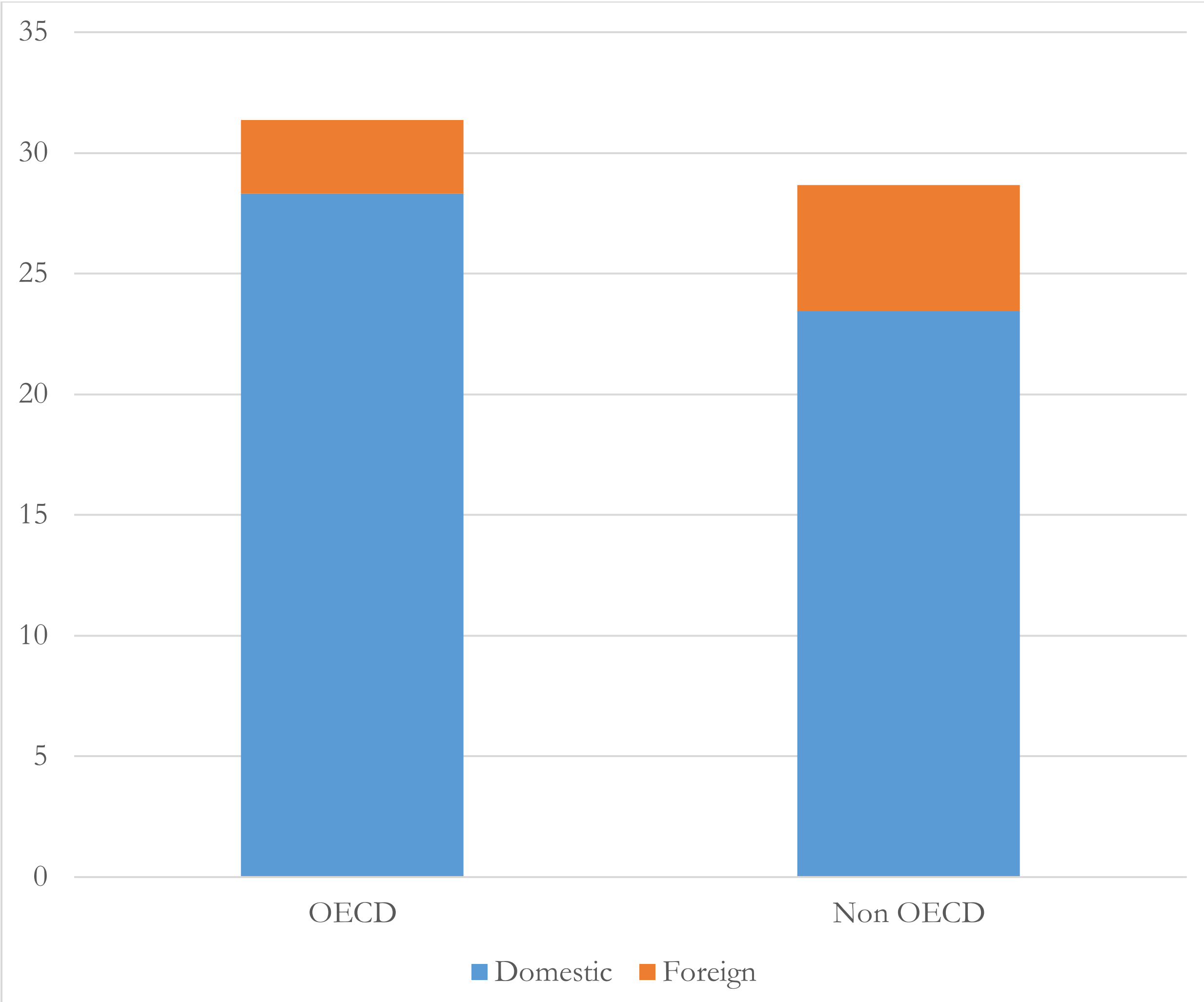
Figure 1: Services as a percent of GDP, by World Bank country income group, 1995-2019.



Source: World Development Indicators.

A final force behind servicification that is less evident in traditional data is the increasing intertwining of services and manufacturing. On the one hand, manufacturers in all subsectors rely increasingly heavily on services inputs sourced from other firms. These inputs include traditional activities like transport and finance, but also relatively new ones such as research and development, engineering (a professional service), and marketing. The net result is that over 30% of the gross value of world trade in manufactured goods is now accounted for by incorporated value added originating in services sectors (Figure 2). Even this figure is an underestimate, as it is derived from input-output tables that only capture extra-firm purchases of services. It would be much higher if it were systematically possible to quantify within-firm services inputs, or services activities performed within manufacturing firms. In particular in high-income countries, but to a certain extent in countries at all income levels, manufacturing firms tend to employ relatively fewer workers on the production line—those performing classic manufacturing activities—and more who are involved in services like design, engineering, after sales service, or even activities like cleaning and security. Firm-level case studies in the Asia-Pacific suggest that once full account is taken of these kinds of transactions, the proportion of services in the total costs of manufacturing firms could be as high as 50% (Low and Pasadilla, 2016).

Figure 2: Value added originating in services sectors, percent of the gross value of manufacturing exports, 2015.



Source: OECD-WTO Trade in Value Added (TiVA) database.

Against this background of a rising tide of direct and indirect services trade, it is striking that researchers know very little about the policy environment that frames services trade in most countries. This point has been made by analysts since the conclusion of the Uruguay Round and the inclusion of services in the World Trade Organization's (WTO's) disciplines through the General Agreement on Trade in Services (GATS). From an analytical perspective, the problem is a non-trivial one: how can we measure hundreds or even thousands of regulatory measures affecting the market incentives of service providers and consumers, and then how can we present that information in convenient summary form?

This paper builds on work by international organizations such as the OECD, the World Bank, and WTO, to examine selected conceptual and practical issues involved in the quantification of services trade policies. Its objective is to highlight the ways in which recent developments in computing technology and statistical methodology can be used to make the task somewhat less daunting than it has historically been seen to be. In particular, I aim to make the process by which policy measures are selected and weighted as systematic as possible, in the interests of increasing transparency and providing a rationale for a unified approach that can be applied in a large number of countries.

The paper proceeds as follows. Section 2 reviews the quantification of restrictions to services trade from a conceptual standpoint, before providing overviews of the major efforts that have been made in this area to date. The following section extends previous work by looking at ways in which newly feasible methods of machine learning can be used to simplify and systematize some of the analytical choices at the core of the procedures discussed in Section 2. Section 4 then provides an example using data for the banking sector in 45 countries. The last section concludes and discusses the possible implications for the policy community.

2 Quantifying Restrictions to Services Trade

With the beginning of the Uruguay Round of GATT negotiations in 1988, there was clear interest among high-income countries in exploring the inclusion of services trade within the ambit of global trade law. By that time, the high income countries were already heavily dependent on their services sector, and saw themselves as having comparative advantage there. At the same time, the traditional work of the GATT had already reduced tariffs in goods sectors within the high-income group to relatively low levels by historical standards. Developing countries saw the matter quite differently, with countries like India being extremely skeptical as to whether or not a future GATS would be in the interests of lower income countries. When negotiations concluded, the "grand bargain" was that services came into the WTO in exchange for further tariff reductions in sectors like agriculture and apparel, but also with the clear understanding that countries would not be required to engage in substantive liberalization as part of the negotiations. The GATS was intended to be a framework for negotiations, rather than an act of liberalization in and of itself.

