Revisiting the Trade-Creating Effects of Non-Tariff Barriers

Gabriel Felbermayr (Austrian Institute for Economic Research (WIFO), Vienna University of Economics and Business (WU), CESifo, GEP)

Feodora Teti (ifo Institute, LMU Munich, CESifo)

Discussion Paper No. 459

November 20, 2023

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

Revisiting the Trade-Creating Effects of Non-Tariff Barriers

Gabriel J. Felbermayr[*] and Feodora A. Teti[†]

November 20, 2023

Abstract

Modern regional trade agreements focus on promoting bilateral exchange mostly by lowering non-tariff barriers to trade. But do existing regional trade agreements actually deliver what they promise? This paper argues that existing results in the literature are upward biased because of measurement error in a crucial control variable: tariff rates. Using a novel data set of high-quality tariff information, the paper shows that, on average, non-tariff barriers reductions in deep regional trade agreements boost services trade but not goods trade. Estimating separate non-tariff barrier effects for each regional trade agreement reveals strong heterogeneity: only 23 percent of all regional trade agreements seem to lower non-tariff barriers. For most regional trade agreements, we fail to find any significant effect, while 9 percent appear to reduce trade, possibly because a more balanced regulation evens out comparative advantages. The trade agreements that foster trade the most include non-discriminatory trade policy changes.

Keywords: RTAs, Non-Tariff Barriers, Trade Policy, Tariffs.

JEL-Classification: F13, F14

[*]Austrian Institute for Economic Research (WIFO), Vienna University of Economics and Business (WU), CESifo & GEP, Arsenal Objekt 20, 1030 Wien, Austria; Phone: +43 1 798 26 01 210; E-mail: felbermayr@wifo.ac.at.

[†]ifo Institute, LMU Munich & CESifo, Poschinger Str. 5, 81679 Munich, Germany; Phone: +49 89 9224 1389; E-mail: teti@ifo.de.

This paper has been partly supported by the Umbrella Facility for Trade trust fund (financed by the governments of the Netherlands, Norway, Sweden, Switzerland, and the United Kingdom). Feodora Teti gratefully acknowledges the financial support received from the German Research Foundation through CRC TRR 190 (project number 280092119). All errors are our own.

1 Introduction

Modern trade talks—multilateral rounds at the level of the World Trade Organization (WTO) as well as bilateral negotiations about regional trade agreements (RTAs)—typically center around non-tariff barriers (NTBs) as the average level of tariff protection between most countries is already very low. Trade agreements that include provisions to address NTBs are often referred to as "deep" agreements.

NTBs are numerous and comprise many different types of trade barriers, ranging from customs procedures, rules of origin and local value content requirements, import or export licensing, to market access regulation including behind-the-border topics such as tax questions, technical standards, or sanitary and phytosanitary measures. They are hard to measure precisely, but their trade-inhibiting effects are usually estimated to be more important than that of tariffs (Anderson and Van Wincoop 2004).

Reducing NTBs is controversial because they usually require deep adjustments of domestic regulatory practice (Lamy 2020). Moreover, tackling them is much harder than tariffs due to their arcane nature. The Brexit process between the European Union (EU) and the United Kingdom (UK) provides a telling example: as a former member of the single market, no tariffs between the UK and the EU were in place and the negotiations were therefore exclusively about topics other than tariffs. They have proven to be extremely difficult. Many other modern trade agreements focus on a wide array of NTBs, as discussed, e.g., by Mattoo, Rocha, et al. (2020), and when no agreement can be reached, it is often due to lacking convergence in that area, as the failed transatlantic trade talks have shown.

Despite their prominent role in trade agreements, it is not theoretically clear whether RTAs should actually reduce trade restricting NTBs. First, countries or trade blocks such as the EU often make the granting of tariff concessions in RTAs dependent on trade partners upgrading their regulations to eliminate "unfair" comparative advantages that a more lenient policy stance may confer (Rodrik 2018). So, when two countries align their market access regulations, it is likely that such a move lowers trade if rules are aligned with the most stringent regulations. In this case, trade flows conditional on tariffs could contract upon entry into force of an RTA. Second, more generally, Yu (2000) shows that NTBs can be a complement or a substitute for tariff liberalizations. When countries agree to lower tariffs at the border, there might be strong political pressure to draw up alternative behind-the-border barriers afterward. Third, RTAs often come with bureaucratic burdens, such as the need to prove the origin of goods, which can easily undo the trade-promoting effect of tariff reductions (Felbermayr, Teti, et al. 2019).

Broadly speaking, there are two fundamental approaches to identify the effects of NTBs on trade: a direct and an indirect one. The direct approach collects observable data on NTBs and estimates how their trade-inhibiting role is altered by RTAs. Measuring the stringency of NTBs over time and space is notoriously hard, and relating such changes to RTAs even more so (De Melo and Nicita 2018; Anderson and Van Wincoop 2004). To overcome these difficulties, an alternative approach consists in taking the definition of NTBs literally by assuming that the entire trade effect of RTAs that is not attributable to changes in tariffs must be due to the change in NTBs. Assuming that importers' preferences for foreign versus domestic goods and services are unaffected by signing an RTA, that view takes NTBs as *residual* barriers to trade. Using structural gravity, their change can then be inferred from changes in trade flows—hence, one may talk about an indirect effect.

Much of the gravity literature includes a dummy variable for the existence of an RTA in the estimated equation. Interestingly, however, many papers actually do not control for tariffs and so the dummy measures the full effect of RTAs (thereby confounding changes in tariffs and in NTBs). Studies that apply the indirect approach and do control for tariffs typically find that RTAs do reduce NTBs and boost trade. Often, the effect is almost exclusively driven by deep trade agreements (e.g., Aichele et al. (2016), Dhingra et al. (2021), and Felbermayr, Gröschl, et al. (2018)).

Like the direct approach, the indirect (or residual) approach is sensitive to data issues: the RTA effect can only correctly identify changes in NTBs when changes in tariffs are accurately measured otherwise the estimated coefficient of the RTA variable combines the NTB effect and the tariff effect. However, as shown by Teti (2020), tariffs from standard sources, such as the World Bank's World Integrated Trade Solutions (WITS), suffer from substantial measurement error. As deep RTAs also feature the most comprehensive tariff cuts, it is impossible to say if the RTA effect identified in a gravity equation measures the NTB reduction caused by the new RTA or instead only captures mismeasured tariff reductions.

This paper uses a new database containing carefully cleaned tariff data to apply the indirect approach. Do RTAs really lower NTBs, and if yes, by how much and through which provisions? Or is all the talk about NTBs only distracting from the lowering or elimination of tariffs? We proceed in two steps. First, we estimate the average effect of RTAs on NTBs across many agreements using the newly available Global Tariff Data (new GTD) provided by Teti (2020) that deals with measurement error. In this exercise, we highlight the important role of general globalization trends. Omitting these overstates the role of NTBs. The second part of the paper sheds light on the heterogeneity between RTAs and on the role of specific provisions implemented in those deals. Following Baier, Yotov, et al. (2019), we estimate a separate coefficient for every RTA that enters into force during the period of observation, while controlling for tariffs, thereby allowing every RTA to have a distinct NTB effect.

Our analysis uses the newly available International Trade and Production Database for Estimation (ITPDE, Release 2) put together by Borchert et al. (2020). Those data include domestic trade flows—this is crucial to identify trade regulations that do not vary across trade partners (Heid et al. 2021); usefully, the data also include services trade. The database has almost universal country coverage and a longer time span than alternative sources such as the WIOD provided by Timmer et al. (2015). To measure the depth of RTAs—and hence their ambition regarding NTBs—we build on two sources that are standard in the literature: first, the World Bank's Deep Trade Agreement Dataset (DTA data) put together by Hofmann et al. (2019), second, the DESTA database provided by Dür et al. (2014).

We estimate a state-of-the-art structural gravity model following all recommendations by Yotov (2022). In essence, the strategy consists in running a Poisson Pseudo Maximum Likelihood (PPML) model on panel data of sectoral bilateral trade flows (including domestic transactions) with the largest possible array of fixed effects to control for other determinants of trade.

In contrast to the existing literature, we cannot confirm a robust trade-creating role for non-tariff provisions in trade agreements (TAs) using gravity, while tariffs—once properly measured—are robustly shown to be potent inhibitors of trade. We present some evidence that, in deep TAs, non-tariff provisions create trade in a non-discriminatory fashion. This is welcome from a global welfare-theoretic point of view. However, if policy makers want to use trade agreements to grant preferences between geopolitically aligned countries, tariffs are the better choice. The result is robust to alternative empirical specifications and is not driven by treatment heterogeneity. Our finding does not necessarily mean that RTAs do not reduce NTBs, instead, the lack of a positive estimate of the NTB effect could be simply the result of yet another measurement error. The standard measure of depth assumes the more provisions an RTA has, the deeper and hence, more trade liberalizing it is. However, more provisions might also mean more regulations that in turn can be harmful to trade. Moreover, the finding does not imply that including NTB-provisions in trade agreements is useless as they may yield more efficient allocations albeit with reduced trade flows.

In the second part, we find widely differing effects across the 258 RTAs covered in our sample. Of the estimated NTB effects, 157 are positive with large differences across RTAs, and 101 turn out to be negative. However, most of these effects are insignificant: only 23% of all RTAs seem to lower NTBs, while 9% appear to even reduce trade, possibly a more balanced regulation evens out comparative advantages; the remaining RTA-specific NTB effects turn out to be insignificant. With these RTA-specific results at hand, we check if the content of trade agreements can be used to predict the estimated NTB effects. To measure the content of trade agreements, we use data put together by Mattoo, Rocha, et al. (2020). This is the most comprehensive data set on the content of trade agreements, distinguishing almost 1,000 distinct characteristics. Applying machine learning tools, we use the Least Absolute Shrinkage and Selection Operator (LASSO) method for the selection of specific provisions to check for persistent patterns in the content of trade agreements for RTAs that reduce NTBs. Most interestingly, all characteristics identified by the LASSO technique share some notion of non-discriminatory NTB or refer to explicit regulations of the WTO, i.e. again trade policy that follows the MFN principle—a result that goes in line with our previous finding that TAs successfully lower NTBs, but mostly in a non-discriminatory way.

Our paper relates to the extensive literature on trade policy and NTBs as surveyed, e.g., in Ederington and Ruta (2016). We also contribute to the large and growing literature on the effects of RTAs on trade flows. Earlier work in this field has almost exclusively focused on the overall effect of trade agreements (Baier and Bergstrand 2007). In a related study, Baier, Bergstrand, and Feng (2014) show that the trade-creating effects of RTAs differ with respect to the level of economic integration: the higher the degree of integration, the larger the effects. In their early work Baier, Bergstrand, and Feng (2014) distinguish between non-reciprocal trade arrangements, partial scope agreements, free trade agreements, customs unions, common markets, or economic unions. Baier, Yotov, et al. (2019) implement an even more flexible approach and allow RTA effects to vary not only by trade agreement but also within pairs. Finally, we contribute to a small but growing literature that uses machine learning and other related methods to study the effects of trade agreements in the gravity context (Breinlich et al. 2021; Baier and Regmi 2021).

