Product Differentiation, Interdependence, and the Formation of PTAs

KIRAN AYUB* PERI SILVA**

October 24, 2023

Abstract

We consider the role played by trade in differentiated inputs in the country-pair decision to form a PTA in goods and in their decision to expand it to trade in services with varying degrees of coverage, which transforms a preferential agreement into an Economic Integration Area (EIA). Our baseline model is very successful in predicting the formation of preferential agreements. Our model correctly predicts 84 percent of the country pairs with PTAs in our dataset and can successfully predict the 83 percent of the country pairs that do not form a PTA. Moreover, our model predicts 78 percent of the observations involving country pairs belonging to an EIA when a PTA exists.

We are grateful for the excellent comments received by two referees and the associate editor, Frank Stähler. We also appreciate the seminar participants at the Fall 2020 Missouri Valley Economics Association, the Spring 2021 Midwest Economic Association, and the Fall 2022 Southern Economic Association annual meetings for their helpful comments and suggestions.

* Visiting Assistant Professor of Economics, Lewis and Clark College, email: kaayub@ksu.edu

** Professor and Head, Economics Department, Kansas State University, fellow at the GEP and Ld'A, email: pdasilva@ksu.edu

"This surge in the importance of input trade seems to have been accompanied by a parallel increase in the share of differentiated products in the total volume of world trade, and an associated fall in the share of homogeneous goods (as measured by the share of goods traded on organized exchanges). Although part of this trend is explained by the changing nature of final good trade, a significant portion of it reflects a disproportionate increase in world trade in differentiated intermediate inputs"

Antràs and Staiger (2012)

1 Introduction

The post-WWII era has witnessed an increasing degree of international economic integration based on a rules-based international trade system.[1] A central piece of this effort has been the General Agreement on Tariffs and Trade (GATT), which has sponsored eight successful rounds of multilateral negotiations, reducing the tariffs on manufactured goods applied by developed countries by more than eighty percent.[2] Despite the success of multilateral negotiations, there has been a surge in the formation of preferential trade arrangements (PTAs) since the establishment of the World Trade Organization (WTO) in 1995.[3] There were 47 active PTAs in 1995, which increased to 72 in 1999 and expanded exponentially to 361 by September of 2023.[4] In this respect, the advances in international trade liberalization in the last twenty-five years are mostly due to the formation of PTAs.[5] At the same time, Antràs and Staiger (2012) point out that most of the growth in trade flows took place in terms of differentiated intermediate inputs. In this paper, we consider the role played by trade in differentiated inputs in the country-pair decision to form a PTA

[1]The Post-WWII rules-based international trade system was established in part to avoid the "trade wars" and, more generally, the beggar-thy-neighbor policies that characterized the 1930s. Douglas Irwin has recently discussed the fragmentation of the international economy during the 1930s, and the main conclusions can be found at `https://www.piie.com/experts/peterson-perspectives/trade-talks-episode-31-trade-wars-and-smoot-hawley-tariff-what-really`.

[2]Bagwell et al. (2016) survey the GATT/WTO literature. In particular, their tables 1 and 2 provide information on the main tariff-reducing objectives of the various rounds of multilateral negotiations and show the substantial decrease in the MFN tariffs applied by developed countries.

[3]In addition, the ninth round of negotiations initiated in Doha (Qatar) has failed to liberalize multilateral trade further.

[4]The reader can check this information at `http://rtais.wto.org/UI/charts.aspx.` These figures correspond to physical agreements where double counting due to goods and service coverage and acceding processes is eliminated.

[5]Nicita et al. (2018) argue that the combination of multilateral and preferential tariff reductions has led the average world exporter to face an average tariff of 2.6 percent. See details at `https://voxeu.org/article/trade-war-will-increase-average-tariffs-32-percentage-points.`

in goods and in their decision to expand it to trade in services with varying degrees of coverage, which transforms a preferential agreement into an Economic Integration Area (EIA).[6] A key component of our analysis is that we consider that country-pair decisions, and PTA formation, in general, are interdependent, creating conditions for the formation of a preferential agreement influence the decision to form other preferential agreements.

Importantly, the literature has identified multiple concerns regarding the expansion of PTAs.[7] Bagwell and Staiger (1999) explain that multilateral agreements sponsored by the WTO and its predecessor agreement, the General Agreement on Tariffs and Trade (GATT), are based on two simple rules. The most-favored-nation (MFN) rule requires that negotiators agree on non-discriminatory tariff commitments (tariff bindings), while the rule of reciprocity requires members to reciprocate increases in market access via decreases in their bound tariffs. Importantly, Bagwell and Staiger (1999) show that these two rules promote an efficient policy outcome by eliminating the terms-of-trade tariff incentive, which is inherently present in otherwise unilateral tariff settings. Instead, PTAs deviate from the MFN rule by applying preferential tariffs, thereby possibly undermining the efficiency of the international trade system.

However, notice that multilateral negotiations sponsored by the WTO tend to focus mostly on trade in goods. In contrast, many PTAs tend to be deeper in policy by covering rules that liberalize and address trade in goods and services and potentially covering additional areas outside the WTO agreement, such as competition policy, investments, labor market regulations, etc. Antràs and Staiger (2012) argue that, in an environment where intermediate differentiated inputs are important, sellers usually customize inputs for intended buyers, creating lock-in effects that are determined by bilateral negotiations rather than by global discipline of market clearing considerations. Likewise, Blanchard (2010) explains that the presence of cross-border ownership essentially creates externalities requiring trade policy negotiations to also consider cross-border ownership rather than only market access

[6] This definition follows the WTO. In particular, note that all country pairs with a preferential agreement covering some areas in services also have an agreement that involves goods, but, in many cases, preferential agreements only involve goods.

[7] Researchers have investigated how preferential trade liberalization has affected multilateral tariffs and how it affects trade between member and non-member countries. Estevadeordal et al. (2008) and Limão (2006) consider whether lowering preferential tariffs tends to affect the MFN applied by Latin American economies and the multilateral tariff cuts negotiated by the U.S., respectively. Instead, Conconi et al. (2018) study the protectionist effects of the rules of origin used under the North American Free Trade Area (NAFTA) on the trade of intermediate products between members and non-members of that agreement.

concessions that characterize negotiations under multilateral agreements. The bottom line is that if countries are trading differentiated inputs primarily with prospective preferential partners and are engaged in foreign direct investments[8] (FDI) related to the production of specific parts to jointly produce other goods, then the non-discriminatory nature of MFN tariffs in goods does not necessarily control for the externalities that exist in setting trade policy. Addressing these specific bilateral externalities then creates an opportunity for PTAs to enhance the efficiency of the international trade system.[9]

It is important to note that Ruta (2017) and Hofmann et al. (2019) show that the only common characteristic across preferential agreements is that they cover preferential tariffs on goods, but their coverage policy varies significantly, and they have become deeper (increased in scope) over time. As such, we also follow the WTO and control for the fact that many preferential agreements also cover preferences in service (i.e., become an EIA), formed sequentially (after the PTA is formed) or jointly. An example can put some of these elements into perspective. Consider the case of a Japanese automaker that outsources differentiated parts of a vehicle to Malaysia and Thailand. The production of the vehicle is now reliant on a specific regional value chain (WTO and IDE-JETRO (2011)).[10] Consequently, it increases the interdependent interests of Japan and Malaysia, Japan and Thailand, and Malaysia and Thailand to go beyond the WTO. They can form a PTA[11] to promote the trade in differentiated inputs, with stronger rules to mitigate trade policy uncertainty[12] and possibly extend it to become an EIA where investment protection and

[8]Blanchard (2010) shows that the presence of FDI implies that the MFN and reciprocity rules do not completely control the negative effects of policy externalities across countries.