Besides the political and political economy issues inherent in expanding the ambit of global trade law, trade negotiators and analysts quickly confronted a major technical obstacle. The starting point for liberalization of goods markets under the GATT was to bind and progressively reduce tariffs, simple taxes applied to imports when they cross the border. Such measures are extremely rare in services markets. Instead, the costs a firm must pay to enter a foreign market are typically governed by regulatory measures, similar to what is covered by the catch-all terminology of non-tariff measures (NTMs) in goods markets.

A concrete example helps to fix ideas. Consider a UK law firm looking to provide advice to a client in India. First, there is the question of how such a transaction can occur at all. The GATS identifies four possibilities. Under Mode 1 trade (pure cross-border trade), both parties remain in their respective countries, and the services are provided remotely (by phone or email, for example). An alternative is that the Indian client physically goes to the UK to receive the advice, which is an example of Mode 2 trade (movement of the consumer). More common is that the law firm sends a lawyer from the UK to India to provide advice, after which time she returns to the UK (Mode 4 trade, temporary movement of service providers). But in many services markets, the most common scenario is that the UK firm establishes a local subsidiary in India that then provides services to local consumers (Mode 3 trade, commercial presence).

As this discussion makes clear, there are complex issues of regulation and even jurisdiction involved in these different types of international services transactions. In the legal field, providers typically need to be licensed, so under what, if any, circumstances can a UK-trained lawyer provide advice physically within the territory of India, or even to an Indian consumer? Can a UK firm advertise legal services in ways that are visible to potential clients in India? Are there limits on foreign investment in the legal services sector that would make it more difficult for the UK firm to establish a local subsidiary?

The list of questions like this is potentially limitless, and would cover the full range of regulatory measures that facilitate or inhibit trade in legal services under the four modes identified above. The problem for analysts is two-fold. First, how is it feasible to identify a list of the most important policy measures that affect trade in each sector? And second, how can those measures be weighted and aggregated into summary measures of restrictiveness that can be easily interpreted by policymakers and, potentially, negotiators?

The first major effort to look at applied, rather than bound, services policies was undertaken by the Australian Productivity Commission (APC) in the early 2000s. The APC's framework is still the reference point for work in this area, but the exercise itself was more in the nature of a proof of concept, as the project undertook exploratory work for a variety of countries and sectors, but was not developed into an ongoing project to track services trade restrictiveness. The APC coined the term "Services Trade Restrictiveness Index" (STRI) as the name for their summary index of policy restrictiveness in each sector, and the term has been used extensively by subsequent researchers.

The APC's basic framework (Figure 3) still informs STRI projects today. The first stage is to produce regulatory questionnaires in the sectors of interest. These documents ask questions similar to those posed for legal services above, but proceed in a systematic way. Development of the questionnaires is non-trivial, given the complexity of the modern services economy, the wide range of potentially relevant policies, and the existence of major legal and institutional differences across countries. But in principle, consultation with government regulators, academic experts, and the private sector can greatly inform this process, so that while it is complex and necessarily subject to the risk of being incomplete, it is nonetheless feasible to put together regulatory questionnaires that capture the major policy measures that affect services trade in a given sector.

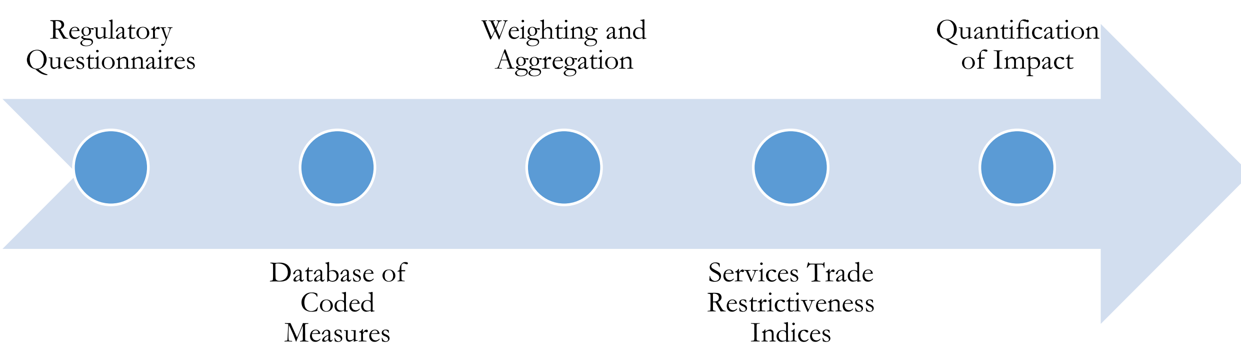
Once the regulatory questionnaires have been designed, they need to be filled in, one sector in one country at a time. Responses have to be coded according to a quantitative key, in which policy measures are assigned a numerical value based on their perceived level of restrictiveness. In some cases, this exercise is straightforward. For limits on foreign equity participation, for example, it is natural to use the percentage of a local firm that can be acquired as a measure of restrictiveness (with a higher score indicating a less restrictive environment). In many other cases, the response is effectively binary (one/zero): if a license is required to practice law, for example, then the questionnaire would be coded as one, whereas if no license is required, it would be coded as zero. While approaches to this stage of the process can differ significantly across implementing agencies, the general trend is to define measures in the regulatory questionnaire as narrowly as possible, so that a maximum number can be given a binary or natural numerical response. Where that approach is impossible, analysts typically define ordinal scales, where a higher score indicates a range of more restrictive possibilities.

Once responses to the individual questions in the regulatory questionnaire have been systematically coded by country-sector pair, the next step is to use that coded database to produce summary measures of restrictiveness, or STRIs. From a technical perspective, the problem is one of weighting and aggregation: analysts need to choose a set of weights applied to individual policy measures in producing summaries, and must select a mathematical function by which the weighted methods are aggregated into a single number.

The weighting and aggregation scheme takes as its input the full set of scores in the coded regulatory database. Its output is a set of STRIs, with scores by country-sector pair. While these measures are convenient measures of restrictiveness, they need to be interpreted carefully. An STRI score of 0.2 on a zero to one scale in one sector is, by definition, more restrictive than a score of 0.1 in the same sector. But, strictly speaking, it is not possible to compare scores across sectors. The reason is that the economic impact of the same measure can be different in different sectors. Similarly, it is important to keep in mind that STRIs are ordinal measures. So while it is possible to say clearly that a score of 0.2 is more restrictive than a score of 0.1 in the same sector, it is impossible to say that it is "twice as" restrictive. Again, the reason is that scores likely do not map to economic impacts in a simple, linear way.

It is recognition of these issues that leads to the final stage of the process. By relating STRIs to appropriate measures of economic performance, it is possible to quantify the impact of services policies in a meaningful way. The APC originally focused on firm-level measures, particularly price-cost margins, with the aim of identifying "rent-creating" and "cost-increasing" policies (see Dee, 2005, for a complete review). Later work has tended to focus on trade data, given the relationship of STRIs to international transactions. A key quantification measure of interest is the ad valorem equivalent (AVE), taken from the literature on NTMs in goods markets. The AVE of a given STRI score is the equivalent ad valorem tariff that would restrict trade to the same degree as the bundle of regulatory measures captured by the STRI. AVEs are not subject to the caveats issued above in relation to interpretation of STRIs: they can be given a cardinal interpretation, and can legitimately be compared across sectors.