The rest of this paper is structured as follows. Section 2 discusses the empirical approach typically used in the literature to determine NTB effects of deep trade agreements and highlights how mismeasured tariffs and globalization trends lead to upwards biased results. Section 4 shows how results change when adequately controlling for tariffs and globalization trends. Section 5 then presents our new empirical strategy to flexibly estimate separate NTB effects across all RTAs and shows the results of the Lasso variable selection procedure. Section 6 concludes.

2 Gravity: From Theory to Estimation

In this section, we describe the empirical strategy that allows us to disentangle between the NTB effect and the tariff effect of trade agreements of different types of agreements. We also discuss how omitted variables, especially caused by measurement error in the tariff data and omitting controls for general trends, biases the NTB effect upwards.

2.1 Identification Strategy

To answer the question whether regional trade agreements (RTAs) lower non-tariff barriers (NTBs) we follow the empirical strategy put forward by the literature, i.e., Aichele et al. (2016), Dhingra et al. (2021), and Felbermayr, Gröschl, et al. (2018). The basic idea of the *indirect approach* is to estimate RTA effects using the gravity model while controlling for changes in tariffs to disentangle between the effect of tariffs and NTBs on trade. In this fashion, NTBs can be interpreted as *residual* policy measures that explain the observed pattern of bilateral trade flows given tariffs. A key assumption is that RTAs do not change bilateral preferences between trade partners for their respective goods and services. Using structural gravity, changes in trade costs due to RTAs can, thus, be indirectly identified from trade data.

Next, we formally derive the main estimation equation starting with a structural gravity equation for bilateral trade flows that explicitly models changes in tariffs and NTBs due to RTAs. Head and Mayer (2014) show that a broad class of general equilibrium trade models with constant elasticity of substitution aggregation leads to a gravity equation of the following form:

$$X_{ijst} = \frac{Y_{ist}E_{jst}}{Y_{st}} \frac{\mathcal{T}_{ijst}}{\Omega_{ist}\Omega_{jst}}. \tag{1}$$

Here, X_{ijst} equals the value of exports of country i to country j in sector s at time t; Y_{ist} is country i's value of production in sector s; E_{jst} is country j's expenditure in sector s; Y_{st} is the value of global output. The terms Ω_{ist} and Ω_{jst} are the multilateral resistance terms and account for the effects of third countries' trade costs on i's exporting and j's importing behavior (Anderson and Van Wincoop 2003); $\sigma > 1$ is the elasticity of substitution among goods from different countries. Sectors s are defined as broad sectors, i.e., agriculture, manufacturing, and services. For some specifications, we also do the analysis at a finer level of aggregation.

Finally, $\mathcal{T}_{ijst} \geq 1$ denotes bilateral trade costs between partners i and j. The total trade costs $\mathcal{T}_{ijst}$ can be decomposed into two parts: *i)* ad valorem tariffs denoted by the tariff factor $(1 + \tau_{ijst})$ and *ii)* other bilateral non-tariff trade costs $\Phi_{ijst} \geq 1$.

$$\mathcal{T}_{ijst} = (1 + \tau_{ijst})^{-\sigma_s} \times (\Phi_{ijst})^{1-\sigma_s}, \tag{2}$$

where the exponent $1-\sigma_s$ on non-tariff trade costs arises due to their iceberg nature (trade flows are defined as including the non-tariff trade costs).

Tariffs equal the Most-Favored-Nation (MNF) tariff τ_{jst} and the bilateral preferential tariff τ^*_{ijst}, whenever an RTA is in place. Given that preferential tariffs are always lower than MFN tariffs, τ_{ijst} equals $min\{\tau_{jst};\tau^*_{ijst}\}$. While tariffs are observable, non-tariff trade costs Φ_{ijst} cannot be directly measured in the data but must be estimated instead. Following common practice, we can further split Φ_{ijst} into trade cost shifters unrelated to policy such as geographical or cultural distance (denoted by T_{ij}) and those that are related to trade policy, that we allow to vary across the different types of trade agreements TA_{ijt} with $p \in [deep, shallow, EnCl]$.

$$\Phi_{ijst} = \prod (T_{ij})^{\delta_s} \times exp\left(\sum^{ps} \beta_{ps}\, TA_{ijt,ps}\right), \tag{3}$$

where β_{ps} measures the semi elasticity of trade costs in sector s as captured by type p.

We distinguish between agreements that were notified under the WTO's enabling clause, i.e., non-reciprocal trade arrangements and partial scope agreements that are denoted as $EnCl_{ijt}$, and agreements that were notified under *Article XXIV*, hence, all customs unions and free trade agreements. The Article XXIV agreements can be further distinguished by their depth. $deep_{ijt}$ is a dummy variable that equals one if the trade agreement is a deep TA, and $shallow_{ijt}$ equals one for the remaining agreements.

Substituting (2) and (3) into (1), taking logs, and following the state-of-the-art methods for gravity summarized in Yotov (2022) yields the following estimation equation:

$$\begin{aligned} X_{ijst} = exp[&-\sigma\ \ln(1+\tau_{ijst}) + \beta_1 deep_{ijt} + \beta_2 shallow_{ijt} + \beta_3 EnCl_{ijt} \\ &+ \sum_{t=2001}^{2015} Glob_{ijt} + \mu_{ijs} + \nu_{ist} + \nu_{jst}] + \epsilon_{ijst} \end{aligned} \tag{4}$$

where we pool the data over sectors so that the sector indices of trade policy parameters δ and β disappear.[1] Following Yotov (2022), in some specifications we add time-varying border dummies, $\sum_{t=2001}^{2015} Glob_{ijt}$, that equal one for international and zero for domestic trade flows to flexibly account for trends in trade policy, trading technology or preferences. The time-varying border dummies minimize the danger of endogeneity bias. On the other hand, they also make it harder to identify trade policy parameters because they reduce the available variance. Clearly, it is only possible to include them when domestic trade is observed as well.

[1] In principle, equation 4 can be estimated separately for each sector, or, within the same equation, sector-level parameters can be estimated.

As is now standard in the literature, we include domestic trade flows (Yotov 2021) as well as zero-trade flows, and estimate the model multiplicatively with the Poisson Pseudo Maximum Likelihood (PPML) estimator (Santos Silva and Tenreyro 2006) using a panel of consecutive years (Egger, Larch, and Yotov 2022); standard errors are clustered three-way (importer, exporter, and year) (Larch et al. 2019).[2] We regress exports X_{ijst} of sector s from source country i to destination j on dummy variables that capture the different types of trade agreements, i.e., $EnCl_{ijt}$, $shallow_{ijt}$, and $deep_{ijt}$, while controlling for the effectively applied tariffs $ln\,(1+\tau_{ijst})$. Including tariffs allows us to disentangle between the effects of tariff cuts and changes in non-tariff trade costs due to RTAs. Additionally to the measures for policy changes, we include exporter-sector-time and importer-sector-time fixed effects ν_{ist} and ν_{jst}, which control for multilateral resistances Ω_{ist} and Ω_{jst} (Anderson and Van Wincoop 2003) as well as for any other country-sector specific trends such exchange rate fluctuations or sector-specific demand or technology shocks. These terms also capture all multilateral policy changes (both of tariffs and of NTBs). Lastly, the country pair-sector fixed effects, μ_{ijs}, absorb all possible time-invariant bilateral trade costs consequently controlling for omitted variables such as bilateral distance or common language.

2.2 Accounting for Omitted Variables

Next, we describe how our empirical approach departs from the traditional identification strategy to account for confounding factors, that are especially caused by measurement error in the standard sources for tariff data as well as general trends in trade policy. Both omitted variables substantially bias the NTB effect of deep trade agreements upwards.

The advantage of the indirect approach is that it is less demanding in terms of necessary data: instead of gathering information on changes in NTBs that are hard to quantify and to compare across different types of barriers, we only need information on tariffs and the presence of trade agreements. However, the indirect approach hinges on the correct measurement of tariffs: If tariffs are mismeasured the agreement dummies pick up not only changes in NTBs but also in tariffs happening simultaneously with the implementation of the trade agreement. Deep trade agreements typically reduce tariffs much more, while shallow ones often exclude entire sectors such as agriculture. Hence, the finding of the literature that only deep trade agreements reduce NTBs could be the result of a classical omitted variable problem whenever the empirical approach does not adequately control for tariffs.

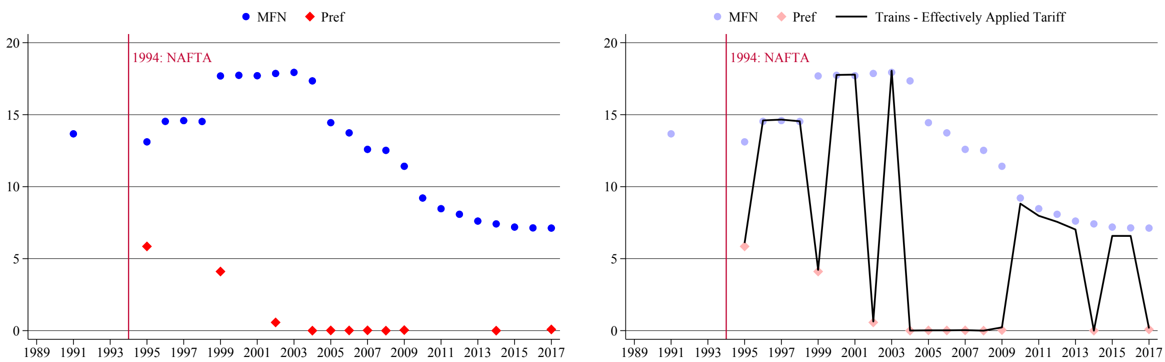
[2] Three-way clustering by importer, exporter, and year is the most conservative approach and yields the largest standard errors. For robustness, we also use alternative clusters that have been used in the existing gravity literature i.e., the pair level.

As shown in Teti (2020), the standard sources for tariff data, i.e., the World Integrated Trade Solution (WITS) that draws on data from the UN and the WTO, suffer from massive measurement error. Existing data include many missing observations because countries do not report tariffs every year. The issue is particularly bad for preferential tariffs and for low-income countries. WITS deals with the missing tariffs by imputing them with MFN tariffs of the respective importing country. For time series data this procedure leads to large irregularities: In years in which the country does not report preferential tariffs, the tariffs jump up from the relatively low preferential tariffs of the previous year to the MFN tariff which is often substantially higher. This type of time variation creates substantial measurement error, in particular in models which identify on time variation as specified in Equation (4)

Figure 1 from Teti (2020) illustrates this point for the tariff that can be found through WITS for exports from the United States to Mexico. Panel (a) shows the simple average of the raw tariffs, MFN and preferential, which can be downloaded through WITS. The average comprises over 5,000 products. The blue dots equal Mexico's MFN tariff, the red dots are the preferential tariffs under NAFTA that entered into force in 1994. The figure shows that neither MFN nor preferential tariffs are available each year. However, the preferential tariffs (red dots) are less likely to be available indicating that Mexico is not reporting the preferential tariffs on imports from the United States every year. The solid line displayed in Panel (b) equals the tariff that WITS reports when one downloads the so-called "effectively applied tariff". Recall, WITS imputes the missing preferential tariffs with MFN tariffs, resulting in the spikes as the effectively applied tariff shoots up to the level of the MFN tariff every time when the data on preferential tariffs are missing but MFN tariffs are available.

Figure 1: Missing Data and Resulting Measurement Error in WITS

(a) Missing Data (b) Measurement Error in WITS



Note: This figure is from Teti (2020) and shows the simple average of the raw MFN and preferential tariffs, as well as the effectively applied tariffs imputed by WITS for the pair Mexico/United States. Details can be found in the corresponding paper.