[9]Bagwell et al. (2016) label these ideas as the Offshoring Theory of PTAs. More specifically, Antràs and Staiger (2012) argue that "...that in the presence of offshoring it is necessary to achieve "deep integration"—extending beyond a narrow market-access focus in ways that we formalize below—in order to arrive at internationally efficient policies."

[10]The WTO and IDE-JETRO (2011. pp. 15) use an example of a Japanese automaker outsourcing key components from four ASEAN countries (Malaysia, Philipines, Indonesia, and Thailand), taking advantage of the ASEAN Free Trade Area (AFTA). More information regarding the example can be found in this publication by WTO and IDE-JETRO `https://www.wto.org/english/res_e/booksp_e/stat_tradepat_globvalchains_e.pdf.`

[11]In this case, the Common Effective Preferential Tariff (AFTA-CEPT) applied by ASEAN members represents a cooperative arrangement among ASEAN member states to reduce intraregional tariffs and remove non-tariff barriers.

[12]For instance, as part of the negotiations that led to the approval of the USMCA, Mexico received guarantees that its access to the U.S. automotive market would not change even if that country decided to impose additional tariffs on imports of cars from other WTO members. This strategy then implies that PTA formation can mitigate trade policy uncertainty. See side letter sent by the United States Trade Representative's office to its Mexican counterpart at "`https://ustr.gov/sites/default/_les/_les/agreements/FTA/USMCA/Text/MXUS` Side Letter on 232.pdf."

the appropriate level of patents are addressed to diminish the risks and policy uncertainty to all interdependent parties.[13] In line with this argument, Johnson and Noguera (2017) show that forming PTAs has disproportionally reduced the ratio between domestic value-added and gross exports since these member economies have been fragmenting the production processes across themselves, forming and enhancing regional and global value chains (GVCs).

We study the decision to form preferential agreements by compiling a dataset with bilateral trade flow information at the 4-digit of the Standard International Trade Classification (SITC) system from the United Nations COMTRADE system, with information on the formation of preferential trade agreements in goods (PTA), as well as agreements involving goods and services (EIA) from Hofmann et al. (2017). This dataset also provides information on policy areas covered beyond goods and services, such as property rights and investments. Our dataset spans from 1980 to 2015 and covers several gravity-related variables measured at the country-pair level with information obtained from the CEPII. Numerous features of the dataset indicate the importance of the topic and the particular elements used in our analysis. It suggests that trade in inputs has grown by almost forty-six percent between 1980 and 2015, reaching a peak of almost fifty-two percent of total trade in 2015. Our data also indicate that about forty percent of the PTAs in our sample only involve trade in goods, while some PTAs in our data first liberalize only in goods and later on also added liberalization in services.[14] As examples, we notice that the original agreement creating the European Union (E.U.) dates back to 1957 with the Treaty of Rome, but the E.U. did not include an agreement in services until 1994. The same is true with the European Free Trade Area (EFTA)-Israel agreement, which became an FTA (only goods) in 1995 but later evolved to add service coverage in 2000.

This paper is part of the literature that focuses on the main drivers of PTA formation, and our empirical strategy builds on its fundamental approaches. Baier and Bergstrand (2004) use a binary econometric specification (Probit) to identify country-pair economic

[13] Prusa and Teh (2010) explain that the formation of PTAs either rules out or significantly constrains the use of temporary trade barriers (Anti-dumping duties, countervailing duties, and safeguard measures) across member countries relative to the WTO negotiations. These agreements then decrease uncertainty relative to applying these measures as well. Tabakis and Zanardi (2019) show that PTA formation also leads to fewer applications of these tools against non-member countries by constraining these temporary trade measures of protection between members.

[14] Information compiled from the WTO files provided at http://rtais.wto.org/UI/PublicAllRTAListAccession.aspx.

characteristics that affect the probability that a country-pair agrees to join or form a PTA. Egger and Larch (2008) extend this binary choice approach by highlighting the importance of interdependence among PTAs in the probability that a country joins an existing PTA or forms a new one. Moreover, Baldwin and Jaimovich (2012) offer a theory-driven measure of PTA interdependence where the country-pair decisions depend on their direct and relative trade exposure to partner countries involved with other PTAs. Similarly, Chen and Joshi (2010) investigate the effects of a particularly important type of interdependence theoretically and empirically on PTA formation. In their case, some prospective PTA members would serve as the "hub" (i.e., a country that participates in many overlapping PTAs), while others would serve as the "spoke" with limited participation in other PTAs. In line with the literature, the role played by PTA interdependence is a central part of our analysis.[15]

Our benchmark econometric model then uses a Probit specification to explain the formation of all comprehensive agreements[16] formed until 2015 while we control for different measures of the share of bilateral trade in inputs. Moreover, we indicate above that some country- pairs that form a PTA may extend it to trade-in services with various degrees of policy coverage. We then follow the approach used by Facchini et al. (2021) and expand our empirical model to study the decision related to forming an EIA with varying degrees of policy coverage using a Probit model with sample selection. In this case, we also control for the interdependence among EIAs since the degree of interdependence among them may differ from the effects of interdependence among agreements that focus solely on goods. This approach is important since EIAs have significant differences in policy coverage, as we outline in more detail in the section dedicated to the data description.

Our empirical results support the idea that bilateral trade in differentiated inputs (rather than inputs in general) across prospective member countries is an important driver of the formation of PTAs.[17] More specifically, we find that the best model to explain the forma-

[15]Of course, there are other essential papers in this literature. Notice that most articles tend to focus on the economic drivers of PTA formation with some notable exceptions. Liu and Ornelas (2014) focus on how the risk of interruption in a democratic regime may lead to an increase in the probability of PTA formation, which, in turn, tends to make democratic regimes last longer. On the other hand, Facchini et al. (2021) consider the political economy role of income inequality and trade imbalances on the decision to form a PTA and how geographic specialization of production affects the decision about PTA type (FTA or CU).