Figure 3: The basic STRI framework.



Source: Author.

As the above discussion makes clear, construction of STRIs is extremely labor intensive, and therefore costly. It also requires considerable analytical capacity. Finally, there is a clear interest in standardizing approaches across countries, so that results can easily be compared. These considerations favor the production of STRIs by international agencies, and two major initiatives are currently in operation. The OECD publishes STRIs in 22 sectors for its member countries, which are mostly in the high-income group, along with major emerging markets. In all, the OECD data cover 45 countries from 2014 to 2019, with data updated annually. The second effort was launched by the World Bank. It has resulted in an STRI for 105 countries for 2008 (released in 2012), and jointly with the WTO a database covering 68 countries for 2016 (STRIs released in 2020). The current WTO-World Bank database and World Bank STRI cover 68 countries, of which 45 use data taken directly, with permission, from the OECD database. So the WTO-World Bank database has expanded country coverage by 23 countries.

While the two projects are similar in overall objective, and use the same acronym to describe their core output, there are some important differences along key dimensions (Table 1). Some of these differences, like data sources and collection methodology, are important from a procedural standpoint, but ideally should not have major analytical implications. Others are more important in terms of understanding how the project outputs work, and how they can be interpreted in concrete economic terms. From this perspective, the key decision confronting both sets of analysts was the choice of an appropriate methodology for weighting individual policy measure scores, and aggregating them into STRIs. OECD engaged in a consultative process, engaging regulators, academics, and the private sector. The World Bank drew on that experience by taking the OECD dataset as its starting point, but allowed greater scope for analyst judgment to play a role. In particular, analyst judgment played a major role in deciding on the aggregation methodology and its parameters. Whereas previous approaches have used some kind of (linear) weighted average, the World Bank has departed from that approach by using a constant elasticity of substitution (CES) function to aggregate policy measures. CES functions are ubiquitous in the international trade literature, most commonly as utility functions. The inspiration for the World Bank's approach is perhaps the trade restrictiveness index of Anderson and Neary (19940), which aggregates tariff line measures using a CES functions where the parameters are derived from import demand functions. This approach is rigorous and theoretically grounded, as the authors develop it from first principles. It is a solving a problem, though, that is fundamentally different from that of an STRI: in goods, a trade restrictiveness index seeks to aggregate a large number of tariffs on different products into a single index number in a way that is economically meaningful, such as by keeping welfare constant. The rationale for transposing this approach, if such is indeed its origin, to the services context is unclear. In services, the primary issue is aggregating measures that affect a single traded service, not thousands of individual products. Those measures are not "consumed" in any meaningful sense, and so do not enter directly into a utility function in the way that goods do in the Anderson and Neary (1994) model. Most significantly, CES is just one commonly used functional form among many. The key consideration for an index produced using it is the source of the elasticities that are applied. Anderson and Neary (1994) provide a rigorous basis for estimating those parameters from data in goods markets. The World Bank STRI, on the other hand, simply applies analyst judgment to choose the parameters.

Table 1: Key similarities and differences between OECD and World Bank STRIs.

OECD	46 countries, 22 sectors.
	2,096 data points per country.
	Data collected directly by consultant lawyers.
	Weights based on expert judgment, aggregation based on weighted average.
	Annual updates, less than one year delay.
World Bank	68 countries, 23 sectors.
	2,602 – 5,219 data points per country.
	Data sourced from OECD (45) collected through surveys administered to local law firms (23).
	Weights based on analyst judgment, aggregation based on CES with parameters based on analyst judgment.
	Updates every eight years, with 3-4 years delay.

Source: Author.

So far, the analysis of the OECD and WTO STRIs has focused on the similarities and differences in the ways in which the indices are constructed. But it is also important to address the last stage in the APC process identified above, namely quantification of impacts. OECD work has used a standard economic model to show that a country's STRI score is negatively associated with both its imports and exports of services, and that differences in regulation within the STRI database, referred to as regulatory heterogeneity, also have a negative association with observed bilateral trade (Nordas and Rouzet, 2017). Subsequent work has looked at estimation of AVEs based on observed STRI scores (Benz, 2017). There is thus an important body of evidence showing that the OECD STRI is correlated with bilateral trade flows, and that it can be used to produce estimates of economic impact in standard AVE form. For the World Bank STRI, Jafari and Tarr (2014) produced estimated AVEs from the original (2008) STRI data; there are no such estimates for the 2016 data, nor for the revised version of the 2008 data using the 2016 methodology. The World Bank itself has never produced a model of the economic impact of its STRIs, but van der Marel and Shepherd (2013) use the 2008 version to show that the measures are indeed associated with observed bilateral trade in services, although results are only statistically significant in some sectors.

Currently, only Hoekman and Shepherd (2019) use a standard economic model to examine the full trade and GDP effects of services reform, measured by changes to the OECD STRI—extended approximately to the 23 new countries included in the WTO-World Bank data—and the World Bank STRI. The bottom line is that a simultaneous 10% cut in each country's score on these indices is associated with a GDP increase of 0.5% (OECD) or 0.6% (World Bank). The difference between these two impact assessments is of little interest from an economic perspective, and does not appear to be statistically significant either. In other words, it is not at all obvious from the data that either STRI does a substantially better job of predicting bilateral trade flows than the other, and that many observed differences in the respective index numbers in fact “wash out” as apparent issues related to scaling that are largely corrected once the indices are used in a standard trade modeling framework.

Policy analysts now have two sources of raw data on applied MFN services policies around the world: the original OECD regulatory database (46 countries, 2014-2019), and the new WTO-World Bank regulatory database (45 OECD countries, 23 additional countries, 2016). The availability of this large amount of data is extremely welcome, and should provide developing country researchers, in particular, with the raw material they need to produce policy-relevant research in their own countries. However, most developing countries are still not covered by any data at all: even considering only WTO members, there are 96 countries that are not covered by either dataset. Both organizations are working on expanding geographical coverage, which is again extremely welcome from a research point of view.

However, two issues loom large in moving forward on measuring and quantifying services trade policies around the world. The first is the need for a more uniform approach to collecting and presenting data. While the WTO-World Bank data are, in principle, compatible with the OECD data, the concordance is not always straightforward, and it seems that not all data points in the original OECD data are covered. There would be a clear benefit to the global research and policy communities if the regulatory questionnaires and coding guidelines used to populate the two databases could be harmonized, in the way that NTM data collection in goods was harmonized through the MAST initiative (MAST, 2019). Given the good amount of data now available, the time seems opportune to make use of data-driven methods as one tool that can help identify the most important measures that account for the bulk of observed variation in the current STRIs. The hope is that it may prove possible to substantially reduce the amount of data that needs to be collected in order to produce a “nearly identical” index to currently available STRIs, which would mean that data collection costs would be drastically lowered. Lower costs would mean that existing budgets could cover more countries, thereby providing some first information in areas where, to date, our knowledge of services policies is close to non-existent.