If we were using WITS to obtain information on tariffs and then estimate Equation (4), we would overstate the NTB effect of the trade agreement because the actually occurring tariff reduction cannot be identified due to the massive measurement error. Instead, the RTA dummy picks up both, the NTB effect and the tariff reduction. If tariff cuts are larger in deep trade agreements than in other ones, the measurement error creates an omitted variable problem that biases the NTB effect upwards: in the extreme, it could be possible that a positive and trade-creating NTB effect could be entirely due to tariff reductions. Therefore, it is crucial to use tariff data that adequately deals with the missing observations in the raw data.

The traditional approach to identifying the NTB effects of deep trade agreements does not control for time-varying unobservable country-pair-specific changes in bilateral trade costs such as general globalization trends that reduce the costs of international relative to domestic trade. As pointed out by Bergstrand et al. (2015), ignoring these forces overstates the effect of RTAs on trade. The resulting omitted variable bias is stronger for trade agreements that have been signed into effect in more recent years when these trends have been more prevalent. RTAs that entered into force more recently, have a higher probability of being a deep trade agreement because the number of policy areas covered in trade deals increases with time: Hofmann et al. (2019) report that the average depth steadily increased starting from 1990 going from an average of 11 provisions for RTAs signed between 1990 and 1994 to 23 provisions for RTAs signed between 2010 and 2015.[3] Hence, not accounting for time trends in trade policy with the help of time-varying border dummies can lead to upwards-biased NTB effects for all RTAs, and the bias is larger for deep trade agreements than shallow ones. Following Yotov (2022), we add time-varying border dummies, $\sum_{t=2001}^{2015} Glob_{ijt}$, that equal one for international and zero for domestic trade flows to flexibly account for trends in trade policy.

2.3 Other Challenges to Identification

Next, we discuss other challenges to identification when quantifying NTB effects and how we deal with them. These issues are not specific to our context but are a problem for any empirical strategy guided by structural gravity that aims at backing out trade policy effects.

Trade agreements are not randomly assigned across country pairs. Baier and Bergstrand (2007) argue that the inclusion of pair-specific fixed effects goes a long way in solving the potential endogeneity bias that results from the non-random treatment of country pairs by trade agreements. In fact, historical ties, geographical proximity, or cultural similarity, which are known to both foster trade and political proximity between countries, are fully accounted for by those terms, as are any other time-invariant country-pair-specific determinants of trade.

[3] Dür et al. (2014) and Kohl et al. (2016) show similar patterns while using different definitions for depth.

Nonetheless, it is still possible that positive trade shocks make the signing and depth of RTAs more likely, creating reverse causation.

The panel structure of our data helps to mitigate concerns about the endogeneity of trade policy, as transitory trade shocks should not affect the likelihood of permanent changes in trade arrangements and permanent trade shocks are well captured by pair-specific fixed effects. Additionally, the inclusion of the $\sum_{t=2001}^{2015} Glob_{ijt}$-terms also helps to reduce endogeneity concerns as they account for all time-varying policy covariates that affect all international trade partners in the same way such as trends in trade policy (i.e., increased probability of successfully signing RTAs in the first place). Moreover, since WTO law obliges RTAs to cover "substantially all trade", uncorrelated sector-specific shocks are unlikely to trigger signing an RTA that covers all sectors. Having said all this, we are aware that some endogeneity bias may still bias our estimates upwards. Given that there are no convincing instruments for RTAs, our (almost) saturated model eliminates that concern as far as possible.

Furthermore, identification relies on the assumption that RTAs do not change deep preferences such that consumers prefer sourcing from an alternative trade partner even if trade costs have not changed. As preferences are unobservable, this assumption cannot be properly tested but it is crucial for the validity of the "indirect" approach to measuring changes in NTBs.

Tariffs are set at the tariff-line level, which corresponds in most countries to eight or even ten-digit-level products. Unfortunately, trade data are only available for higher levels of aggregation.[4] Therefore, in virtually all gravity settings, tariffs need to be aggregated to match the level of aggregation that can be found in the trade data. One might be worried about aggregation bias of the tariff variable. Particularly in cases when tariffs are only reduced for a subset of tariff lines, the aggregated tariff variable might be measured very noisily—in the extreme, it might be impossible to separate the effects of a reduction in tariffs and NTBs. Hence, the RTA-dummy catches both yielding upward-biased estimates for the NTB effect.

Recall that we are differentiating between trade agreements notified under Article XXIV and under the enabling clause; information about the depth of trade agreements is only available for the former. Given the fact that all trade agreements notified under Article XXIV must liberalize "substantially all trade", we are not particularly worried about aggregating tariffs. Instead, for agreements notified under the enabling clause (partial scope agreements and non-reciprocal trade arrangements) we expect a stronger upward bias for the coefficient of interest,

[4] Although for some countries trade data can be found at the same level, it is impossible to gather global coverage.

i.e. aggregation bias in the tariff variable should make it easier to find a significant NTB effect for TA notified under the Enabling Clause.

All the challenges discussed above work in favor of finding a positive NTB-effect. Hence, our NTB coefficients should be interpreted as upper bounds.

3 Data

To revisit the trade-creating effects of NTBs, we combine trade data with information on ad valorem tariffs and on the depth of trade agreements. For tariffs, we use the new Global Tariff Database (GTD) put together by Teti (2020) that deals with misreporting of tariffs (details can be found in the corresponding paper). We aggregate the data that are available at the HS6-digit level to two broad sectors, i.e. agriculture and manufacturing, and for the disaggregated analysis to 18 sectors, respectively; in the area of services, tariffs are set to zero. For the baseline results, we use simple averages, results using trade-weighted averages are available upon request.

We want to relate our new estimates to the ones the literature has reported in the past and get a better understanding of how severe the bias due to mismeasured tariff data is. To do so, we download the effectively applied tariff from WITS.[5] Using Trains as a primary source, we download the effectively applied tariff for ISIC Rev. 3 at the four-digit level and aggregate it to the sector level. We use the ISIC Rev. 3 classification instead of the HS classification because it involves fewer steps for concording across different classifications and nomenclatures to get to the sectors that are used in the trade data, i.e. ITPDE and WIOD.

We use the International Trade and Production Database for Estimation (ITPDE, Release 2) put together by Borchert et al. (2020) for our baseline results for three reasons: first, to estimate Equation 4 theory-consistent and to be able to include trends for globalization, we need data on international and domestic trade flows (Yotov 2021). In contrast to other standard sources for trade data such as UN Comtrade or CEPII's Baci (Gaulier et al. 2010), the ITPDE covers both, inter- and intranational trade. Second, deep trade agreements typically cover, trade in both goods and services (Mattoo, Rocha, et al. 2020) while tariffs are only applicable for trade in goods. It could be possible that trade agreements do not reduce NTBs for goods trade but do so for services. To get a full picture of the trade-creating effect of NTBs we would therefore like to include also services. Lastly, the ITPDE covers the whole world while the comparable

[5] WITS is available at https://wits.worldbank.org/.

WIOD data only report trade for 44 countries (Timmer et al. 2015). The ITPDE is available through the gravity portal of the US ITC.[6] For further robustness checks, we also use WIOD.

While Release 2 of the ITPDE includes trade data for manufacturing and agriculture since 1989, information on services is only available starting in 2000. Hence, our sample spans the years 2000 to 2015, a time crucial for the development of RTAs: According to the WTO, the number of RTAs in force more than tripled in this period.[7] For computational ease we restrict the sample to the 120 largest countries measured in terms of GDP in 2019 taken from the World Bank's World Development Indicators.[8]

To estimate Equation 4, we further need information on the presence of RTAs and their type. For partial scope agreements, free trade agreements, and customs unions, we draw information from Mario Larch's RTA database (Egger and Larch 2008). For non-reciprocal arrangements the database on Economic Integration Agreements maintained by Jeffrey Bergstrand and Scott Baier is the main source (Bergstrand et al. 2015), we update it to 2014 ourselves using information from the WTO's PTA database.[9] With these two data sets at hand we know whether there was any RTA in force in the first place and we can distinguish between agreements that were notified under the Enabling Clause, i.e. non-reciprocal trade arrangements and partial scope agreements, and those that were notified under Article XXIV, i.e., free trade agreements and customs unions.

Our analysis builds on the World Bank's Deep Trade Agreement Dataset (DTA data) provided by Hofmann et al. (2019) to determine which Article XXIV agreements can be considered to be "deep". The data set on the content of RTAs maps 52 provisions in 279 PTAs notified to the WTO and signed between 1958 and 2015. It also includes information about the legal enforceability of each provision. We define agreements as deep when they cover more than 20 provisions, hence, the variable $deep_{ijt}$ equals one if a trade agreement covers more than 20 provisions and zero otherwise, $shallow_{ijt}$ is one if at most 20 provisions are covered. The richness of the data allows for distinguishing between the simple count of all provisions and those that are legally enforceable. For the second type of depth measure, we classify trade agreements to be deep when they contain at least 20 legally enforceable provisions. These definitions are comparable to Hofmann et al. (2019) and Mattoo, Mulabdic, et al. (2022).

One drawback of the DTA data is their coverage: only agreements that are still in force in 2015 are included. This is mostly an issue for deepening agreements as the shallower ones

---

[6] https://www.usitc.gov/data/gravity/itpde.htm

[7] http://rtais.wto.org/UI/PublicMaintainRTAHome.aspx.

[8] The World Bank's World Development Indicators are available at https://databank.worldbank.org/source/world-development-indicators, we use GDP, constant 2010 US$ (time series: NY.GDP.MKTP.KD).

[9] http://ptadb.wto.org/.

that have been superseded by a deeper one are not included. For example, the association agreements of the EU with the new member states that preceded the Eastern Enlargement are not included. As we do not have any information about their depth, we assume that they are shallow ones.

As an alternative measure for depth, we use on the DESTA database provided by Dür et al. (2014), which focuses on only seven policy areas (c.f., Hofmann et al. (2019) for a discussion of the two methodologies). We classify an agreement as deep whenever the depth index is larger than three. DESTA provides for almost all agreements information about their depth, if no information is available, we assume that the trade agreement is shallow.

4 The NTB Effects of Deep Trade Agreements Revisited

Next, we show how accounting for misreported tariffs and globalization trends affects the finding of the literature that deep trade agreements reduce NTBs. We also discuss potential treatment heterogeneity and perform numerous robustness checks.

4.1 Baseline Results

Table 1 shows the baseline results. We start by estimating Equation (4) without controlling for $(1 + \tau_{ijkt})$, which yields the combined effect of tariff reductions and NTB effects without accounting for globalization trends. The results can be found in column (1). To tease out the pure NTB effect, we control for tariffs, first using information downloaded from WITS (column (2)), then from the cleaned GTD that accounts for misreporting (columns (3) and (4)). Lastly, we add the border dummies to control for globalization trends (columns (5) and (6)). All specifications include importer-sector-time, exporter-sector-time, and pair-sector fixed effects.

Panel (A) includes data for the three broad sectors of agriculture, manufacturing, and services and the years 2000 to 2015. When using all provisions of the World Bank's DTA data to determine the depth of trade agreements, we find that deep agreements increase overall trade by 25% while shallow ones and agreements notified under the Enabling Clause do not activate trade (column (1)). The size of the effect is very standard in the literature. The size of the sample used to obtain these results is large: $N = 474,312$ sector-pair combinations. If one accounts for ad valorem tariffs from the standard WITS data set in the specification, the sample shrinks to $N = 360,522$ observations as there are many missing values.