[16]In this paper, we focus on free trade areas (FTAs) and customs unions (CUs) since these PTAs usually aim at implementing duty-free trade among members. As such, we do not use partial scope agreements as a PTA in testing the robustness of our results using data provided by Egger and Larch (2008) and Baier et al. (2014).

[17]Egger et al. (2008) study how the endogenous formation of PTAs affects intra-industry trade. Instead, our focus is the effects of the bilateral share of trade on differentiated and differentiated-manufactured products, alongside

tion of PTAs from a statistical point of view uses the share of bilateral trade in the form of differentiated-manufactured inputs alongside a control for their interdependence. This result aligns with the findings discussed by Johnson and Noguera (2017) that the manufacturing industries have seen the most significant drops in the ratio between value-added and gross exports. Likewise, we find that controlling for this specific share of bilateral trade while controlling for the interdependence of EIAs yields the best statistical results in explaining the decision to form EIAs for country-pairs that have decided to join an existing PTA or create a new one. In this regard, we also find that the share of bilateral trade in differentiated-manufactured inputs is particularly powerful in explaining the formation of EIAs that cover all major WTO-related policy areas (GATS, TRIMs, and TRIPs) plus additional policy items related to investment and intellectual property rights.

Importantly, our baseline model is very successful in predicting the formation of preferential agreements. It correctly predicts 84 percent of the country pairs with PTAs in our dataset and can successfully predict about 83 percent of the country pairs that do not form a PTA. Moreover, our model predicts 78 percent of the observations involving country pairs belonging to an EIA when a PTA exists. We consider numerous robustness tests of these results, including alternatively defining the presence of PTAs using the datasets by Egger and Larch (2008) and Baier et al. (2014) and by measuring the definition of interdependence using the theory-driven measure explored in Baldwin and Jaimovich (2012). Moreover, we consider the results under a linear probability model with a comprehensive set of fixed effects, testing the model for different sub-samples of years, using different cross-sections of the data rather than the panel structure across years used in the baseline results, among others. These robustness tests confirm the baseline empirical model results, suggesting that forming PTAs may enhance the efficiency of the multilateral trade system.

We notice the strong relationship between our paper and the literature focusing on preferential agreements in services. Cole and Guillin (2015) are the first to note the potential interplay between preferential agreements in goods and services since the common characteristic across preferential agreements is the presence of preferential tariffs in goods. They find that the more effective and stable the political system of prospective members, the higher the likelihood that an agreement on goods and services is formed. Egger and Shin-

interdependence, on PTA and EIA formation. Thus, our papers study different causalities and rely on different dependent and explanatory variables while we consider additional policy coverage across preferential agreements.

gal (2021) investigate the role played by the country-level services restrictiveness index (STRI) in the formation of preferential agreements in services. Importantly, they control for the endogeneity of STRI in determining preferential agreements. They find that the greater the asymmetry of STRI in a country pair, the less likely a preferential agreement in services is formed.[18]

Our paper explores the role of bilateral trade in differentiated inputs on PTA and EIA formation. Thus, it relates to the literature that discusses the relationship between GVCs and PTAs since trade in inputs is also a measure of the GVC intensity.[19] Laget et al. (2020) find that GVC linkages in the form of the share of bilateral trade in inputs and the foreign value-added content in gross exports are important determinants of bilateral trade flows. Likewise, Mulabdic et al. (2017) conclude that the stronger the GVC ties, the greater the effects of BREXIT on E.U.-U.K. trade flows. Instead, Fontagné and Santoni (2021) investigate the effects of GVCs on PTA formation. In their paper, GVCs are captured by the importance of indirect value-added exports between prospective members relative to their economy sizes. They find that the greater the importance of indirect value-added exports, the more likely PTAs are formed.

Our paper investigates the role of differentiated-manufactured inputs, highlighted in Antràs and Staiger (2012), and interdependence on PTA and EIA formation. Thus, it contributes to the literature by emphasizing this important contributor to linkage across countries (differentiated inputs), properly measuring and quantifying the direct effect of the inter-related web of PTAs and EIA on their formation while studying agreements with different policy coverage depths.[20]

The paper is further divided as follows. Section 2 discusses the data and the construction of our variables. Instead, section 3 focuses on our econometric methodology. Section

[18] Egger and Shingal (2017) explore the role of the intensity of bilateral trade on goods and services in the formation of preferential agreements in services. They also consider the role of these variables in determining whether a preferential agreement in goods will be formed jointly or sequentially, given the decision to form an agreement in services. They find that the size of bilateral merchandising trade and bilateral trade in services are important determinants in forming preferential agreements.

[19] Ruta (2017) provides an excellent review of the dataset discussed in Hofmann et al. (2017) and the role of GVC in promoting bilateral trade flows.

[20] Tsirekidze (2021) incorporates trade in inputs to the "competing exporters" model outlined in Bagwell and Staiger (1999). He shows that forming FTAs creates a free riding incentive for non-members countries due to lower tariffs on final goods while continuing their free-trade reliance on the non-member country for inputs. This result then makes free trade less likely as the importance of GVC (trade on inputs) grows. The paper then shows that sufficiently restrictive RoOs can be used to diminish this free-rider incentive, by shifting input demand towards FTA members. Thus, RoOs can be used as a policy device to increase the likelihood of free trade in equilibrium.

4 then describes our main empirical results, while in section 5, we present several robustness results. Section 6 concludes.

2 Data and Variable construction

We investigate the role of trade in intermediate-differentiated products and interdependence in forming PTA and EIAs by employing a large panel dataset with country-pair information every five years covering 189 countries from 1980 to 2015. The decision to use data every five years follows in the footsteps of Baier and Bergstrand (2007), Egger and Larch (2008), Subramanian and Wei (2007), Facchini et al. (2021), among others, which observed that the presence of PTAs does not change much at a yearly level and preferential tariffs are phased in with substantial time lags. Notice that these characteristics are also shared by other trade agreement commitments, such as the decision to join the WTO.[21] Moreover, Fontagné and Santoni (2021) argue that using five-year lagged-explanatory variables is important to control for reverse causality since the median PTA negotiation time is 4.5 years. Our dataset consists of country-pair information about the presence of PTAs and EIAs, measures of these agreements' interdependence, measures of the degree of bilateral trade in differentiated (manufactured and non-manufactured) intermediate goods, and control variables used in other studies. We describe each of these sets of variables below.

2.1 Data on formation of PTAs and EIAs

Our baseline results define the presence or absence of PTAs and EIAs across country pairs based on information provided by Hofmann et al. (2017), which was discussed in Hofmann et al. (2019). In this case, their dataset defines the presence of PTAs and EIAs according to a de facto definition. For this reason, we test the robustness of our results using alternatively the PTA information provided by Egger and Larch (2008), which uses a de jure definition in line with the WTO and the information in Baier et al. (2014), which is also based on de facto criteria. However, Baier et al. (2014) do not define the presence of EIAs. [22]

[21] Likewise, Chen and Joshi (2010) use three-year time intervals for the same reason.