The second issue is a basic one, but which the research and policy communities would benefit from stating explicitly: what economic variable is the STRI supposed to do the best job of explaining? Borchert et al. (2019) do not answer this question in such explicit terms, but their argument is that their STRI is designed to measure the restrictiveness of services trade policies. A similar starting point is implicit in the OECD’s methodology paper (Geloso-Grosso et al., 2015). Measurement is, of course, an important objective in and of itself. But as discussed above, the production of indices as measurement tools is not free from complications: it is impossible to compare scores across sectors at all, and scores can only be compared within sectors across countries on an ordinal, not cardinal, basis. Given the extensive work that has now been done from a measurement perspective, there is scope for the research community to approach the index creation process from a fundamentally different standpoint. For instance, the weighting and aggregation can be posed not as one of simple dimensionality reduction—summarizing thousands of input data points in one output data points—but as an optimization problem: how can these thousands of input data points be aggregated into a single output data point that explains as much of possible of the observed variation in an economic outcome of interest? There has been interest in related approaches before, but they have been approached from the standpoint of technical criteria rather than economic outcomes (e.g., Dihel and Shepherd, 2007). Providing a clear relationship between services policies and an economic variable of interest would greatly assist interpretation by non-technical users, who are confronted with descriptive papers that focus on changes in index scores rather than changes in economic impact (e.g., Borchert et al., 2020).

The remainder of this paper explores these two issues in greater detail, and presents some initial answers based on the application of machine learning tools.

3 How Can Machine Learning Help in Producing STRIs?

The OECD and World Bank STRIs are based on large amounts of input data, in the form of observations of individual policy measures for each country-sector combination. The complete OECD database, for example, consists of 2,096 observations per country per year. Some of those observations are, in effect, repeated, as they capture measures that apply horizontally, i.e. across all sectors. But there is nonetheless an exceptional amount of policy detail in this database. As a result, the cost of collecting it is substantial, in particular for countries in the developing world. Collecting comprehensive sectoral data for a single country can cost tens of thousands of dollars, and requires a considerable level of technical expertise. This exercise is impossible for many developing countries, despite the importance of services and services trade to their economies. As such, they have depended to date on efforts by international organizations to fill the gap.

It is important to distinguish the two main outputs of an STRI project. One is the set of coded answers to sectoral regulatory questionnaires, which constitutes a regulatory database. This output is of value in and of itself, and can often serve as a goldmine of information for qualitative researchers, as well as quantitative modelers. The second important output is the numerical indices—STRIs—that summarize in some way the information in the regulatory database. In light of international experience, the technology for producing the indices can be regarded as reasonably settled. There are detailed issues of methodology that are dealt with differently between the OECD and World Bank versions, but the final outputs correlate strongly (rho = 0.8), which means that they contain substantially similar information; further evidence for this contention is that counterfactual reform simulations using the two different STRIs in fact produce results that are very close (Hoekman and Shepherd, 2019).

From the perspective of developing countries, therefore, the primary barrier to extending coverage of existing STRIs is the cost of data collection for the regulatory database. The World Bank, it would seem, has tried to reduce this cost by using law firm surveys rather than direct collection by consultants. However, this approach has only proved feasible twice in 12 years, and has been associated with long delays—three to four years—between data collection and publication. Experience does not seem to suggest that it is a viable way forward in terms of reducing cost, and the related objectives of improving coverage and observing policies at regular intervals to facilitate a virtuous cycle in the policymaking process, as well as better empirical work.

Hoekman and Shepherd (2019) use the WTO-World Bank regulatory dataset, but started their work before the STRI itself was made public. The problem they addressed was therefore how to rapidly and at low cost create an STRI using this data, ensuring comparability with the existing OECD index in so far as possible. An added element of the task is that the OECD algorithm used to weight and aggregate individual measures into composite STRIs is not fully public, although weights can be obtained by the use of online tools. Concretely, then, the question was how could a large number of individual policy measures be weighted and aggregated into an STRI that could be compared with the OECD STRI in circumstances where the weights could not be directly observed?

Hoekman and Shepherd (2019) approached this task as a prediction problem, as the objective was to use observations on policy measures to predict an STRI score. The set of techniques known as machine learning offers a powerful perspective on such problems. That paper limited consideration to regression-like techniques, specifically Lasso, Ridge, and Elastic Net. Each of these techniques can be understood as similar to a regression problem, as is often solved by ordinary least squares (OLS). However, OLS is unavailable in the STRI context, because the number of individual data points used as inputs is greater than the number of STRI observations that are the output. Thankfully, such situations are common in machine learning applications, where in addition to the key objective of producing accurate predictions, there is also the subsidiary concern of identifying the most important data points for making those predictions, so that the amount of data used as an input can potentially be reduced.

Machine learning refers broadly to a range of statistical techniques that can be used to automate important aspects of the task of prediction. The basic problem is typically to turn some set of inputs into a prediction of one or more outputs. Machine learning algorithms need to be "trained" by using inputs to predict outputs on a sub-sample of the data for which outputs are observed. Performance is then typically assessed by allowing the algorithm to make predictions for observations outside the training sample, and using some criterion of goodness of fit to identify the optimal model. Optimization in most machine learning models refers to selection of variables to use as inputs, as well as the application of weights to those inputs so as to predict the output with as little error as possible.

The three machine learning methods used by Hoekman and Shepherd (2019) follow this general workflow. Each of them can be considered as a regression-based machine learning algorithm, as the fundamental setup is similar to a linear regression problem that can be solved with OLS. There are two major differences. First, the main purpose of an OLS regression is typically inference, not prediction. In other words, the researcher's main purpose is to obtain precise estimates of the relationship between certain input variables and the output variable; predicting the output variable is at most a secondary consideration, whereas in a machine learning context, it is very much the primary objective. Second, the three machine learning techniques referred to above apply a penalty to the coefficients estimated by OLS, which has the effect of "shrinking" some of those estimates towards zero. The result is that some variables effectively drop from the model, to a much greater extent than with OLS. The models apply a tradeoff between information inputs (the more information, the better), and prediction accuracy (the lower the errors, the better) to reach an optimal tradeoff in terms of some criterion function. For example, Hoekman and Shepherd (2019) start from the complete WTO-World Bank dataset for eight sectors that can easily be concorded to OECD sectors. They then create interaction terms, to produce a total of 1,606 variables with informational content, with which to predict 272 observations of the OECD STRI for those eight sectors in the training subsample. Such a problem cannot be solved using OLS, as the number of observations is less than the number of explanatory variables. But it can be solved using Lasso, Ridge, and Elastic Net methods. The authors' preferred technique, the Elastic Net, retains only 59 of the 1,606 variables, but is successful in producing an estimated STRI that correlates very closely with the original OECD index, and explains over 70% of the observed variation in the OECD STRI in the prediction subsample. The approach is therefore highly successful both in quickly and inexpensively extending the OECD index to additional countries even without access to the weighting and aggregation algorithm, and also in identifying a small subset of the available input data that does most of the work in terms of powering the observed STRI, thereby opening up the prospect of a less intensive data collection process in future.