Specification (2) suggests that deep TAs continue to remain statistically significant determinants of bilateral trade, with a trade-creating effect of 16%. Shallow TAs and the enabling clause still do not appear to matter in a statistically significant fashion. The estimated (absolute value of the) tariff elasticity is 1.08 and different from zero at the 10%-level of significance. To make the effect easier to understand, we translate it into the trade-creating effect of a tariff reduction of 3 percentage points, a very standard magnitude across sectors and trade agreements; a tariff reduction of 3 percentage points adds on average 3.24% to trade.[10] Model (2), therefore, seems to suggest that both the tariff reduction (across all types of agreements) and the non-tariff provisions in deep agreements create trade between countries, but the non-tariff provisions matter more. Note, also, that the point estimate of the tariff coefficient can be interpreted as the trade elasticity σ. The value of 1.08 is implausibly low compared to the usual assumptions in the literature.

Specification (3) keeps the sample size constant but replaces the noisy tariff measure from WITS with the one obtained from the new GTD. The tariff elasticity is estimated much more precisely (the t-value doubles) and doubles in size to 2.99—more closely in line with parameter choices in the macroeconomic literature. This is as expected, as the use of the GTD lowers measurement error in the tariff variable, thereby lowering the attenuation bias that drives the tariff elasticity in Specification (2) towards zero. Now, a 3 percentage point reduction in tariffs boosts bilateral trade by 9%. The effects of the dummy variables, however, no longer differ from zero in terms of usual statistical significance.

This is an unexpected result. It suggests that the measurement bias present in (2) inflates the estimate of the deep trade agreement dummy as the trade-creating effect of the agreements is not appropriately accounted for by the mismeasured tariff variable. Taking Specification (3) at face value, non-tariff provisions in TAs do not *per se* increase bilateral trade. In principle, this can be due to two reasons: either the non-tariff provisions in TAs do not lower trade costs at all, or they lower trade costs in a non-discriminatory (MFN) fashion so that trade costs fall with all trade partners (not just the ones in an FTA). Since we include domestic trade flows in our analysis, so that different treatment of domestic versus international trade flows is possible, the failure to find a significant TA effect would imply that TAs put foreign trade partners at par with domestic ones. In contrast, tariffs—by their very definition discriminatory—have a strong and precisely identifiable effect. One interpretation of this finding is that, if policy makers want to use trade preferences for geostrategic purposes, they should focus on tariff reductions. Deep non-tariff provisions tend to diffuse and benefit all trade partners, potentially including those that the agreement does not directly address.

The larger sample used in Specification (4) resuscitates a marginally statistically significant effect of deep TAs, which now have an independent effect of increasing bilateral trade by 15%.

[10] $(1 + 3/100)^{-(-1.08)} - 1 = 0.0324$

A tariff cut of 3 percentage points increases trade by 8.33%, an effect in the same ballpark as with the smaller sample. Hence, we have strong evidence that (discriminatory) tariff cuts boost bilateral trade, and some, but weaker, evidence for discriminatory trade creation by deep non-tariff trade provisions.

Next, we want to test if TAs might reduce NTBs in a non-discriminatory way. The trade policy areas covered by TAs can be classified into two groups: discriminatory and non-discriminatory. While discriminatory trade policy areas, such as tariffs or countervailing measures, only affect the signatory countries, for other areas discrimination is simply not possible. For example, a reform of the customs procedures also benefits third countries as well (c.f., Mattoo, Mulabdic, et al. (2022) for a more extensive discussion). Our current estimation strategy does not allow to identify any non-discriminatory NTB-reduction due to TAs, which currently is absorbed by the multilateral resistance terms, ν_{ist} and ν_{jst}. However, as shown by Heid et al. (2021), with domestic trade it is possible to identify non-discriminatory trade policy changes, such as changes in MFN-tariffs, by leveraging the fact that trade policy only applies to international trade flows but not to domestic ones.

To quantify the effect of non-discriminatory provisions on bilateral trade, we identify policy areas that make it impossible to discriminate against TA members and third countries, i.e., policy areas with an MFN character. We follow the classification first put forward by Baldwin et al. (2009) and classify the following twelve trade policy areas $m \in [1; 12]$ to be non-discriminatory: Customs, SPS, TBT, STE, State Aid, TRIMs, GATS, TRIPs, Competition Policy, IPR, Investment, and Movement of Capital. Resulting in twelve dummy variables $provMFN_{ijt}^m$ that equal one if the respective MFN-provision is included in the TA and zero otherwise. Since the provisions $provMFN_{ijt}^m$ are non-discriminatory, we only expect importer j to reduce NTBs for the first TA when the respective provision is contained in the legal text of the agreement. Formally, this corresponds to the maximum of the respective MFN-provision m over all trade partners i at time t, i.e., $\max_{i,t}\{provMFN_{ijt}^m\}$.

To identify whether importer j has agreements in place that led to a non-discriminatory reduction of NTBs, we define a new variable, $MFN\ Openness_{ijt}$,

$$MFN\ Openness_{ijt} = \sum_{m=1}^{12} \max_{i,t}\{provMFN_{ijt}^m\} \times international_{ij}. \tag{5}$$

First, we count the cumulated number of the twelve MFN-provisions in trade agreements the first time they get signed. Put differently, whenever a country signs a new TA that includes MFN provisions not included in earlier TAs, the count variable grows by one unit. Note that so far the measure of MFN openness does not have a country pair dimension because MFN provisions are per definition non-discriminatory across all exporters i. Following Heid et al.

(2021), we leverage the fact, that for domestic trade TAs do not change domestic trade costs, and interact the count variable with an international border dummy $international_{ij}$.

Specification (5) shows that adding the MFN Openness variable to Equation (4) undoes the discriminatory effect of deep TAs, while the tariff elasticity is still precisely estimated. The table illustrates the effect of one additional MFN provision in place. In specification (5) this leads to additional trade by 4% with all (international) trade partners. Note that, even with importer and exporter-year-sector fixed effects (present in all specifications), the effect can be identified, because the MFN openness variable varies between domestic and foreign trade partners.

Specifications (6) and (7) are very demanding in that they include time-varying border controls, $\sum_{t=2001}^{2015} Glob_{ijt}$, to account for general trends in the trading technologies (i.e., transportation or telecommunication), bilateral preferences, as well as trends in the content of trade agreements. This further narrows endogeneity concerns, but makes inference difficult as very little variation is left unexplained by the combination of fixed effects and globalization controls. It is, therefore, not overly surprising that, compared to Specification (5), this more demanding model does not lead to different conclusions regarding the effects of non-tariff provisions in trade agreements. While the discriminatory tariffs remain statistically significant, but with lower estimated effects, the MFN openness variable in Specification (7) fails to be statistically different from zero.

Table 1: The Role of Omitted Variables: Tariffs and Globalization Trends

	All Provisions							Leg. Enforcable Provisions						
	(1) No Tariffs	(2) WITS	(3) New GTD	(4) New GTD	(5) MFN	(6) Glob	(7) Glob-MFN	(8) No Tariffs	(9) WITS	(10) New GTD	(11) New GTD	(12) MFN	(13) Glob	(14) Glob-MFN
(A) Baseline (2000-2015), FE: i-s-t, j-s-t, i-j-s														
Deep TAs	0.25***	0.21**	0.14	0.15*	0.07	0.08	0.06	0.27**	0.22*	0.14	0.17*	0.07	0.08	0.06
	(0.10)	(0.09)	(0.09)	(0.08)	(0.07)	(0.06)	(0.07)	(0.12)	(0.12)	(0.12)	(0.10)	(0.08)	(0.08)	(0.08)
Shallow TAs	0.02	-0.01	-0.07	-0.06	-0.04	-0.01	-0.01	0.06	0.03	-0.03	-0.02	-0.02	0.01	0.01
	(0.10)	(0.10)	(0.10)	(0.09)	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)
Enabl. Clause	0.05	0.04	0.00	0.00	0.03	0.04	0.05	0.04	0.03	-0.00	-0.01	0.03	0.04	0.05
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
$ln\,(1+\tau)$		-1.08*	-2.99***	-2.90***	-1.65***	-1.22*	-0.98**		-1.10*	-2.98***	-2.88***	-1.64***	-1.22*	-0.97*
		(0.59)	(0.81)	(0.84)	(0.60)	(0.63)	(0.50)		(0.59)	(0.81)	(0.84)	(0.61)	(0.64)	(0.51)
MFN Openness					0.04***		0.01					0.04***		0.01
					(0.01)		(0.01)					(0.01)		(0.01)
N	474,312	360,522	360,522	474,312	474,312	474,312	474,312	474,312	360,522	360,522	474,312	474,312	474,312	474,312
(B) Long Sample (Manuf. and Agri., 1989-2015), FE: i-s-t, j-s-t, i-j-s														
Deep TAs	0.30***	0.15**	0.05	0.16**	0.03	0.05	0.01	0.36***	0.15*	0.03	0.20**	0.05	0.08	0.02
	(0.08)	(0.06)	(0.05)	(0.08)	(0.06)	(0.06)	(0.07)	(0.10)	(0.09)	(0.09)	(0.10)	(0.07)	(0.09)	(0.08)
Shallow TAs	0.06	0.03	-0.05	-0.04	-0.05	-0.04	-0.05	0.09	0.06	-0.02	-0.01	-0.04	-0.03	-0.04
	(0.09)	(0.11)	(0.11)	(0.09)	(0.07)	(0.08)	(0.07)	(0.08)	(0.09)	(0.08)	(0.08)	(0.06)	(0.07)	(0.06)
Enabl. Clause	0.01	-0.00	-0.05	-0.05	-0.02	-0.01	-0.00	-0.00	-0.01	-0.05	-0.06	-0.02	-0.01	-0.00
	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.04)	(0.04)	(0.06)	(0.05)	(0.06)	(0.05)
$ln\,(1+\tau)$		-1.65***	-3.55***	-3.19***	-1.96***	-1.39***	-1.05**		-1.67***	-3.58***	-3.15***	-1.95***	-1.35***	-1.03**
		(0.54)	(0.62)	(0.58)	(0.51)	(0.43)	(0.45)		(0.54)	(0.63)	(0.58)	(0.52)	(0.44)	(0.47)
MFN Openness					0.05***		0.02**					0.05***		0.02**
					(0.01)		(0.01)					(0.01)		(0.01)
N	683,290	392,839	392,839	683,290	683,290	683,290	683,290	683,290	392,839	392,839	683,290	683,290	683,290	683,290

Note: The information on the depth is provided by Hofmann et al. (2019) (all provisions and only legally enforceable provisions), definitions of the different types of agreements (enabling clause, shallow TA, and deep TA) can be found in the main text. All columns include importer-sector-time, exporter-sector-time, and importer-exporter-sector fixed effects, columns (6), (7), (13), and (14) include time-varying border controls to account for general trends in globalization. In Panel (A) the data includes three broad sectors (agriculture, manufacturing, and services) for the years 2000 to 2015, Panel (B) has no services-trade but a longer time frame (1989-2015). Standard errors are clustered three-way for importer, exporter, and year. * for $p < 0.1$, ** for $p < 0.05$ and *** for $p < 0.01$.