[22] Mario Larch's database defines that an EIA is present if the preferential agreement involves liberalization of service products included in the General Agreement on Trade in Services (GATS). Details about his database can be found at `https://www.ewf.uni-bayreuth.de/en/research/RTA-data/index.html`. Instead, Baier et al.

Hofmann et al. (2017) define preferential agreements as FTAs, and CUs and provide information on policy coverage for 50 additional policy areas (i.e., beyond preferential trade in goods). In their dataset, all preferential agreements include preferential treatment in manufactured goods, and very few exclude preferential trade in agricultural products.[23] Thus, the only universal policy shared by all preferential agreements is preferential access to goods. The coverage of other policies varies significantly across agreements. We disregard partial agreements as a type of PTA since, by definition, these agreements may liberalize trade in a few products and clearly do not imply that members are pursuing internal duty-free trade.[24]

Thus, in our dataset, any country pair that belongs to a CU or an FTA is considered a PTA. Of those countries that belong to a PTA, we then look at those country pairs that also share a preferential agreement in service, i.e., belong to an EIA.[25] We then construct two dependent variables PTA_{abt} and EIA_{abt}, to denote the presence of a PTA and an EIA. The former takes a value of one if at time t a PTA in goods is present between countries a and b. The latter variable is defined for those country-pairs with a PTA in place. It takes a value of one if at time t there is a PTA in goods that also includes services between countries a and b, and zero if there is a PTA in place without including services. Notice below that we outline further variations of the variable EIA_{abt} with the intent of testing the effects of trade in differentiated-intermediate products and interdependence on the policy scope covered by EIAs.

[TABLE 1 here]

We layout the descriptive statistics for the variables used in our baseline results in Table 1. Column 1 offers information on the average and standard deviation for each

(2014) define the presence of unilateral trade agreements (e.g., GSP, African Growth and Opportunity Act, etc.), partial PTAs, FTAs, CUs, Common Markets, and Economic Unions. This dataset includes information for 195 country pairs from 1950 to 2012 and uses these six categories to provide information on the depth level of preferential liberalization between a country pair. Their dataset is available at https://sites.nd.edu/jeffrey-bergstrand/database-on-economic-integration-agreements/.

[23] Hofmann et al. (2019) confirms this information.

[24] Ornelas and Tovar (2022) investigate the decision to set preferential trade margins across FTAs and CUs. They explain that preferential tariffs are not necessarily set to zero because of PTAs created under the Enabling Clause, for instance. Their dataset focuses on several Latin American PTAs during the 1990s, where positive internal tariffs deviating from duty-free access across members were widespread.

[25] According to the WTO's Regional Trade Agreement Database, there are no countries pairs where an EIA exists for a country-pair, while it does not also share a PTA. In this case, only the European Economic Association (EEA) corresponds to a standalone EIA. However, notice that its members (Switzerland and the EU) have a separately negotiated FTA.

variable in the entire sample, whereas column 2 shows equivalent information focusing only on country-pairs belonging to a PTA. Moreover, Column 3 offers similar information on country-pairs that form an agreement in goods and services (i.e., belonging to an EIA). Column 1 of Table 1 indicates that we use a total of 69824 country-pair observations in our baseline results. The information shown in column 2 suggests that about 12 percent of the total sample (8364 out of 69824) is represented by country- pairs that have a PTA in place.[26] Column 3 of Table 1 shows that about 60 percent of the country pairs with a PTA in place also liberalize preferentially in services (5013 out of 8364), i.e. form an EIA.

As discussed earlier, PTAs have become very numerous in the last three decades, but they can differ across multiple lines. On the one hand, they can differ concerning their tariff policy. Facchini et al. (2013) and Facchini et al. (2021) explain that the formation of FTAs is more common than CUs since many countries have significantly different productive structures. In this case, the presence of common external tariffs under CUs leads representative democracies to delegate power to more protectionist representatives.[27] They show that the delegation process can then lead FTAs to become politically more palatable than CUs. Baier et al. (2014) distinguish agreements according to the degree of tariff liberalization (partial agreements versus FTAs/CUs) and to the degree of coordination in external tariff policy (FTAs versus CUs). They find that full-blown PTAs in the form of FTAs and CUs lead to more trade among members than partial agreements, while CUs tend to promote more trade than FTAs.

In this paper, we also investigate the degree of PTA depth by considering variations in the definition of EIAs that might be related to economic integration via trade in intermediate goods. Major differences across preferential arrangements exist beyond their policy towards trade in goods, which is very understudied. Horn et al. (2010) and WTO (2011) discuss that an increasing number of preferential arrangements cover not only goods and services but also other policy areas, such as competition policy, intellectual property rights protection, investment, money laundering, mining, etc. Some of these policy areas maybe vital to relate the importance of trade in intermediate goods [28], the interdependence across

[26]For comparison, the PTA database used in Baier et al. (2014) suggests that 7.5 percent of the total sample corresponds to country-pairs that belong to the same PTA.
[27]Lake and Yildiz (2016) show that higher trade costs between members (distance) can turn CUs welfare-inferior relative to FTAs.
[28]Mattoo et al. (2022) show that deep agreements lead to more trade creation and less trade diversion than shallow agreements.

preferential agreements, and the decision to form an EIA and its policy coverage (i.e., the EIA's depth). For this purpose, we use the database provided by Hofmann et al. (2017), and also discussed in Hofmann et al. (2019),[29] to analyze the decision to form and the policy coverage of EIAs in detail.[30]

Horn et al. (2010) divide the provisions covered across preferential agreements into WTO+ (WTPplus) and WTO-X (WTOextra) policy areas. The former group corresponds to policy areas such as customs regulations, the General Agreement on Trade in Services (GATS), the Agreement on Trade-Related Investment Measures (TRIMS), the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), antidumping, technical barriers to trade (TBT), among others, and fall under the current mandate of the WTO. Instead, the latter group represents provisions outside the WTO mandate and includes policy areas related to investments, environmental laws, nuclear safety, etc. In addition to these specific policy areas, Hofmann et al. (2017) also provide information on whether these policies are legally enforceable. These correspond to provisions that are expressed in clear legal language such that it can be appealed in a dispute settlement proceeding. The idea is that vague legal language used to detail provisions may not be legally enforceable WTO (2011).