An alternative approach to machine learning is to use an artificial neural network (ANN). These algorithms are the basis of familiar "deep learning" applications in artificial intelligence, from the algorithms that power search suggestions on familiar websites, to image recognition software, to the AlphaGo system that has defeated top ranked human players in the complex game of Go. ANNs do not start from a regression framework, but are instead designed to mimic human thinking on the assumption that it is often based on statistical processes. An ANN receives a set of inputs, applies weights to them through a set of "neurons" in "hidden layers", and produces a prediction that can be compared to an observation in a training dataset. Through a large number of iterations, weights are adjusted so as to produce predictions that are closer and closer the actual observations, typically until there is evidence that further iterations do not materially improve the predictions.

ANNs provide a more flexible framework for machine learning applications than the regression-based techniques considered by Hoekman and Shepherd (2019). The reason is that ANNs can manipulate inputs in complex ways when assigning weights through one or mode hidden layers, with the result that the predicted output can be nonlinear in the input data. While the basic setup of the problem is the same—taking input data (policy measures) and transforming them as accurately as possible into an output (STRIs)—most ANN applications stress prediction accuracy rather than efficiency in data use. While an ANN can be "pruned" to identify the combination of input variables that has the best tradeoff between accuracy and parsimony, typical applications use as much data as possible with the aim of producing highly accurate predictions.

A key difference between an ANN and a regression-based machine learning methodology is that the former is, in some sense, a "black box". A regression model has an intuitive sense behind it, so the selection of variables or estimation of weights can be sense checked against intuition or external information. In an ANN, by contrast, the key processing step is passing data through the hidden layers, which apply weights in potentially complex ways that may not convert easily into intuition based on economic models. The analyst therefore needs to choose the methodology carefully so as to strike the right balance between accuracy and transparency in particular empirical contexts.

Hoekman and Shepherd (2019) was the first paper to apply machine learning to the problem of estimating STRIs from data on individual measures but without the OECD's weighting algorithm. As such, they privileged simple and transparent regression-based methods that rest on foundations familiar to most social scientists. While their approach proved reasonably accurate in terms of reproducing the OECD STRI, it was hampered by the fact that the World Bank and WTO did not collect the full dataset of OECD measures for the 23 additional countries they covered. It appears they collected a subset of data, based on their analyst's views as to the most important policies from a trade perspective. To more comprehensively assess the role of machine learning in this context, it would be desirable to start from the full OECD dataset, as representing the international community's most comprehensive effort to systematically collect data on the services sector. With that in mind, the present paper uses a dataset generously supplied by the OECD Secretariat. It contains the coded response for every policy measure in every sector, for every country, for every year in the STRI dataset. It is the totality of policy information held by OECD as part of its STRI project.

A second way in which the present paper builds on the foundation laid by Hoekman and Shepherd (2019) is that it goes beyond reproduction of the OECD index to ask what economic outcome should be explained by an STRI. Many answers are possible, but the obvious starting point is bilateral trade: given the nature and purpose of an STRI, it should have strong explanatory power in relation to observed bilateral trade flows. Whereas Hoekman and Shepherd (2019) used machine learning to connect individual measures to aggregate STRIs produced by OECD, this paper uses the OECD's dataset of measures to produce an STRI that explains as much as possible of observed bilateral trade. To distinguish this index from the OECD STRI, it is termed "OSTRI" for Optimal STRI. Of course, optimality can only be defined in relation to a benchmark, so in this case, it is limited to having the highest possible level of explanatory power for observed bilateral trade. There is an obvious, though by no means exact, parallel with work such as Anderson and Neary (2003) in the goods space.

A third area of difference with Hoekman and Shepherd (2019) is that the present paper completely leaves to one side the objective of simplifying the task of collecting regulatory data for the STRI. The analysis makes use of all data points in the OECD universe, rather than selecting the most pertinent in terms of their explanatory power so as to privilege model parsimony. Given this choice, as well as the objective of explaining an outcome variable, bilateral trade, that has unknown and potentially complex relationships with the inputs, it is natural to use an ANN rather than a regression-based machine learning model.

The objective of the empirical analysis in the present paper is not to undertake a comprehensive analysis of the links between services policies and bilateral trade. Instead, it is to provide proof of concept for the application of an ANN to solving this kind of problem, in the knowledge that techniques and approaches will be refined over time. To simplify the analysis and speed up computations, consideration is limited to a single sector, commercial banking. The rationale for choosing this sector is that financial services are an important component of global services trade, and there is also a close correspondence between the sectoral definition of the OECD policy data (commercial banking) and the available bilateral trade data (financial intermediation and business services). While the trade data cover a larger aggregate than just commercial banking, it is reasonable to assume that that activity accounts for a substantial share of observed trade. The next section therefore provides a sample empirical application using the banking sector, with the objective of comparing outputs between the OECD approach (STRI), and the ANN-based approach in which policies are linked to observed bilateral trade (OSTRI).

4 Example Application: The Banking Sector

The starting point for the empirical analysis is OECD's coded regulatory database for the banking sector. The full dataset is 27, 324 observations (46 countries * 1 sector * 6 years * 99 policy measures). From this, I take a single year of data, namely 2015. That year corresponds to the most recent bilateral trade data available from the Eora input-output table. I limit consideration to a single year to abstract from time series considerations, and to thereby simplify the model. I reshape the dataset to contain 46 observations, one for each country, with the 99 policy measures separately identified as independent variables.

The trade data come from the Eora multi-region input-output table. The reason for using this source rather than standard trade data sources, such as UN Comtrade, is twofold. First, services trade data in bilateral disaggregation are lacking for many countries. Eora fills in the trade matrix using information from the national accounts and external sources, so that it covers 183 countries and 26 sectors, including financial intermediation and business services as an aggregate. Second, the recent literature on the determinants of bilateral trade flows emphasizes the importance of including self-trade—production that is produced and consumed in the same country—in the model, for a variety of analytical reasons, and this approach is now commonplace (e.g., Baier et al., 2019).

Economists' standard framework for analyzing bilateral trade flows is the gravity model. It takes the following form, considering a single year and single sector cross-section only:

$$(1)\ X_{ij} = F_i F_j t_{ij}^{-\theta} e_{ij}$$

Where: Xij is exports from country i to country j; the F terms are exporter and importer fixed effects; t_{ij} is bilateral trade costs; θ is a parameter capturing the sensitivity of demand to cost; and e_{ij} is an error term satisfying standard assumptions. Numerous theoretical frameworks are consistent with this model, including as the Armington-type model of Anderson and Van Wincoop (2003), the Ricardian model of Eaton and Kortum (2002), and the heterogeneous firms model of Chaney (2008). Arkolakis et al. (2012) and Costinot and Rodriguez-Clare (2014) show that a wide class of quantitative trade models, including the canonical ones just cited, have the same macro-level implications for the relationship between trade flows and trade costs even though their micro-level predictions are quite different.

At first glance, it would seem tempting to simply substitute the 99 policy measures from the OECD database into equation (1) as part of the trade costs term, and estimate the model as usual. A drawback of this approach from the current perspective is that it would not provide an obvious way of retrieving an OSTRI from the data, nor of applying an ANN, given the importance of estimation methodology in obtaining gravity model estimates that are consistent with the constraints imposed by theory (Fally, 2015).