4.2 Robustness and Treatment Heterogeneity

In our gravity analysis, we cannot confirm the robust trade-creating role for non-tariff provisions in TAs that has been reported by the literature so far, while tariffs—once properly measured—are robustly shown to be potent inhibitors of trade, yielding plausible trade elasticities in the estimation. We show next, that this result is very robust to alternative empirical specifications and is not driven by treatment heterogeneity.

We test if the results might be driven by the definition of depth. To do so, we use alternative measures that are standard in the literature. The results remain virtually unchanged in columns (8) to (14), for which we use the number of legally enforceable provisions to determine the depth of trade agreements. When using DESTA (results can be found in Table A1, Panel (A)) to determine the depth, the NTB effects of the "no tariffs" and "WITS" columns are much smaller than when using the World Bank's DTA data—the coefficients almost halve. This result is most likely due to the more precise measure of depth in the DTA data compared to DESTA; nevertheless, the NTB effect of deep trade agreements does not survive when controlling adequately for tariff cuts while the discriminatory changes in tariffs remain to be a significant driver for trade.

The results also hold for a longer time horizon. The second release of the ITPDE extends the number of available years and starts as early as 1989. The advantage of the longer sample is that we can cover more trade agreements entering into force. However, data on services are only available starting in 2000. With this extended data at hand, we redo the baseline analysis including the years 1989 to 2015 for the two broad sectors manufacturing and agriculture. The coefficients, which are shown in Panel (B) of Table 1, report the same pattern as in the baseline and are also robust to alternative definitions of depth (c.f., columns (8) to (14) and Table A1, Panel (B)).

We follow recent methodological developments of the gravity literature and cluster standard errors three-way by importer, exporter, and year (Larch et al. 2019). The three-way clustered standard errors are the most conservative approach. Hence, they work against finding significant NTB effects because the standard errors are relatively large. To make sure that we are not only reporting insignificant effects of non-tariff provisions in deep trade agreements due to conservative clustering, we also cluster standard errors by country-pair, which can be considered as the approach yielding the smallest standard errors and corresponds to the level of the treatment. Despite pushing standard errors down, we still find that the non-tariff provisions mostly matter when non-discriminatory provisions are included, while discriminatory tariffs still have a significant effect. Also clustering by importer-sector, exporter-sector, and year does not alter the results.

Lastly, we want to check if treatment heterogeneity within deep TAs might be the reason for the insignificant and small results of deep TAs when adequately accounting for tariff reductions. To do so, we interact the deep TA variable with five indicator variables, that equal one for *(i)* services trade, *(ii)* G7 countries[11], *(iii)* high-income countries[12], *(iv)* the EU-28 countries, and *(v)* the United States. Figure 2 summarizes the results of the interaction terms, details can be found in the Appendix in Table A3. Most remarkably, almost all coefficients of the respective interaction terms are very close to zero and have large confidence intervals. Only for services we find suggestive evidence that deep trade agreements defined as the number of non-tariff provisions have a positive and significant effect on trade flows. Therefore, our baseline results are not driven by treatment heterogeneity within deep TAs but instead, it seems that simply adding non-tariff provisions to TAs does not have, on average, a positive effect on trade.

5 Heterogeneous NTB Effects: Let the Data Speak

The results so far suggest that deep trade agreements do not have any trade-creating effect on trade through the reduction of NTBs once accounting for tariffs that are corrected for measurement error and globalization trends. Does this mean that trade agreements never reduce NTBs? Not necessarily as it might be simply a question of yet another measurement error: By defining "deep" agreements as agreements with many provisions, we make the implicit assumption that more provisions are automatically better for trade. However, more provisions might also mean more regulations that in turn can actually be harmful to trade. In this section, we first set up an alternative empirical strategy that is independent of the content of trade agreements to determine which RTA has positive NTB effects, and discuss the results. In a second step, we aim at exploring if the content of RTAs can be useful to explain heterogeneous NTB effects using LASSO.

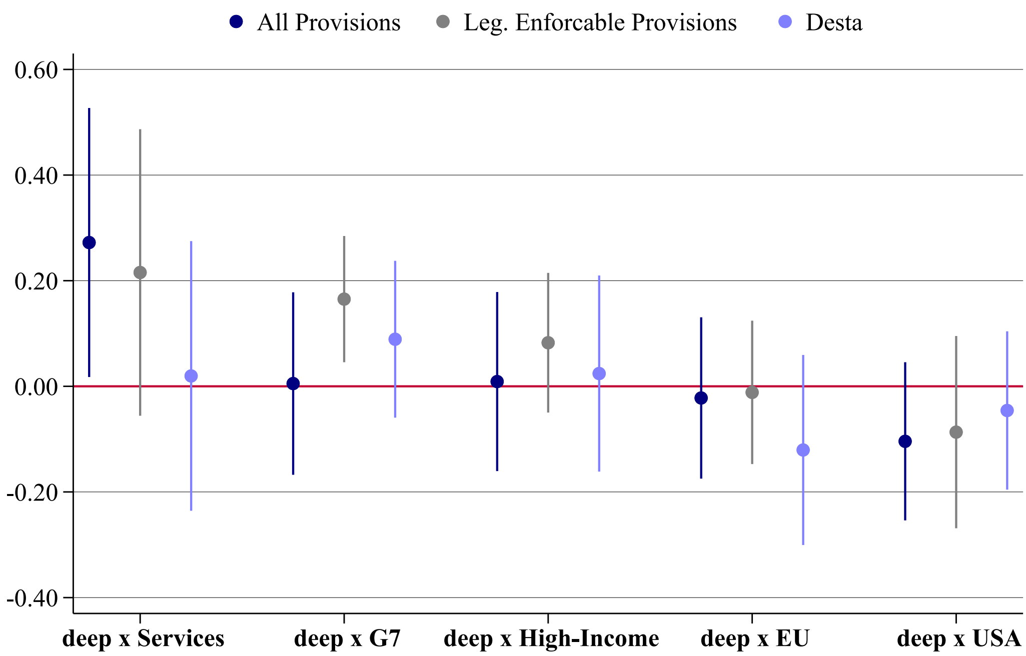
5.1 Estimating RTA-Specific NTB Effects

To check, which trade agreement successfully reduces NTBs, we make Equation (4) more flexible and include every RTA separately into the regression, an approach inspired by Baier, Yotov, et al. (2019) and Kohl et al. (2016). This way, we do not impose any structure on the

[11] The group of G7 countries consists of the European Union (including the United Kingdom), Canada, Japan, and the United States.

[12] We classify the following 43 countries as high-income: the European Union (including the United Kingdom); Australia; Canada; Hong Kong SAR, China; Iceland; Israel; Japan; Liechtenstein; Macau SAR, China; New Zealand; Norway; Singapore; the Republic of Korea; Switzerland; Taiwan, China; and the United States.

Figure 2: Treatment Heterogeneity: Interacting deep TAs



Note: This Figure shows the coefficients when interacting deep trade agreements using the three different definitions for depth (details can be found in the main text) with an indicator variable that takes the value one for *(i)* services trade, *(ii)* G7 countries, *(iii)* high-income countries, *(iv)* the European Union (EU), and *(v)* the United States (USA). The confidence intervals at the 90%-level are included. All estimated coefficients can be found in Table A3.

type of the trade agreement and allow all agreements to have potentially positive—or even negative—effects. We always control simultaneously for tariffs and globalization trends to net out only the NTB effect of the respective trade agreement. Equation (6) formalizes the approach:

$$X_{ijst} = exp[-\sigma \; ln\,(1+\tau_{ijst}) + \sum \beta^{TA} RTA^{TA}_{ijt} + \mu_{ijs} + \nu_{ist} + \nu_{jst} + \sum_{t=2001}^{2015} Glob_{ijt}] + \epsilon_{ijst}. \tag{6}$$

The term $\sum \beta^{TA} RTA^{TA}_{ijt}$ equals the sum of separate dummy variables for every RTA that enters into force during the period of observation. For example, $RTA^{EU-Korea}_{ijt}$ is equal to unity for trade between the members of the European Union and the Republic of Korea in the years after 2011, i.e., when the trade agreement enters into force, and zero otherwise. $\beta^{EU-Korea}$ then gives us the NTB effect of this specific trade agreement, σ determines the average effect of tariff reductions for trade agreements. The remainder of the empirical specification remains the same: the same set of fixed effects are included (μ_{ijs}, ν_{ist}, and ν_{jst}), and we estimate the RTA-specific NTB effects for the sample that covers the years 2000 to 2015 and the three broad sectors manufacturing, agriculture, and services, we use information on tariffs from the new GTD, and include time-varying border controls $\sum_{t=2001}^{2015} Glob_{ijt}$. The approach allows us to separately identify the NTB effect for every RTA while simultaneously controlling for tariffs and other omitted variables such as general globalization trends.

Our baseline results deliver distinct NTB effects for 258 RTAs. Table 2 provides summary statistics for the 258 estimated coefficients. Besides the number of coefficients, the tables also display the mean, median, the 25^{th}- and 75^{th}-percentile. Table A4 in the Appendix shows the same summary statistics when only including estimates that are significant at the 90%-level and at the 99%-level, respectively.

Table 2: Estimated NTB-Effects: Summary Statistics

	All Estimates					95%-Level				
	(1) Nr.	(2) Mean	(3) pc(50)	(4) pc(25)	(5) pc(75)	(6) Nr.	(7) Mean	(8) pc(50)	(9) pc(25)	(10) pc(75)
All	258	0.121	0.089	-0.113	0.319	83	0.308	0.376	-0.218	0.623
Positive	157	0.350	0.276	0.128	0.459	59	0.604	0.515	0.348	0.731
Negative	101	-0.235	-0.167	-0.318	-0.081	24	-0.421	-0.341	-0.467	-0.260

Note: The table summarizes the results for the 258 RTA-specific NTB-effects when estimating Equation (6).

Examining the separately estimated coefficients, shows that the NTB effects are extremely heterogeneous across RTAs: first, although most of the estimated NTB effects are positive, for some RTAs they turn out to be negative. This finding is in line with Baier, Yotov, et al. (2019) who find "widely differing" effects for RTAs, too. Second, from the large difference between the 25^{th}- and 75^{th}-percentile as well as the large difference between the mean and median we can conclude that even within the set of RTAs featuring positive NTB effect, the magnitude varies drastically. Lastly, more than two-thirds of the RTAs do not have an effect on NTBs at all, as the number of insignificant coefficients suggests.

For 68% (175/258) of all RTAs we find insignificant NTB effects. Out of the remaining 83 significant RTA-specific NTB effects, 59 are significantly positive (5%-level) and 24 negative. Hence, we can only identify positive and significant NTB effects for 23% (59/258) of all RTAs that entered into force between 2000 and 2015. For these RTAs, the trade-creating effect due to the reduction of NTBs is large: on average, trade increases due to the RTA by 60%, and the median NTB effect translates into a 51% increase. For 9% of all RTAs, we find negative NTB effects, so, conditional on tariff reductions, countries traded less after the respective RTA entered into force. One potential explanation could be over-regulation through RTAs.