Since the assumption is that increased trade in differentiated-intermediate goods increases the likelihood that prospective members form a PTA, we also consider its role in the formation of an EIA and in determining its policy coverage. As discussed in the introduction, trade in differentiated-intermediate goods exposes producers to greater uncertainty by delegating parts of the production process to other countries (i.e., a hold-up problem), be it through FDI or from foreign outsourcing [31]. As such, we focus on the following five areas beyond trade in goods that relate to FDI and intellectual property rights. We refine the definition of an EIA by using three WTO+ areas (GATS, TRIMS, and TRIPS), and two WTO-X areas (intellectual property rights (IPR) and investments).[32][33]

[29] This dataset is available at https://datatopics.worldbank.org/dta/table.html on the world bank site.

[30] The dataset covers 52 policy areas and their legal enforceability in 279 PTAs among 189 countries up till 2015.

[31] Osnago et al. (2019) find that the depth of trade agreements is correlated with vertical foreign direct investment. The relationship being positive when the provisions improve the contractibility of inputs provided by suppliers, and negative when these provisions relate to intellectual property rights and investment protection.

[32] TRIMS covers Trade-Related Investment Measures that affect trade in goods. Hofmann et al. (2017) indicate that "These are provisions on requirements for local content and export performance on FDI and applies only to measures that affect trade in goods". Moreover, TRIPS covers provisions that reinforce WTO commitments, including harmonization of standards and provisions related to intellectual property rights.

[33] Investment under WTO-X relates to provisions on promotion, protection, and liberalization of investment mea-

To control for these areas, we construct two additional variants of our EIA (binary) variable: $EIA\_WTO-deep_{abt}$, and $EIA\_WTO\_leg_{abt}$. Our variable $EIA\_WTO-deep_{abt}$ takes a value of one if, at time t, a PTA in goods and services (EIA) is present between countries a and b and the agreement covers simultaneously all three major WTO+ areas (GATS, TRIMS, and TRIPS) and the two fundamentals WTO-X areas (IPR and investments), and zero otherwise. Furthermore, the variable $EIA\_WTO\_leg_{abt}$ takes on a value of 1 if, at time, t a PTA in goods and services (EIA) is in place between countries a and b and the provisions for GATS, TRIMS, TRIPS, IPR, and investment are legally enforceable. Notice that these variables represent subsets of country pairs with an EIA in place where the agreements display an increasing policy coverage. As such, we find that only 6.14 and 5.78 percent of the country-pairs with a PTA present a value of 1 for these variables, respectively.

2.2 Bilateral Trade Ratios

To determine the formation of a PTA, we develop some important explanatory variables related to the share of bilateral trade on inputs and the interdependence among PTAs and EIA. We built measures of the bilateral trade share on inputs using the information on bilateral trade flows at the industry level from the UN COMTRADE database. More specifically, we downloaded information on import values at the 4-digits of the SITC (revision 2) from UNCOMTRADE for the years 1980 to 2015. We then merged this dataset with James Rauch's Classification[34] so that we could identify differentiated products, a product characteristic key to our analysis.[35] This step is important as preferential trade can be efficiency-enhancing if products traded bilaterally are differentiated from products found in the world marketplace. Moreover, our bilateral trade measures require us to distinguish products between final and intermediate goods. Thus, we use the information on Broad Economic Categories (BEC) codes from the United Nations (U.N.) Trade statistics to identify goods for final and intermediate use. Finally, we also wish to identify trade in manufactured products since Johnson and Noguera (2017) point out that fragmentation of the production process has particularly accelerated for manufacturing industries. We

sures or common investment policies, whereas IPRs includes any provision beyond that covered under TRIPS.

[34] This dataset is available at https://econweb.ucsd.edu/~jrauch/rauch_classification.html

[35] His commodity classification distinguishes among homogeneous ("w"), reference pricing ("r"), and differentiated goods ("n"). We use his code "n" to identify differentiated products.

define a manufactured product if its 4-digit SITC code is above 5000 and below 9000.

We define our trade ratio in inputs by dividing the trade between countries a and b in intermediate products by their total bilateral trade. We create six different variants of this basic ratio by considering different definitions of bilateral intermediate trade, i.e., the numerator of this trade ratio. First, we consider the definition of trade in inputs by excluding (including) capital goods from the definition of inputs, which is denoted by subscript "1" ("2") in the following formulas.[36] Notice that more information can be found in Table A of the Appendix:

$$RatioINP_{abt}\_1 = \frac{bilateral\,inputs_{abt}\_1}{Total\,bilateral_{abt}} \qquad RatioINP_{abt}\_2 = \frac{bilateral\,inputs_{abt}\_2}{Total\,bilateral_{abt}} \tag{1}$$

where we notice that $RatioINP_{abt}\_1$ is lower than $RatioINP_{abt}\_2$ since the latter measure includes capital goods as part of the intermediate products group, while the former measure does not.

The essential part of our analysis focuses on the content of trade in differentiated inputs. Our definition of the bilateral trade ratio controlling for the presence of differentiated products is denoted by the subscript "diff". We can then modify the numerator of expressions (1) to control for bilateral trade only in differentiated goods following James Rauch's classification as follows,

$$\begin{aligned} RatioINP_{abt}\_diff1 &= \frac{bilateral\,inputs_{abt}\_diff1}{Total\,bilateral_{abt}} \\ RatioINP_{abt}\_diff2 &= \frac{bilateral\,inputs_{abt}\_diff2}{Total\,bilateral_{abt}} \end{aligned} \tag{2}$$

Notice that a direct comparison between $RatioINP_{abt}\_1$ and $RatioINP_{abt}\_diff1$ indicates that the former is greater than the latter since it includes both homogeneous and differentiated trade flows in the numerator. In contrast, the latter includes differentiated products. A similar analysis applies to $RatioINP_{abt}\_2$ and $RatioINP_{abt}\_diff2$. Last, we also control for the presence of manufactured goods in the bilateral ratios as suggested by

[36] The BEC labels intermediate goods as group "1", final goods as group "2", and capital goods as group "3".

the following expressions:

$$\begin{aligned} RatioINP_{abt}\_diffman1 &= \frac{bilateral\,inputs_{abt}\_diffman1}{Total\,bilateral_{abt}} \\ RatioINP_{abt}\_diffman2 &= \frac{bilateral\,inputs_{abt}\_diffman2}{Total\,bilateral_{abt}} \end{aligned} \quad (3)$$

A direct comparison suggests that the measure $RatioINP_{abt}\_diff1$ is greater than the measure $RatioINP_{abt}\_diffman1$ since its numerator includes manufactured and nonmanufactured products. Notice that expressions (1)-(3) are primarily measured using import values, since export values are less reliable, but we fill in missing import values with export values.