An alternative is to apply the methodology of Novy (2013) to isolate trade costs from the other determinants of bilateral trade. He shows that a model in the general form of (1), but ignoring the error term, can be rearranged and simplified to eliminate the exporter- and importer-specific terms, thereby yielding a simple expression for bilateral trade costs in ad valorem equivalent terms:

$$(2)\ \tau_{ij} = \left(\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}}\right)^{\frac{1}{2}} - 1 \equiv \left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{\frac{1}{2\theta}}$$

Where terms are defined as before. The measure of trade costs τ_{ij} is a geometric average of trade costs between two countries in either direction relative to internal trade costs in both countries. Intuitively, it is clear that the measure is higher when a country trades relatively more with itself than with an external partner. Its form means that it is not a pure measure of the t term in the original gravity model (1), but it nonetheless provides a useful and transparent measure of trade costs that has been used extensively in subsequent work (e.g., Arvis et al., 2016).

Using Eora data, I calculate the trade costs measure in (2) for the financial intermediation and business services sector, as a proxy for commercial banking, for all 33,306 (183 * 182) pairs of countries. In the absence of any published research measuring the trade elasticity θ for this sector, I assume that it is equal to 8.25, which is the average across the sectors analyzed by Caliendo and Parro (2015). In line with previous work, such as Miroudot et al. (2013), trade costs in this sector are high in absolute terms: an average of 367%, with a range from 32% to 733%. Of course, not all of these costs are due to policies that could potentially be liberalized. A significant proportion is likely linked to non-compressible sources of trade costs, particularly geographical and historical factors.

The first step in preparing the trade costs data for an analysis that can produce an OSTRI is to purge them of the influence of non-policy factors. Gravity models of trade typically include data on geographical and historical factors that are believed to influence trade costs, and there is strong evidence over a long period supporting the importance of these factors. Their influence on trade costs needs to be removed so that the OSTRI can then be based on the links between policy factors in the OECD database and the part of bilateral trade costs that is not explained by geographical and historical factors. To do that, I use trade costs as the dependent variable in an OLS regression with geographical and historical variables as regressors, as in Arvis et al. (2016). The error term from that regression can be considered to be the variation in bilateral trade costs that is not accounted for by variation in geographical and historical factors.

The model takes the following form:

$$(3)\ \ln \tau_{ij} = b_0 + b_1 EIA_{ij} + b_2 \ln Distance_{ij} + b_2 Contiguous_{ij} + b_3 Colony_{ij} + b_4 CommonColonizer_{ij} + b_5 CommonLanguage_{ij} + b_6 SameCountry_{ij} + w_{ij}$$

Where τ_{ij} is trade costs between countries i and j, as defined above; EIA is a dummy variable equal to unity if the two countries have a GATS Economic Integration Agreement (EIA) in force between them; Distance is the great circle distance between the commercial centers of the two countries; Contiguous is a dummy variable equal to unity if the two countries share a common land border; Colony is a dummy variable equal to unity if one country was once a colony of the other; CommonColonizer is a dummy variable equal to unity if the two countries were once colonized by the same power; CommonLanguage is a dummy variable equal to unity if at least 9% of the population of each country has a language in common; SameCountry is a dummy variable equal to unity if the two countries were once part of the same country; and w is an error term satisfying standard assumptions. The EIA dummy is sourced from Mario Larch's RTA Database (Egger and Larch, 2008), and the remaining variables come from the CEPII Distance Dataset.

Table 2 presents regression results. Each coefficient shows the extent to which trade costs are increased (positive number) or decreased (negative number) when the corresponding variable is changed. Distance can be interpreted as an elasticity, so a 10% increase in the distance between two countries is associated with trade costs that are 0.2% higher. The dummy variables can be interpreted by exponentiation, so the impact of signing an EIA is to reduce trade costs by exp(-0.302) – 1 = 26%. All explanatory variables have coefficients that are statistically significant at the 5% level or better, except for the dummy indicating country pairs that were once part of the same country. All signs are as expected, except for the common colonizer dummy, which has a surprising positive coefficient.

Table 2: Trade costs regression results.

	(1)
EIA	-0.302 ***
	(0.014)
Ln(Distance)	0.024 ***
	(0.005)
Contiguous	-0.223 ***
	(0.041)
Colony	-0.301 ***
	(0.037)
Common Colonizer	0.113 ***
	(0.010)
Common Language	-0.022 **
	(0.010)
Same Country	0.043
	(0.049)
Constant	6.006 ***
	(0.041)
Observations	33,306
R2	0.089

Note: Estimation is by OLS with dependent variable log(trade costs). Standard errors adjusted for clustering by country pair are in parentheses below parameter estimates. Statistical significance is indicated as follows: * (10%), ** (5%), and *** (1%).

Source: Author.

Using the estimated coefficients, I create residuals from the above regression, the w term in equation (3). This variable then becomes the observed output that the ANN tries to predict using the OECD policy database. I randomly choose 67% of the available data as a training sample, meaning that the ANN uses it to update weights and make predictions based on the OECD policy data. I then conduct a simple cross-validation exercise by comparing simple measures of goodness of fit for the training subsample and the remaining 33% of the data not used in training, known as the prediction subsample. The ANN consists of a single hidden layer with 50 nodes, with that number obtained by applying a simple rule of thumb that suggests a number of nodes in the hidden layer equal to the average of the number of input and output nodes. I run 10,000 repetitions of the ANN algorithm, with weights updated at each iteration and a learning parameter of unity. The model parameters—speed of learning, number of hidden layers, and number of nodes—could all potentially be optimized by grid search using standard cross-validation tools, but I leave that for future research as the primary interest of this paper is in providing proof of concept.

Table 3 summarizes results from the ANN. R2, which summarizes the proportion of observed variation in trade costs that is accounted for by variation in the ANN's output variable, OSTRI, is 0.123 in the training sample, and only slightly lower at 0.107 in the prediction sample. The difference is suggestive of a small amount of overfitting in the model, but the two figures are relatively close so the problem appears to be a minor one. By contrast, these R2s are low by the standards of ANNs, which typically perform very well in prediction tasks. It is possible that performance could be improved by optimizing the model, as mentioned above. But the results also tend to suggest that policy is only one among many factors shaping the pattern of trade costs across countries.

In terms of the model's predictive ability, the root mean squared error (RMSE) provides an indication of overall accuracy. It is very similar between the training and prediction subsamples, and is indeed equal at the two decimal point level. There is therefore little evidence of overfitting based on this statistic. However, differences between the actual and predicted values of log trade costs are substantial: the variable's mean is zero, by definition, with a range from -1.98 to 1.14, so an "average" error of 0.38 is substantial in absolute terms.

Table 3: Goodness of fit statistics for the ANN.