The individual NTB effects for each RTA are summarized in Figure 3, Panel (a). On the x-axis, the 258 RTAs that entered into force between 2000 and 2015 are displayed and ordered by the size of the NTB effect, the y-axis shows the corresponding estimated NTB effects. For the significant agreements, 95% confidence intervals are included. As a robustness check, we estimate Equation (6) using the data starting in 1989 (only manufacturing and agriculture). Panel (b) plots the two sets of NTB coefficients against each other. If all of them were identical, they would perfectly align on the 45-degree line. Eyeballing already tells us that the two estimates are very similar. In fact, the correlation between the two distributions of the NTB-coefficients is 0.85 indicating that the results are robust to extending the period of observation. The comparison between the two sets of coefficients provides an additional interesting insight as it highlights the role of services trade: The NTB effects of the short panel, which includes services trade, are more often larger than the NTB effects of the long panel (only agriculture and manufacturing).

5.2 Predicting NTB Effects Using the Lasso Method

Next, we ask whether we can find persistent patterns in the content of TAs that help predict the NTB effects—positive as well as negative ones. We would like to answer the question of whether there is a systematic difference in the content of the TAs that generate high NTB effects compared to those that have no or even a negative NTB effect. To do so, we need detailed information on the content of TAs. The World Bank has put together a new dataset that analyzes in detail 18 key policy areas for roughly 200 RTAs (Mattoo, Rocha, et al. 2020). For each agreement and policy area, the database provides granular information on the specific provisions covering stated objectives and substantive commitments, as well as aspects relating to transparency, procedures, and enforcement.[13] Mattoo, Rocha, et al. (2020) have coded almost 1,000 distinct characteristics of the content of trade agreements.[14] With this new and detailed data on the content of RTAs at hand we want to determine next if any specific combination of provisions is more prevalent in RTAs with a high NTB effect. To remove ourselves from the model selection problem, we use the Least Absolute Shrinkage and Selection Operator (LASSO) to determine the most important provisions.

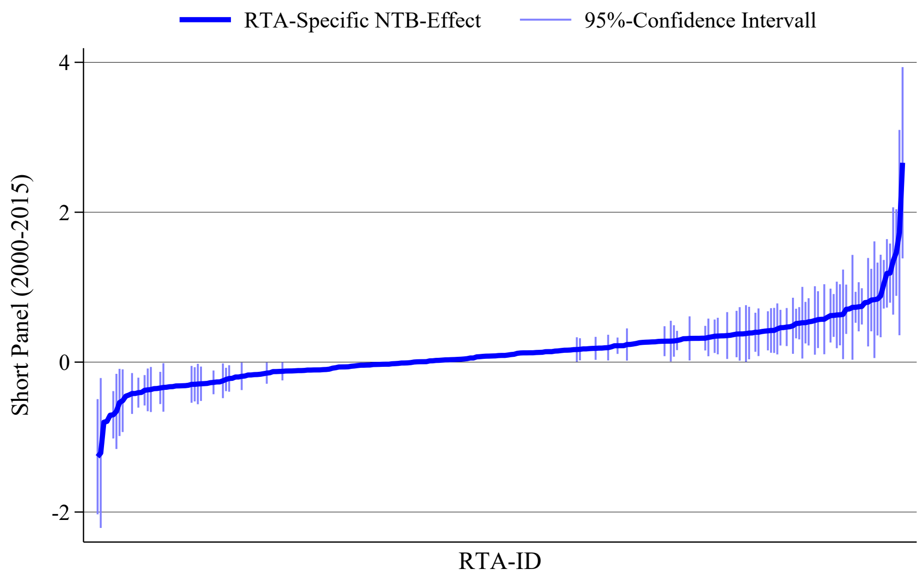
Due to the high dimensionality of the covariates, OLS would surely overfit the data. LASSO offers a way out of this by regularizing the estimation process. More formally, we estimate the parameters by solving the problem

[13] Breinlich et al. (2021) provide a good summary of the data.

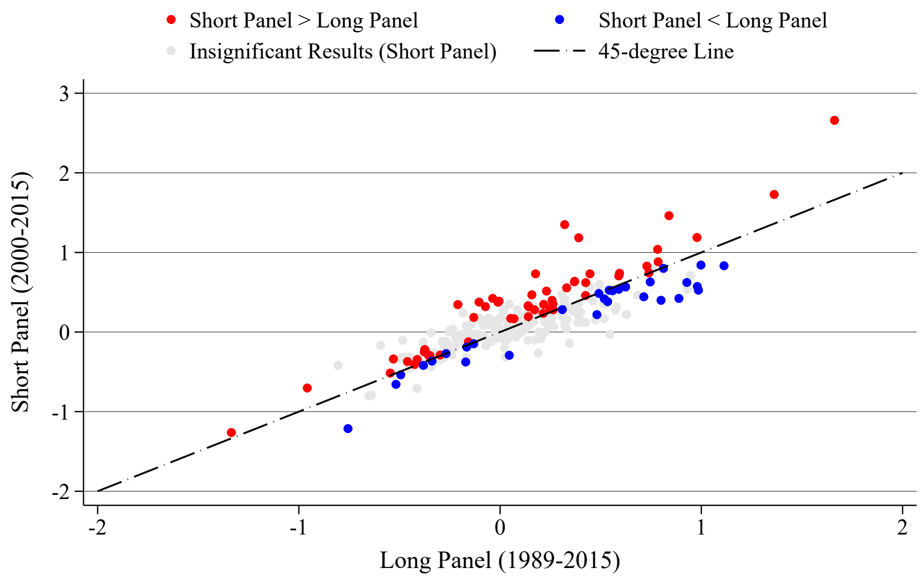
[14] We treat missing observations as not covered by the RTA and set it therefore to zero.

Figure 3: Heterogeneous NTB-Effects across RTAs

(a) Baseline: Short Panel (2000-2015)



(b) Robustness: Long Panel (1989-2015)



Note: This figure summarizes the RTA-specific NTB-effects when estimating Equation (6). Panel (a) shows the baseline results, i.e., when we include all three sectors (agriculture, manufacturing, and services). The confidence intervals are only included for the coefficients that are significant on the 95%-level. Panel (b) correlates the results of the long panel (1989-2015) that only includes agriculture and manufacturing with the coefficients from Panel (a).

$$\min \sum_{i=1}^{n} \left(y_i - \sum_{j}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{7}$$

Thus, compared to OLS, LASSO adds a penalty term $\lambda \sum_{j=1}^{p} |\beta_j|$. λ is the so-called regularization parameter that controls the penalty and minimizes the mean cross-validation error. To alleviate the problems caused by the high dimensionality of the data and the high level of correlation across the included in the agreements, we follow Breinlich et al. (2021) and focus the analysis on a subset of "essential" provisions. Essential provisions are substantial provisions that require specific integration/liberalization commitments and obligations plus the disciplines among procedures, transparency, enforcement or objectives, which are viewed as indispensable and complementary to achieving the substantive commitments (Breinlich et al. 2021). The classification is based on experts' knowledge and is included in the original data provided by Mattoo, Rocha, et al. (2020). We further exclude all provisions that are non-numeric. This leaves us with 273 provisions that we can use for the LASSO.

Table 3 reports the ten most often selected explanatory variables x_{ij}. We perform 50 reps of LASSO to make sure that the chosen characteristics of the TA are not a result of randomness. The ten most often picked provisions are summarized in Table 3. We include all NTB effects (positive, negative, and insignificant ones) to not lose too many observations, which would drastically reduce the sample. Most interestingly, all chosen characteristics have some notion of non-discriminatory NTB or refer to explicit regulations of the WTO, i.e., again trade policy that follows the MFN principle—a result that goes in line with our previous finding that TAs can lower NTBs but mostly in a non-discriminatory way.

Table 3: Variable Selection: Can the Content of Trade Agreements Predict the NTB-Effects?

Description	Policy Area	Total
Does the transfer provision explicitly exclude 'good faith and non-discriminatory application of its laws' related to prevention of deceptive and fraudulent practices?	Movement of Capital	26
Prohibits new export taxes, but with refererence to exceptions mentioned in the provision	Export Taxes	17
Does the Agreement refer to the WTO SPS Agreement?	Sanitary and Phytosanitary	16
Does the agreement specify supremacy of MEA obligations over PTA obligations?	Environmental Laws	14
Prohibits voluntary export restraints inconsistent with GATT Article VI	Export Taxes	14
Freedom of transit for goods	Trade Facilitation and Customs	12
Origin verification measures	Trade Facilitation and Customs	10
Recognizes the Joint Recommendation Concerning Provisions on the Protection of Well-Known Marks	Intellectual Property Rights	9
Does the agreement contain explicit provisions on the prohibition of offsets?	Public Procurement	8
Prohibits all export taxes between the Parties, but with reference to certain exceptions mentioned in the provision that are WTO-plus	Export Taxes	8

Note: The tables shows the ten most often picked characteristics that can predict best the RTA-specific NTB-effects when estimating Equation 6. We perform LASSO 50 times, column (3) shows the total number of times the respective characteristic has been chosen.

6 Conclusion

Deep non-tariff provisions in trade agreements have gained a lot of policy attention. In this paper, we revisit an old question: do RTAs lower NTBs? In our gravity analysis, we do not find a robust independent trade-creating role for them, while tariffs—once properly measured—are robustly shown to be potent inhibitors of trade. Crucially, to obtain this result, measurement error in the tariff data, which tends to inflate estimates of NTBs, must be removed. Our result suggests that, across a large set of trade agreements, tariff reductions are still quantitatively important drivers of trade creation while non-tariff provisions might be less trade-creating. This does not mean, however, that such provisions are not important as they may actually reduce distortions in bilateral trade. Moreover, they can be the price to be paid for tariff reductions.

Interestingly, we find that, in deep FTAs, non-tariff provisions create additional trade in a non-discriminatory fashion. This is welcome from a welfare-theoretic point of view as inefficient trade diversion effects can be avoided. RTAs with this function can be seen as building blocks of multilateral trade liberalization. However, if policy makers want to use trade agreements to grant preferences between geopolitically aligned countries, tariffs are the better choice.

7 Comments & To Do's

Comments

- convince people that indirect approach is better than direct approach (especially the direct approach guys) –> residual approach
- could it be that to get market access to partners like the EU countries have to agree to all the deep provisions? "tit for tat"
  - is there a way to show this more formally? it would have strong policy implications
- how to address endogeneity?
  - bilateral time fixed effects are not possible because that would absorb our NTB effect, wouldn't it? We have the boarder trends + pair FE but I honestly do not know what else we can do here...
  - we could include pair-trends (linear, square and cubic) BUT this will absorb NTB effects

To Do's

- consistent terminology: RTAs vs. FTAs vs. TAs
- EQ (3): TAs do not vary across sectors, why do we need the s index?
- we should estimate EQ (4) by sector
- can we test if negative NTB effects occur whenever there is an FTA between north-south, i.e. whenever high-income countries force low-income countries to level the regulatory playing field?
- how should we cluster standard errors? Treatment is at the pair level for RTAs, but pair-sector for tariffs, threeway clustering might be too conservative (works against finding an effect for deep)
- missing robustness checks
  1. WIOD aggregated
  2. BACI aggregated
  3. disaggregated sectors: ITPDE short, long, WIOD, and BACI
  4. trade-weighted averages for tariffs: ITPDE short
  5. include interaction between MFN and deep
- missing stuff for the appendix
  1. list of 120 countries in sample
  2. list of 18 sectors plus concordance to 3 broad sectors plus con-cordance to ISIC 3
  3. add list of RTA-specific effects to the Appendix

References

Aichele, Rahel, Gabriel J. Felbermayr, and Inga Heiland (2016), "TTIP and Intra-European Trade: Boon or Bane?" *Jahrbucher fur Nationalokonomie und Statistik* 2366, 639–664.