An assessment of columns 1, 2, and 3 of Table 1 shows that the average ratio of trade in inputs based on the measure $RatioINP_{abt}\_1$ for a country pair that belongs to a PTA (0.46) is higher than the average for a country pair for the entire sample (0.36) and is similar to the average for a country pair belonging to an EIA (0.45). This fact is in line with the idea that trade in intermediate inputs is an essential driver to the formation of a PTA and an EIA. Notice that a similar assessment applies to all other measures of trade ratios described in expressions (1)-(3), but the difference between the average for the full sample and the PTA/EIA sample is even relatively more important when controlling for manufactured-differentiated products (see Table 1 for details).[37] Figure 1 suggests that the difference between the average value of the ratios for these samples applies over time. A comparison between the left-hand side panel (PTA sample) and the right-hand side panel (entire sample) of Figure 1 shows that the average ratio of trade in inputs across years tends to be higher for the country-pairs with a PTA than for the entire sample.

[Figure 1 here]

2.3 Interdependence and Other Control Variables

An essential point made in the literature (see Egger and Larch (2008)) indicates that PTA formation could encourage other countries to form new PTAs or lead to the enlargement of existing PTAs. We construct different measures of interdependence separately for the presence of PTAs and EIAs. Our baseline measure follows Egger and Larch (2008)

[37] In this case, the average of $RatioINP_{abt}\_diffman1$ for PTA country pairs is 32 percent higher than the equivalent average for the entire sample.

and describes for each country the average share of other country pairs that have a PTA in place using the distance between country pairs as weights. We name this variable $INTERDPTA_{abt}$. Table 1 suggests that this variable's average value for the PTA sample (0.413) is much greater than its value for the entire sample (0.207).

We also measure interdependence across EIAs. As explained above, if a country-pair adopts property rights and investment-protection specific regulations, other regional country-pairs that jointly produce different parts or components for various industries will more likely join an EIA both by enlarging an existing one or forming a new one. We measure the interdependence among EIAs using the variable $INTERDEIA_{abt}$, whose concept is equivalent to the variable $INTERDPTA_{abt}$.[38] Again, its average value for the EIA sample (0.01) is greater than its value for the entire sample (0.002). This result confirms that interdependence may be an essential driver of PTAs and EIAs. Notice that we also measure interdependence following the theory-driven approach proposed by Baldwin and Jaimovich (2012) for robustness analysis.

Our empirical specifications control for other essential drivers of the formation of PTAs, according to Baier and Bergstrand (2004) and Egger and Larch (2008). We include the variable $Natural_{ab}$, which corresponds to the inverse of the distance between two country pairs. We expect it to have a positive effect on PTA formation. Table 1 suggests that the average value for the EIA sample (-7.77) is greater than its value for the entire sample (-8.68). Next, we control for the size of the country pairs economy with the variable $GDPSUM_{abt}$, and we expect it to have a positive effect on the probability of PTA formation. We find that the average value for the PTA sample (11.148) is slightly greater than its value for the entire sample (11.07). $GDPSIM_{abt}$ calculates the similarity in the economic size of a country pair. We can verify that the average value for the PTA sample (-1.90) is greater than its value for the entire sample (-2.528). The variable $DCONT_{ab}$ takes on the value one if the country pair is located on the same continent and zero otherwise. We expect both variables to have a positive effect on the likelihood of PTA formation. Looking at the averages again, we see that the average value for the PTA/EIA sample (0.655/0.657) is much higher than its value for the entire sample (0.245).

We also consider the degree of remoteness of a country pair with the variable

[38]The only difference is that this measure focuses on the share of other country pairs with an EIA in place rather than a PTA

$REMOTE_{abt}$. It corresponds to the weighted average distance between each country in a pair and its trade partners. This variable is expected to have a positive effect on PTA formation, according to Baier and Bergstrand (2004) and has an average value of (8.758) for the PTA sample, which is slightly lower than its value for the entire sample (8.928). However, we get a negative sign in our analysis. The variable DKL_{abt} calculates the difference in the capital-labor ratio between countries a and b at time t. $SDKL_{abt}$ is the square value of DKL_{abt}. The latter variable is expected to have a positive effect on the probability of PTA formation up to a point, which implies that $SDKL_{abt}$ should have a negative effect. Both have average values for the PTA sample (1.311, 2.70), which are lower than their counterparts for the entire sample (1.881, 5.275). $DROWKL_{abt}$ calculates the average relative differences in factor endowments between a country pair and its trade partners. Here, we also find that the average value for the PTA sample (1.305) is higher than its value for the entire sample (1.251). These averages seem well in line with existing literature (see Egger and Larch (2008) and Baldwin and Jaimovich (2012)). The details about the construction for all these variables can be found in Table A in the appendix.

3 Econometric Methodology

In this section, we discuss the econometric strategy used to investigate the role of trade in differentiated inputs and the role of interdependence in determining the formation of PTAs and EIAs. Moreover, we also discuss some of the robustness tests we use in light of some alternative methodologies used in the literature.

Using our panel data with country-pair information from 1980 to 2015, we model the formation of a PTA using a Probit model, given the binary nature of the variable we want to investigate. As discussed above, we also consider the drivers of the formation of PTAs that cover services and become an EIA. Section 2.1 makes clarifies that about 60 percent of the PTA country pairs become an EIA, and some famous PTAs first covered only goods to add services to their coverage years later. As such, we extend the baseline Probit model to investigate the formation of a PTA using a Probit model with sample selection. As robustness, we also test our results using Logit and Linear Probability Models. Facchini et al. (2021) use a similar approach in explaining the decision to form a PTA and the choice between FTAs and CUs. We can then interpret the decision to form a PTA and an EIA as a two-step process. Initially, the countries make a decision to form a PTA or not, and then agree on its depth, i.e., on whether the PTA will also cover services. Thus, the first stage involves self-selection into a PTA, and the decision to form an EIA is made in the second (latent) stage if a PTA is formed in the first stage.

The empirical approach employed in this paper follows the Probit models proposed by Baier and Bergstrand (2004), Egger and Larch (2008), among others. But, there are a few differences between these papers. Our approach relies on a panel dataset spanning through 35 years, close to the approach used in Egger and Larch (2008) to investigate the role of interdependence on PTA formation. Instead, Baier and Bergstrand (2004) and Bergstrand and Egger (2013) rely on a cross-sectional dataset to investigate the determinants of PTA formation. For this reason, we discuss below a robustness tests that uses a cross-section of our dataset.

One common concern in using economic country-pair characteristics to explain the decision to form PTAs is the potential for endogeneity problems that create biases in our estimates. The first problem is potential reverse causality. The formation of PTAs affects trade flows, which can have potential effects on incomes, for instance. We first control for reverse causality by using five-year lagged values of country-pair characteristics in determining the probability of PTA formation.[39] A second problem is missing variables that may be correlated to our country-pair characteristics. In this case, controlling for country-pair fixed effects may lead to the incidental parameters' problem, given the binary nature of our dependent variables. To address this problem, we follow Egger and Larch (2008) and Facchini et al. (2021), which use Chamberlain's (1980) strategy and use country-pair specific averages across years as separate regressors in our econometric strategy.