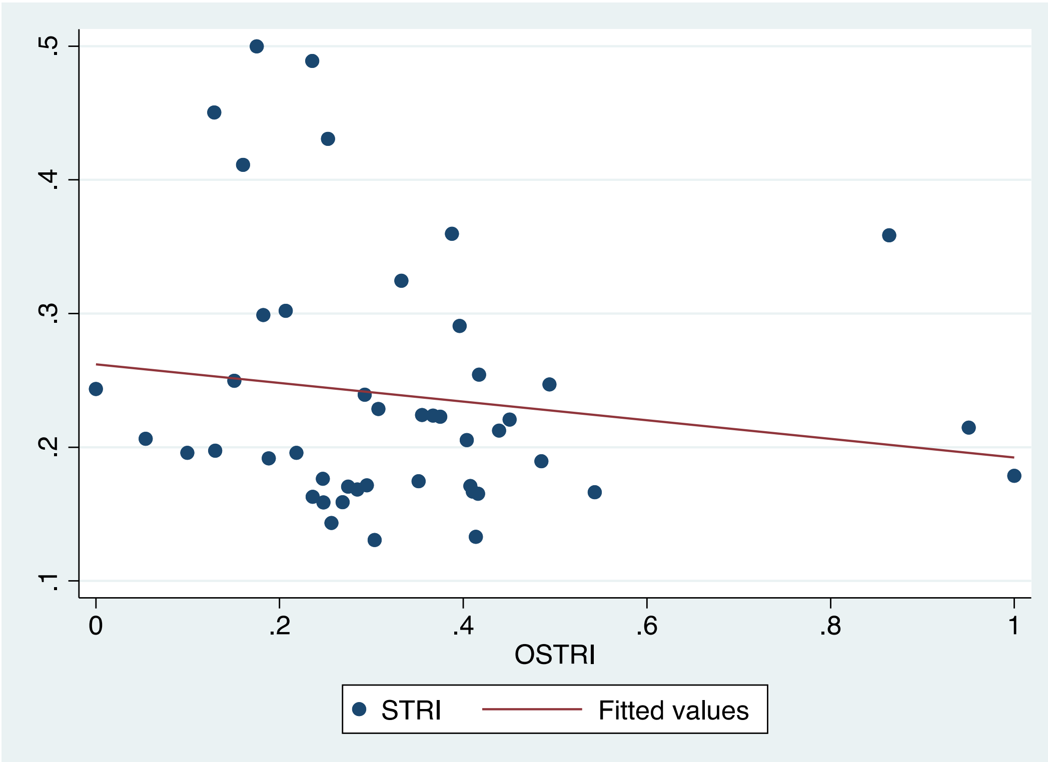
	R2	RMSE
Training	0.123	0.380
Prediction	0.107	0.378

Source: Author.

While there is perhaps scope to improve the model's predictive ability, or to simplify the task by using country-level rather than country-pair data, I take the ANN's output and rescale it to lie between zero and one, defining that index as the OSTRI.

A question that naturally arises relates to the relationship between the STRI and the OSTRI. Figure 4 shows that the correlation between them is in fact negative, which is unexpected. In other words, there is a slight tendency for a higher STRI score to be reflected in a lower OSTRI score, whereas a natural expectation would be for the reverse to be true. The conclusion to be drawn is that the ANN methodology provides a very different way of aggregating the underlying data than the approach used by OECD.

Figure 4: OSTRI vs. STRI.



Source: OECD; and Author.

The key question in deciding between the two approaches should be their performance in empirical settings. To examine this issue rigorously, I use a standard gravity model in the form of equation (1). I specify trade costs in (1) in terms of observables as follows:

$$(4) -\theta \ln \tau_{ij} = b_0 + b_1 EIA_{ij} + b_2 \ln Distance_{ij} + b_2 Contiguous_{ij} + b_3 Colony_{ij} + b_4 CommonColonizer_{ij} + b_5 CommonLanguage_{ij} + b_6 SameCountry_{ij} + b_7 STRI_j * Intl_{ij} + b_8 Intl_{ij}$$

Where variables are defined as above, and STRI is alternately the OECD STRI and the OSTRI, and Intl is a dummy variable equal to one for observations where the exporter and importer are not the same country (i.e., international versus intra-national trade). I estimate the gravity model defined by (1) and (4) on the Eora dataset, using Poisson Pseudo-Maximum Likelihood (PPML), for the reasons set out in Fally (2015) and as per latest practice such as Baier et al. (2019).

Table 4 presents results. The first two columns use the full sample of 183 exporters and 46 importers, while the last two use a square sample of 46 importers and exporters. Results are very similar for the unbalanced and square samples, so the following discussion focuses on the full sample only, namely columns 1 and 2. The most striking difference is between the STRI and OSTRI coefficients: the STRI does not have a statistically significant coefficient, whereas the OSTRI has a coefficient that is negative and statistically significant at the 1% level. In this case, it could be argued that the OSTRI outperforms the STRI in terms of predicting bilateral trade flows, as it is reasonable to expect that a measure of service sector restrictiveness should have a statistically significant and negative coefficient in a standard gravity model. Results on other variables are largely in line with previous work, with historical and geographical controls arguably performing slightly better with the OSTRI than with the STRI.

Table 4: Gravity model regression results.

	(1)	(2)	(3)	(4)
EIA	0.326 ***	0.625 ***	0.314 ***	0.654 ***
	(0.108)	(0.106)	(0.114)	(0.112)
Ln(Distance)	-0.482 ***	-0.519 ***	-0.475 ***	-0.522 ***
	(0.065)	(0.053)	(0.067)	(0.054)
Contiguous	0.400 **	0.639 ***	0.363 **	0.601 ***
	(0.161)	(0.150)	(0.176)	(0.160)
Colony	0.155	0.309 **	0.107	0.297 *
	(0.171)	(0.148)	(0.192)	(0.160)
Common Colonizer	0.353	-0.252	0.256	-0.543 *
	(0.335)	(0.233)	(0.300)	(0.282)
Common Language	0.469 ***	0.115	0.466 ***	0.090
	(0.107)	(0.110)	(0.113)	(0.119)
Same Country	0.292	0.464 **	0.376	0.531 **
	(0.326)	(0.226)	(0.340)	(0.217)
STRI*Intl	-0.729		-0.990	
	(0.687)		(0.751)	
OSTRI*Intl		-5.655 ***		-5.750 ***
		(0.415)		(0.442)
Intl	-5.487 ***	-4.217 ***	-5.430 ***	-4.187 ***
	(0.260)	(0.191)	(0.268)	(0.195)
Constant	24.150 ***	24.381 ***	24.113 ***	24.407 ***
	(0.400)	(0.328)	(0.418)	(0.334)
Observations	8418	8418	2116	2116
R2	0.998	0.998	0.997	0.998
Exporter Fixed Effects	Yes	Yes	Yes	Yes
Importer Fixed Effects	Yes	Yes	Yes	Yes

Note: Estimation is by PPML with dependent variable exports. Standard errors adjusted for clustering by country pair are in parentheses below parameter estimates. Statistical significance is indicated as follows: * (10%), ** (5%), and *** (1%).

Source: Author.

Interpreting the STRI and OSTRI coefficients in Table 4 is not straightforward. As shown by Anderson and Van Wincoop (2003), a change in services policy restrictiveness has indirect (general equilibrium) as well as direct effects, so the estimated coefficient is not an accurate summary of the total impact. I use the solution algorithm developed by Baier et al. (2019) to conduct a simple counterfactual that takes full account of general equilibrium effects. Concretely, I consider a 10% reduction in the restrictiveness of services trade policies in all countries for which data are available, simultaneously. It is obviously a highly stylized example, but serves to fix ideas and provide an indication of the extent to which policy matters for bilateral trade.