Anderson, James E. and Eric Van Wincoop (2003), "Gravity with Gravitas : A Solution to the Border Puzzle." *The American Economic Review* 931, 170–192.

— (2004), "Trade Costs." *Journal of Economic Literature* 423, 691–751.

Baier, Scott L. and Jeffrey H. Bergstrand (2007), “Do Free Trade Agreements actually Increase Members’ International Trade?” *Journal of International Economics* 711, 72–95. URL: https://ideas.repec.org/a/eee/inecon/v71y2007i1p72-95.html.

Baier, Scott L., Jeffrey H. Bergstrand, and Michael Feng (2014), “Economic Integration Agreements and the Margins of International Trade.” *Journal of International Economics* 932, 339–350. URL: http://dx.doi.org/10.1016/j.jinteco.2014.03.005.

Baier, Scott L. and Narendra Regmi (2021), “Using Machine Learning to Capture Heterogeneity in Trade Agreements.” *Mercatus Working Paper*.

Baier, Scott L., Yoto V. Yotov, and Thomas 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. URL: https://doi.org/10.1016/j.jinteco.2018.11.002.

Baldwin, Richard E., Simon Evenett, and Patrick Low (2009), “Beyond Tariffs: Multilateralizing Non-Tariff RTA Commitments.” *Multilateralizing Regionalism: Challenges for the Global Trading System*. Ed. by Richard E. Baldwin and Patrick Low. Cambridge University Press.

Bergstrand, Jeffrey H., Mario Larch, and Yoto V. Yotov (2015), “Economic Integration Agreements, Border Effects, and Distance Elasticities in the Gravity Equation.” *European Economic Review* 78, 307–327. URL: https://www.sciencedirect.com/science/article/pii/S0014292115000847.

Borchert, Ingo, Mario Larch, Serge Shikher, and Yoto V. Yotov (2020), “The International Trade and Production Database for Estimation (ITPD-E).” *International Economics*.

Breinlich, Holger, Valentina Corradi, Nadia Rocha, Michele Ruta, J.M.C. Santos Silva, and Tom Zylkin (2021), “Machine Learning in International Trade Research.” *World Bank Policy Research Working Paper 9629* April. URL: https://openknowledge.worldbank.org/handle/10986/35451.

De Melo, Jaime and Alessandro Nicita (2018), “Non-Tariff Measures: Data and Quantitative Tools of Analysis.” *F.R.E.I.T. Working Paper*. URL: UNCTAD/DITC/TAB/2017/2.

Dhingra, Swati, Rebecca Freeman, and Hanwei Huang (2021), “The impact of non-tariff barriers on trade and welfare.” 17.

Dür, Andreas, Leonardo Baccini, and Manfred Elsig (2014), “The Design of International Trade Agreements: Introducing a New Dataset.” *Review of International Organizations* 93, 353–375.

Ederington, Josh and Michele Ruta (2016), "Nontariff Measures and the World Trading System." *Handbook of Commercial Policy 1b*. Ed. by Kyle Bagwell and Robert W. Staiger. 1st ed. Vol. 1A. Elsevier B.V., 211–277. URL: http://dx.doi.org/10.1016/bs.hescop.2016.04.010.

Egger, Peter H. and Mario Larch (2008), "Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis." *Journal of International Economics* 762, 384–399. URL: http://dx.doi.org/10.1016/j.jinteco.2008.08.003.

Egger, Peter H., Mario Larch, and Yoto V. Yotov (2022), "Gravity Estimations with Interval Data: Revisiting the Impact of Free Trade Agreements." *Economica* 89353, 44–61.

Felbermayr, Gabriel J., Jasmin Gröschl, and Thomas Steinwachs (2018), "The Trade Effects of Border Controls: Evidence from the European Schengen Agreement." *Journal of Common Market Studies* 56213, 335–351. URL: http://www.cesifo-group.de/DocDL/wp-2016-213-felbermayr-groeschl-steinwachs-trade-effects-of-border-controls.pdf.

Felbermayr, Gabriel J., Feodora A. Teti, and Erdal Yalcin (2019), "Rules of Origin and the Profitability of Trade Deflection." *Journal of International Economics* 121, 103248.

Gaulier, Guillaume, Julien Martin, Méjean Isabelle, and Soledad Zignago (2010), "BACI: International Trade Database at the Product-Level. The 1994-2007 Version." CEPII Working Papers No. 2010-23. URL: http://www.cepii.fr/CEPII/fr/publications/wp/abstract.asp?NoDoc=2726.

Head, Keith and Thierry Mayer (2014), "Gravity Equations: Workhorse,Toolkit, and Cookbook." *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Elsevier B.V. Chap. 3, 131–195.

Heid, Benedikt, Mario Larch, and Yoto V. Yotov (2021), "Estimating the Effects of Non-discriminatory Trade Policies within Structural Gravity Models." *Canadian Journal of Economics* 541, 376–409.

Hofmann, Claudia, Alberto Osnago, and Michele Ruta (2019), "The Content of Preferential Trade Agreements." *World Trade Review* 183, 365–398.

Kohl, Tristan, Steven Brakman, and Harry Garretsen (2016), "Do Trade Agreements Stimulate International Trade Differently? Evidence from 296 Trade Agreements." *World Economy* 391, 97–131.

Lamy, Pascal (2020), "Foreword." *Handbook of Deep Trade Agreements*. Ed. by Aaditya Mattoo, Nadia Rocha, and Michele Ruta. 1st ed. Chap. Foreword.

Larch, Mario, Joschka Wanner, Yoto V. Yotov, and Thomas Zylkin (2019), "Currency Unions and Trade: A PPML Re-assessment with High-dimensional Fixed Effects." *Oxford Bulletin of Economics and Statistics* 813, 487–510.

Mattoo, Aaditya, Alen Mulabdic, and Michele Ruta (2022), "Trade Creation and Trade Diversion in Deep Agreements." *Canadian Journal of Economics* 55August, 1598–1637.

Mattoo, Aaditya, Nadia Rocha, and Michele Ruta (2020), *The Evolution of Deep Trade Agreements.*

Rodrik, Dani (2018), "What Do Trade Agreements Really Do?" *Journal of Economic Perspectives* 322, 73–90. URL: https://doi.org/10.1257/jep.32.2.73.

Santos Silva, J.M.C. and Silvana Tenreyro (2006), "The Log of Gravity." *The Review of Economics and Statistics* 884, 641–658. URL: https://doi.org/10.1162/rest.88.4.641.

Teti, Feodora A. (2020), "30 Years of Trade Policy: Evidence from 5.7 Billion Tariffs." *ifo Working Papers* 334.

Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. *de* Vries (2015), "An Illustrated User Guide to the World Input-Output Database: the Case of Global Automotive Production." *Review of International Economics* 233, 575–605. URL: http://doi.wiley.com/10.1111/roie.12178.

Yotov, Yoto V. (2021), "The Variation of Gravity within Countries (or 15 Reasons Why Gravity Should Be Estimated with Domestic Trade Flows)." *Drexel Economics Working Paper Series WP 2021-12.*

— (2022), "Gravity at Sixty: The Bijou of Trade." *Drexel Economics Working Paper Series WP 2022-01.* URL: https://ideas.repec.org/p/ris/drxlwp/2022%7B%5C_%7D001.html.

Yu, Zhihao (2000), "A Model of Substitution of Non-Tariff Barriers for Tariffs." *The Canadian Journal of Economics / Revue canadienne d'Economique* 334, 1069–1090. URL: http://www.jstor.org.emedien.ub.uni-muenchen.de/stable/2667392.

A Appendix

Table A1: The Role of Omitted Variables: Alternative Definition of Depth

	(1) No Tariffs	(2) WITS	(3) New GTD	(4) New GTD	(5) MFN	(6) Glob	(7) Glob-MFN
(A) Baseline (2000-2015), FE: i-s-t, j-s-t, i-j-s							
Deep TAs	0.18**	0.14*	0.07	0.08	0.05	0.06	0.05
	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
Shallow TAs	-0.05	-0.09	-0.15**	-0.14**	-0.15***	-0.08*	-0.09*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Enabl. Clause	0.05	0.04	0.01	0.00	0.03	0.04	0.04
	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
$ln\,(1+\tau)$		-1.17**	-3.09***	-3.06***	-1.65***	-1.27**	-0.98**
		(0.59)	(0.78)	(0.82)	(0.55)	(0.59)	(0.46)
MFN Openness					0.04***		0.01
					(0.01)		(0.01)
N	474,312	360,522	360,522	474,312	474,312	474,312	474,312
(B) Long Sample (Manuf. and Agri., 1989-2015), FE: i-s-t, j-s-t, i-j-s							
Deep TAs	0.23***	0.13*	0.04	0.09	0.02	0.03	0.00
	(0.08)	(0.08)	(0.08)	(0.08)	(0.06)	(0.07)	(0.06)
Shallow TAs	0.00	-0.06	-0.14*	-0.10	-0.10	-0.09	-0.10
	(0.08)	(0.09)	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)
Enabl. Clause	0.01	-0.01	-0.05	-0.05	-0.02	-0.01	-0.00
	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
$ln\,(1+\tau)$		-1.70***	-3.58***	-3.36***	-1.98***	-1.43***	-1.05**
		(0.54)	(0.62)	(0.60)	(0.52)	(0.47)	(0.49)
MFN Openness					0.05***		0.02**
					(0.01)		(0.01)
N	683,290	392,839	392,839	683,290	683,290	683,290	683,290

Note: The information on the depth is provided by DESTA (Dür et al. 2014), definitions of the different types of agreements (enabling clause, shallow TA, and deep TA) can be found in the main text. All columns include importer-sector-time, exporter-sector-time, and importer-exporter-sector fixed effects, columns (5) and (7) include time-varying border controls to account for general trends in globalization. In Panel (A) the data includes three broad sectors (agriculture, manufacturing, and services) for the years 2000 to 2015, Panel (B) has no services-trade but a longer time frame (1989-2015). Standard errors are clustered three-way for importer, exporter, and year. * for $p < 0.1$, ** for $p < 0.05$ and *** for $p < 0.01$.