We use the following equation to describe the decision to form a PTA in the first stage of the selection process:

$$PTA_{abt} = \beta_0 + \beta_1 RatioINP_{abt-5}\_diff1 + \beta_2 INTERDPTA_{abt-5} + \beta_3 X_{abt-5} + \varepsilon_{abt} \quad (4)$$

PTA_{abt} is a binary variable that takes a value of 1 if a country-pair belongs to a preferential trade agreement in year t, and zero otherwise. We use country pair clusters in estimating the error term.

The variable $RatioINP_{abt-5}\_diff1$ measures the share of differentiated inputs on bilateral trade and follows expressions (2). We use different measures of the share of differen-

[39] Notice that Egger and Larch (2008), Chen and Joshi (2010), Bergstrand and Egger (2013), and Fontagné and Santoni (2021), among others, also rely on lagged variables to control for reverse causality.

tiated input products on bilateral trade for our analyses, as explained in Section 2.2. They include the share of bilateral trade in differentiated-manufactured inputs ($RatioINP_{abt-5}\_diffman1$ and $RatioINP_{abt-5}\_diffman2$) following expressions (3). Likewise, we show robustness tests, which include capital goods among input products ($RatioINP_{abt-5}\_diff2$) in Table C of the appendix. For completeness, we also discuss the effects of the content of trade on inputs ($RatioINP_{abt-5}\_1$ and $RatioINP_{abt-5}\_2$) following expressions (1). Notice that our baseline results estimate expression (4) controlling for the lagged value of interdependence among PTAs ($INTERDPTA_{abt-5}$). Matrix X contains all other control variables discussed in the previous section and is used extensively in the existing literature (see Section 2.3 and Table A for details). Moreover, it also contains the average of these variables at the country-pair level, which follows Chamberlain's strategy (1980).

Our first hypothesis can then be expressed in the following way:

Hypothesis 1 – The greater the share of differentiated-input products on bilateral trade, the higher the probability that a PTA is formed between countries a and b at year t (β_1 > 0).

As discussed above, a significant share of PTAs also covers services, i.e., becoming an EIA. Following the evidence, country-pairs that decide to form a PTA in the first stage, described by expression (4), then decide to extend the agreement and become an EIA. In other words, we model the decision to form an EIA as a second-stage decision for country pairs that have self-selected into a PTA. A similar approach is used by Facchini et al. (2021). We model the second stage decision for country pairs that have decided to form a PTA using the following expression:

$$EIA_{abt} = \alpha_0 + \alpha_1 RatioINP_{abt-5}\_diff1 + \alpha_2 INTERDEIA_{abt-5} + \gamma_{abt} \tag{5}$$

Our second stage analysis investigates the effects of the share of differentiated-input products on bilateral trade and the role of interdependence on the formation of EIAs. The dependent variable EIA_{abt} is a binary variable that equals one if an EIA is in place for a country-pair in year t and zero if, instead, there is only an agreement in goods. As such, we use a Probit model with sample selection to investigate the formation of EIAs. This strategy follows the econometric model proposed by Van de Ven and Van Praag (1981). The latent expression (5) controls for the share of bilateral trade that corresponds to differentiated inputs ($RatioINP_{abt-5}\_diff1$) and for the interdependence among EIAs

$(INTERDEIA_{abt-5})$[40]. The error terms ε_{abt} and γ_{abt} are assumed to have zero mean and are normally distributed. We estimate these error terms using country-pair clusters.

We can now summarize our second hypothesis:

Hypothesis 2 – If countries a and b have a PTA in place at year t, the greater their share of differentiated-input products on bilateral trade, the higher the probability that an EIA between them is in place ($\alpha_1 > 0$). Moreover, the greater the share of EIA country-pairs, the more likely countries a and b join or form an EIA ($\alpha_2 > 0$).

As described above, we consider numerous robustness tests. Expressions (4) and (5) are used to obtain the main results of the paper and are based on several papers in this literature. Still, we may want to test the robustness of our results by controlling for additional factors that may not be fully accounted for using the strategy that controls for country-pair averages outlined in Chamberlain (1980). We then alternatively run our main specification (4) using a Linear Probability Model, and with the assistance of country-pair and year-fixed effects, we estimate the following expression:

$$PTA_{abt} = \beta_t + \beta_{ab} + \beta_1 RatioINP_{abt-5}\_diff1 + \beta_2 INTERDPTA_{abt-5} + \beta_3 X\prime_{abt-5} + \varepsilon_{abt} \tag{6}$$

where β_t and β_{ab} represent year and country-pair fixed effects.[41] Notice that these fixed effects control for any time-invariant variables. These are the cases of some variables included in matrix $\boldsymbol{X}$ (see expression (4)) such as $Natural_{ab}$ and $DCONT_{ab}$. We then relabel the matrix of other controls based on the literature as ***X'*** in expression (5).

Likewise, we carry numerous robustness tests by using alternative datasets to measure the presence of PTAs and EIAs. In this case, we consider the measures used by Egger and Larch (2008) and Baier et al. (2014). In addition, we test our results using a more 'liberal' definition of differentiated goods following Rauch (1999). Importantly, we investigate the robustness of the importance of interdependence among PTAs and EIAs using the theory-driven measure provided by Baldwin and Jaimovich (2012).

Another set of robustness tests involves modifying the econometric strategy used to

[40] Our baseline results use the two right-hand-side controls in expression (5) while applying our Probit with sample selection strategy. This strategy reflects limited information on determinants of policy related to services, investments, and IPR relative to the numerous literature on PTAs. Notice that we include all elements of matrix X in the latent equation (expression (5)) as a robustness test in Table D of the appendix.

[41] This strategy controls for about 16500 fixed effects in total out of a sample of 69824.

estimate expressions (4) and (5). A set of tests focuses on estimating the formation of PTAs and EIAs simultaneously using a bivariate Probit model since many agreements create preferential access to goods and services simultaneously. Moreover, we notice that PTAs, once formed, usually are not dissolved. Thus, we estimate expressions (4) and (5) dropping country-pair observations that have been part of the same PTA. Likewise, we have employed five-year lagged values for our explanatory variables to address reverse causality concerns. We then assess our results using 10-year lagged variables to control for longer causes of PTA and EIA formation.

Furthermore, Baier and Bergstrand (2004) and Bergstrand and Egger (2013) argue that the predictions born out of these models are better understood as long-run predictions and, in their assessment, could be tested by using a cross-section of the data where the formation of PTAs in year t is explained by the value of PTA drivers for a particular year in the past. We then test hypotheses 1 and 2 using a cross-sectional model where we investigate the role of trade in differentiated-input products and interdependence measured using information from the year 2010 on the formation of PTAs in the year 2015. As discussed below, these robustness tests support hypotheses 1 and 2, indicating they are robust to these different econometric strategies and various measures of the dependent and explanatory variables.