Results by country are available by request. At a global level, the counterfactual sees a trade gain of 10.6%. This number is not at all unreasonable, but is higher than what is typically seen in work using the STRI (e.g., Hoekman and Shepherd, 2019). In Table 4, the STRI does not have a statistically significant coefficient, so its impact on bilateral trade is estimated very imprecisely. Subject to this caveat, I conduct the same counterfactual for the STRI to provide a point of comparison for the OSTRI's performance. Focusing again on global results, this exercise suggests that a 10% cut in the restrictiveness of services policies is associated with a trade gain of only 1.6%. Comparing these two numbers shows that aggregation methodology makes a major difference when it comes to assessing the trade impacts: when the emphasis is on aggregating optimally, in the sense of explaining as much as possible of observed bilateral trade costs, the observed sensitivity of trade with respect to policy is around 10 times higher than when aggregation is based to a large extent on expert judgment and opinion.

5 Conclusion and Policy Implications

In a world that is undergoing rapid servicification, services policies are increasingly important to developed and developing countries alike. As such, it is a welcome development that international agencies such as OECD, the World Bank, and WTO are turning their attention to the collection of significant amounts of services policy data. While all involved share the objective of expanding coverage in terms of sectors, countries, and years, there are significant differences between the two major data collection programs currently active. These differences cover not only the countries covered and regularity of updating, but also technical issues surrounding the problem of weighting and aggregating individual policy measures to produce STRIs at the sectoral level.

This paper has argued that there is a role to use a more systematic approach than has previously been possible to support weighting and aggregation efforts, and also to potentially reduce the time and cost associated with data collection. Previous research has not used statistical techniques to derive weights for individual policy measures in aggregate STRIs principally because this kind of problem cannot be solved using standard techniques like OLS. The reason is that there are typically more explanatory variables (policy measures) than there are observations in the dataset. However, this feature of the problem does not prevent the application of machine learning techniques, which is the focus of this paper. Machine learning potentially opens two possibilities for future STRI work. The first is the systematic selection of policy measures based on a tradeoff between explanatory power and parsimony. Hoekman and Shepherd (2019) show that simple approaches can greatly reduce the data collection burden, which translates into time and cost savings, and thus greater coverage for a given investment of public resources.

The second is relating the problem of choosing weights to some criterion of optimality, rather than relying on expert or analyst judgment. That is the emphasis of the present paper, which has used an ANN to derive STRI weights based on the desire to explain as much as possible of the observed variation in bilateral trade costs, taking the banking services sector as an example. While there are many ways in which implementation could be improved in future work, there is a clear proof of concept, in the sense that the policy measures can be aggregated in this way to produce an OSTRI that, although substantially different from the OECD STRI based on the same data, nonetheless has significant explanatory power in a standard model of bilateral trade.

Ultimately, choosing among candidate STRIs is not a straightforward issue. One virtue of the present paper is that it posits an objective clearly and transparently. Other objectives, perhaps better ones, could be used in the same way. A key advantage of this general approach is that it provides a rigorous basis for believing that the policy index explains an economic variable of interest. At the same time, it makes it possible to move beyond competing methodologies in the existing literature that are equally plausible from a purely theoretical standpoint.

Future research could usefully concentrate on examining the impacts of particular policies at the microeconomic level. The reason for doing so is that some criterion is needed to choose among different weighting and aggregation techniques. Examining in detail the economic effects of particular policy measures would help provide a "truth check" on the different magnitudes of economic impacts found in global simulation exercise like the one conducted here. While such exercises are challenging in terms of data requirements and econometric methods, they promise major rewards on a technical level, and would ultimately help support better-informed policy choices.

References

Anderson, J., and P. Neary. 1994. "Measuring the Restrictiveness of Trade Policy." *World Bank Economic Review*, 8(2): 151-169.

Anderson, J., and P. Neary. 2003. "The Mercantilist Index of Trade Policy." *International Economic Review*, 44(2): 627-649.

Anderson, J., and E. Van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170-192.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94-130.

Arvis, J.-F., Y. Duval, B. Shepherd, C. Utoktham, and A. Raj. 2016. "Trade Costs in the Developing World: 1996-2010." *World Trade Review*, 15(3): 451-474.

Baier, S., Y. Yotov, and T. Zylkin. 2019. "On the Widely Differing Effects of Free Trade Agreements: Lessons from Twenty Years of Trade Integration." *Journal of International Economics*, 116: 206-226.

Benz, S. 2017. "Services Trade Costs: Tariff Equivalents of Services Trade Restrictions using Gravity Estimation." Trade Policy Paper No. 200, OECD.

Caliendo, L., and F. Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies*, 82(1): 1-44.

Chaney, T. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review*, 98(4): 1707-1721.

Costinot, A., and A. Rodriguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In G. Gopinath, E. Helpman, and K. Rogoff (eds.) *Handbook of International Economics Volume 4*, Amsterdam: Elsevier.

Dee, P. "A Compendium of Barriers to Services Trade." 2005. Working Paper, https://www.gtap.agecon.purdue.edu/resources/download/3492.pdf.

Eaton, J., and S. Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741-1779.

Egger, P., and M. Larch. 2008. "Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis." *Journal of International Economics*, 76(2): 384-399.

Fally, T. 2015. "Structural Gravity and Fixed Effects." *Journal of International Economics*, 97: 76-85.

Geloso-Grosso, M., F. Gonzales, S. Miroudot, H. Nordas, D. Rouzet, and A Ueno. 2015. "Services Trade Restrictiveness Index (STRI): Scoring and Weighting Methodology." Trade Policy Paper No. 177, OECD.

Helble, M., and B. Shepherd (eds.). 2019. *Leveraging Services for Development: Policies and Prospects.* Tokyo and Manila: Asian Development Bank and Asian Development Bank Institute.

Hoekman, B., and B. Shepherd. 2019. "Services Trade Policies and Economic Integration: New Evidence for Developing Countries." Working Paper DTC-2019-1, Developing Trade Consultants.

Jafari, Y., and D. Tarr. 2017. "Estimates of Ad Valorem Equivalents of Barriers against Foreign Suppliers of Services in Eleven Sectors and 103 Countries." *World Economy*, 40(3): 544-573.

Low, P., and G. Pasadilla. 2016. *Services in Global Value Chains: Manufacturing-Related Services.* Singapore: World Scientific.

MAST. 2019. *International Classification of Non-Tariff Measures.* Geneva: United Nations.

van der Marel, E., and B. Shepherd. 2013. "Services Trade, Regulation, and Regional Integration: Evidence from Sectoral Data." *World Economy*, 36(11): 1393-1405.

Miroudot, S., J. Sauvage, and B. Shepherd. 2013. "Measuring the Cost of International Trade in Services." *World Trade Review*, 12(4): 719-735.

Nordas, H., and D. Rouzet. 2017. "The Impact of Services Trade Restrictiveness on Trade Flows." *World Economy*, 40(6): 1155-1183.

Novy, D. 2013. "Gravity Redux: Measuring International Trade Costs with Panel Data." *Economic Inquiry*, 51(1): 101-121.

Shepherd, B. 2020. "Death by Digital? Evidence from the Global Book Market." Working Paper DTC-2020-1, Developing Trade Consultants.