Table A2: The Role of Omitted Variables: Alternative Clustering of Standard Errors

	(1) No Tariffs	(2) WITS	(3) New GTD	(4) New GTD	(5) MFN	(6) Glob	(7) Glob-MFN
(A) SE Clustered by Country Pairs, FE: i-s-t, j-s-t, i-j-s							
Deep TAs	0.25***	0.21***	0.14***	0.15***	0.07	0.08*	0.06
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Shallow TAs	0.02	-0.01	-0.07	-0.06	-0.04	-0.01	-0.01
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
Enabl. Clause	0.05	0.04	0.00	0.00	0.03	0.04	0.05
	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
$ln\,(1+\tau)$		-1.08***	-2.99***	-2.90***	-1.65***	-1.22**	-0.98*
		(0.27)	(0.58)	(0.57)	(0.56)	(0.56)	(0.58)
MFN Openness					0.04***		0.01*
					(0.01)		(0.01)
N	474,312	360,522	360,522	474,312	474,312	474,312	474,312
(B) SE by Importer-Sector, Exporter-Sector, and Year, FE: i-s-t, j-s-t, i-j-s							
Deep TAs	0.25***	0.21**	0.14	0.15*	0.07	0.08	0.06
	(0.09)	(0.09)	(0.09)	(0.08)	(0.06)	(0.06)	(0.06)
Shallow TAs	0.02	-0.01	-0.07	-0.06	-0.04	-0.01	-0.01
	(0.10)	(0.11)	(0.10)	(0.09)	(0.09)	(0.08)	(0.07)
Enabl. Clause	0.05	0.04	0.00	0.00	0.03	0.04	0.05
	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
$ln\,(1+\tau)$		-1.08*	-2.99***	-2.90***	-1.65***	-1.22*	-0.98*
		(0.59)	(0.78)	(0.82)	(0.58)	(0.69)	(0.58)
MFN Openness					0.04***		0.01
					(0.01)		(0.01)
N	474,312	360,522	360,522	474,312	474,312	474,312	474,312

Note: The information on the depth is provided by Hofmann et al. (2019) (all provisions), definitions of the different types of agreements (enabling clause, shallow TA, and deep TA) can be found in the main text. All columns include importer-sector-time, exporter-sector-time, and importer-exporter-sector fixed effects, columns (5) and (7) include time-varying border controls to account for general trends in globalization. In Panel (A) the data includes three broad sectors (agriculture, manufacturing, and services) for the years 2000 to 2015, Panel (B) has no services-trade but a longer time frame (1989-2015). In Panel (A) the standard errors are clustered by country pairs, in Panel (B) by importer-sector, exporter-sector, and year. * for $p < 0.1$, ** for $p < 0.05$ and *** for $p < 0.01$.

Table A3: NTB effects: Counting Provisions to determine Depth of Trade Agreements

	All Provisions				Leg. Enforcable Prov.				Desta			
	(1) No Tariffs	(2) WITS	(3) New GTD	(4) New GTD	(5) No Tariffs	(6) WITS	(7) New GTD	(8) New GTD	(9) No Tariffs	(10) WITS	(11) New GTD	(12) New GTD
(A) Broad Sectors, (FE: i-s-t, j-s-t, i-j-s)												
Deep TAs	0.12*	0.13*	0.09	0.08	0.12	0.13	0.09	0.08	0.10	0.10*	0.06	0.08
	(0.07)	(0.07)	(0.07)	(0.06)	(0.08)	(0.10)	(0.09)	(0.08)	(0.07)	(0.05)	(0.04)	(0.06)
Shallow TAs	0.02	-0.00	-0.04	-0.01	0.04	0.03	-0.00	0.01	-0.05	-0.05	-0.10**	-0.10**
	(0.08)	(0.08)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Enabl. Clause	0.06	0.03	0.01	0.04	0.06	0.02	0.00	0.04	0.06	0.03	-0.00	0.02
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
$ln(1+\tau)$		-0.12	-1.46**	-1.22*		-0.12	-1.44**	-1.22*		-0.17	-1.62**	-1.76**
		(0.51)	(0.68)	(0.63)		(0.51)	(0.71)	(0.64)		(0.51)	(0.67)	(0.68)
N	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312
(B) Interaction with Services, (FE: i-k-t, j-k-t, i-j-k)												
Deep TAs	0.08	0.06	0.00	0.03	0.08	0.03	-0.04	0.03	0.10	0.10	0.06	0.07
	(0.07)	(0.06)	(0.05)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)
Services × deep	0.24	0.28**	0.33***	0.27*	0.18	0.27**	0.33**	0.22	-0.04	-0.01	0.04	0.02
	(0.16)	(0.13)	(0.13)	(0.15)	(0.17)	(0.13)	(0.13)	(0.16)	(0.15)	(0.14)	(0.15)	(0.16)
Shallow TAs	0.02	0.02	-0.02	-0.01	0.04	0.05	0.01	0.01	-0.05	-0.06	-0.10*	-0.10**
	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
Enabl. Clause	0.06	0.03	-0.00	0.03	0.06	0.03	0.01	0.04	0.06	0.03	-0.00	0.02
	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
$ln(1+\tau)$		-0.18	-1.75***	-1.43**		-0.17	-1.76***	-1.38**		-0.16	-1.66**	-1.78**
		(0.50)	(0.61)	(0.61)		(0.50)	(0.65)	(0.62)		(0.50)	(0.65)	(0.70)
N	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312
(C) Interaction with High-Income Countries, (FE: i-k-t, j-k-t, i-j-k)												
Deep TAs	0.12	0.17*	0.12	0.07	0.04	0.41***	0.38***	-0.00	0.09	0.13	0.10	0.05
	(0.10)	(0.09)	(0.09)	(0.09)	(0.07)	(0.12)	(0.11)	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)
High Income × deep	-0.00	-0.04	-0.03	0.01	0.09	-0.28*	-0.30*	0.08	0.00	-0.03	-0.04	0.02
	(0.10)	(0.10)	(0.10)	(0.10)	(0.08)	(0.16)	(0.16)	(0.08)	(0.11)	(0.09)	(0.10)	(0.11)
Shallow TAs	0.02	-0.00	-0.04	-0.01	0.04	0.03	-0.00	0.01	-0.05	-0.05	-0.10*	-0.10**
	(0.08)	(0.08)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Enabl. Clause	0.06	0.03	0.01	0.04	0.06	0.02	0.00	0.04	0.06	0.03	0.00	0.02
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
$ln(1+\tau)$		-0.12	-1.46**	-1.22*		-0.12	-1.45**	-1.22*		-0.16	-1.62**	-1.76**
		(0.51)	(0.68)	(0.63)		(0.51)	(0.71)	(0.64)		(0.51)	(0.68)	(0.69)
N	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312
(D) Interaction with G7 Countries, (FE: i-k-t, j-k-t, i-j-k)												
Deep TAs	0.11	0.13	0.09	0.08	-0.03	-0.01	-0.07	-0.08	0.03	0.07	0.03	0.01
	(0.09)	(0.09)	(0.08)	(0.08)	(0.06)	(0.05)	(0.04)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
G7 × deep	0.01	0.01	0.00	0.01	0.16**	0.15*	0.16**	0.17**	0.10	0.04	0.05	0.09
	(0.10)	(0.11)	(0.11)	(0.11)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.09)	(0.09)	(0.09)
Shallow TAs	0.02	-0.00	-0.04	-0.01	0.04	0.03	-0.00	0.01	-0.05	-0.05	-0.10	-0.10
	(0.08)	(0.08)	(0.08)	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Enabl. Clause	0.06	0.03	0.01	0.04	0.05	0.02	0.00	0.04	0.05	0.02	-0.00	0.01
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
$ln(1+\tau)$		-0.12	-1.46**	-1.22*		-0.12	-1.45**	-1.23*		-0.17	-1.61**	-1.73**
		(0.51)	(0.69)	(0.63)		(0.51)	(0.70)	(0.63)		(0.52)	(0.67)	(0.69)
N	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312	474,312	340,221	340,221	474,312
(E) Interaction with EU Countries, (FE: i-s-t, j-s-t, i-j-s)												
Deep TAs	0.13*	0.14**	0.10	0.09	0.13	0.14*	0.10	0.09	0.13*	0.12*	0.10	0.10
	(0.07)	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.07)	(0.06)	(0.07)
EU × deep	-0.02	-0.01	-0.01	-0.02	-0.02	-0.02	-0.01	-0.01	-0.13	-0.12	-0.10	-0.12
	(0.09)	(0.13)	(0.13)	(0.09)	(0.08)	(0.14)	(0.14)	(0.08)	(0.11)	(0.12)	(0.12)	(0.11)
Shallow TAs	0.02	0.01	-0.03	-0.01	0.04	0.03	0.00	0.01	-0.07	-0.06	-0.12**	-0.13**
	(0.08)	(0.08)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)
Enabl. Clause	0.06	0.06	0.04	0.04	0.05	0.05	0.03	0.04	0.06	0.06	0.02	0.02
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
$ln(1+\tau)$		-0.12	-1.42**	-1.22*		-0.14	-1.41**	-1.22*		-0.20	-1.99***	-1.78***
		(0.49)	(0.68)	(0.63)		(0.49)	(0.69)	(0.64)		(0.50)	(0.70)	(0.68)
N	474,312	360,522	360,522	474,312	474,312	360,522	360,522	474,312	474,312	360,522	360,522	474,312
(F) Interaction with United States, (FE: i-s-t, j-s-t, i-j-s)												
Deep TAs	0.15*	0.17*	0.13	0.11	0.14	0.17	0.13	0.11	0.11	0.10	0.08	0.09
	(0.08)	(0.09)	(0.09)	(0.08)	(0.10)	(0.13)	(0.13)	(0.10)	(0.08)	(0.08)	(0.07)	(0.07)
US × deep	-0.10	-0.11	-0.12	-0.10	-0.09	-0.10	-0.10	-0.09	-0.05	-0.03	-0.03	-0.05
	(0.09)	(0.10)	(0.10)	(0.09)	(0.11)	(0.14)	(0.14)	(0.11)	(0.09)	(0.10)	(0.09)	(0.09)
Shallow TAs	0.03	0.01	-0.02	-0.00	0.05	0.04	0.01	0.02	-0.04	-0.04	-0.10*	-0.10**
	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)
Enabl. Clause	0.07	0.06	0.04	0.05	0.06	0.05	0.04	0.04	0.06	0.06	0.02	0.02
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
$ln(1+\tau)$		-0.12	-1.42**	-1.22*		-0.13	-1.42**	-1.22*		-0.18	-1.98***	-1.76**
		(0.49)	(0.69)	(0.64)		(0.49)	(0.71)	(0.64)		(0.49)	(0.71)	(0.69)
N	474,312	360,522	360,522	474,312	474,312	360,522	360,522	474,312	474,312	360,522	360,522	474,312

Note: Column (1) to (8) use information on the depth provided by Hofmann et al. (2019), column (9) to (12) by Dür et al. (2014), respectively. Definitions of the different types of agreements (enabling clause, shallow TA, and deep TA) can be found in the main text. The sample includes three broad sectors (agriculture, manufacturing, and services) for the years 2000 to 2015. In Panel (B) to Panel (E) we interact deep TA with a services trade dummy, an indicator variable if one of the signatories of the TA is a high-income country, a G7 country, a member of the EU or the United States. All columns include importer-broad sector-time, exporter-broad sector-time, importer-exporter-broad sector fixed effects, and time-varying border controls to account for general trends in globalization. Standard errors are clustered three-way for importer, exporter, and year. * for $p < 0.1$, ** for $p < 0.05$ and *** for $p < 0.01$.

Table A4: Estimated NTB-Effects: Summary Statistics, only Significant Results (95% and 99%-Level)

	90%-Level					99%-Level				
	(1) Nr.	(2) Mean	(3) pc(50)	(4) pc(25)	(5) pc(75)	(6) Nr.	(7) Mean	(8) pc(50)	(9) pc(25)	(10) pc(75)
All	99	0.278	0.348	-0.218	0.572	44	0.477	0.516	0.285	0.741
Positive	70	0.570	0.463	0.320	0.709	36	0.694	0.594	0.420	0.814
Negative	29	-0.428	-0.338	-0.515	-0.264	8	-0.502	-0.392	-0.561	-0.308

Note: The table summarizes the results for the significant (95% and 99% confidence-level) RTA-specific NTB-effects when estimating Equation 6.