4 Empirical Results

4.1 The Formation of PTAs

Table 2 contains our baseline results consisting of the estimation of expression (4) using a Probit model. In this case, we investigate the drivers of PTA formation, and then PTA_{abt} represents our dependent variable. Columns (1) through (4) access hypothesis 1 without controlling for interdependence ($INTERDPTA_{abt}$), while columns (6) through (9) control for it. We provide marginal effects of the PTAs drivers for the specification shown in columns (3) and (8) using columns (5) and (10), respectively. Notice that the number in parenthesis represents standard errors as usual.

[Table 2 here]

The specifications used in columns (1) and (6) control for the share of bilateral trade on inputs, $RatioINP\_1$ (expression (1)). Results shown in these columns indicate that the coefficient of this variable is positive, albeit not statistically significant. Instead, the specifications in columns (2) and (7) control for the share of bilateral trade in differentiated inputs, $RatioINP\_diff1$ (expression (2)). In this case, results suggest this variable has a positive coefficient on the probability that a PTA is formed as expected, and, importantly, it is statistically significant. These results support Hypothesis 1 and suggest that it is not inputs, in general, that drive PTA formation, but, more specifically, it indicates that the bilateral content in differentiated inputs leads to this result. Next, we notice that columns (3) and (8) control for the ratio of differentiated-manufactured inputs, $RatioINP\_diffman1$ (expression (3)). The results described in columns (3) and (8) suggest that the content of bilateral trade in differentiated-manufactured inputs positively affects the formation of PTAs, which is statistically significant. Importantly, the specifications in columns (4) and (9) control for the share of bilateral trade in differentiated inputs originating from the manufacturing sectors ($RatioINP\_diffman1$) alongside those originating from other sectors ($RatioINP\_diffnonman1$). These results suggest that the share of differentiated inputs from non-manufacturing sectors does not significantly affect PTA formation. In conclusion, the results from Table 2 indicate that the share of bilateral trade in differentiated inputs positively affects trade, particularly the inputs originating from the manufacturing sector. Non-differentiated inputs and differentiated ones that do not originate from manufacturing do not have a statistically significant effect on PTA formation. These results provide strong support for Hypothesis 1.

Notice that, as expected, the results shown in columns (6)-(10) indicate that interdependence ($INTERDPTA$) positively affects the probability of PTA formation. It also matters in determining the effect of the share of bilateral trade in differentiated inputs. For instance, the coefficient of $RatioINP\_diffman1$ is much smaller in column (8) than in column (3) (β_1=0.180 versus β_1 =0.248). These results suggest that ignoring interdependence has significant implications for the effects described in Hypothesis 1. Below, we mostly focus on the specification shown in column (8) to assess the economic importance of the results. As explained in Section 2, the measure $RatioINP\_diffman1$ represents the share of differentiated-manufactured inputs on bilateral trade, where we do not add trade in capital goods as part of the inputs traded. It then represents a more conventional measure, although the results are qualitatively similar and robust to including capital goods as part of the traded inputs, as discussed below. This finding strongly supports our hypothesis 1. Notice that our results are economically important. Let us consider the marginal effects shown in column (10). We conclude that a one-standard-deviation increase in the share of bilateral trade in differentiated-manufactured products increases the likelihood of PTA formation by (0.0202*0.202*100= 0.29) 0.40 percentage points. This result makes sense since the probability of a country pair belonging to a PTA is only about 12 percent in our data.[42]

More importantly, the specification used in column (8) can be used to measure the success rate of the model in explaining the share of country pairs with a PTA ("true positives") and the success rate in explaining the absence of PTAs across country pairs without a PTA ("true negatives"). In doing so, we follow Bergstrand and Egger (2013), which use the unconditional probability of PTA formation (12 percent in our case) as a threshold to determine when a country pair should form a PTA. If our model's predicted probability is greater (lower) than 12 percent, it indicates that countries should (not) form a PTA. Our analysis suggests that the model successfully predicts about 84 percent of the country pairs with a PTA in place and 83 percent of the country pairs that do not have a PTA in place. In sum, the model can correctly predict the presence and absence of PTAs in most cases.

The coefficients for variables used in other studies (Matrix X) align with the literature. More specifically, we find that a PTA is more likely if two countries are closer geographically ($NATURAL$), are on the same continent ($DCONT$), and if other country-pairs are already part of PTAs ($INTERDPTA$). Moreover, our results confirm that the likelihood of PTA formation rises with the country pair's market size ($GDPSUM$), the similarity in market sizes ($GDPSIM$), and asymmetry in factor endowments (DKL). Instead, our results do not corroborate the expected effects of the degree of remoteness ($REMOTE$). As in Egger and Larch (2008), we also find that the probability of creating an agreement is directly related to the difference in the relative factor endowment of the countries in a pair relative to the rest of the world ($DROWKL$).

We examine two readily available extensions from Table 2. Table B in the appendix uses the same specifications across columns while using a Logistic model like in Fontagné and Santoni (2021).[43] The results are similar, showing that the bilateral share of differentiated inputs positively affects the probability of PTA formation, notably inputs originating from the manufacturing sector. Likewise, the marginal effects shown in column (10) are very similar to the counterparts from Table 2, suggesting that a one-standard-deviation in-

[42] For the interpretation:
1. Margins in the table are: 0.0202 and 0.0315 for differentiated manufactured inputs with and without interdependence respectively.
2. Standard deviation in the descriptive table is 0.202 manufactured differentiated inputs
3. Multiply the reported standard deviation with the marginal effects

[43] We appreciate the suggestion made by referees to include this extension using Logit.

crease in the share of differentiated-manufactured inputs increases the probability of PTA formation by (0.0207*0.202*100) about 0.42 percentage points. Instead, Table C in the appendix relies on a definition of inputs that includes capital goods (e.g., $RatioINP\_diff2$ (expression 2)). These results again confirm the results outlined in Table 2 regarding the role played by differentiated inputs, particularly those originating from the manufacturing sector, in promoting PTA formation. As suggested above, the difficulty in finding replacement inputs in world markets makes the share of differentiated inputs an essential predictor of PTAs. Essentially, most differentiated products originate in manufacturing industries, which coincides with the greater fragmentation in producing these goods, a finding highlighted by Johnson and Noguera (2017).

4.2 The Formation of EIAs

As discussed in Section 2.1, we have 12 percent of our sample consisting of countries with a PTA, and about 60 percent of these countries have also chosen to have an agreement in Services, EIA. No country-pair in our sample has an agreement in services but not goods, a finding in line with Hofmann et al. (2019). Moreover, some of these agreements were formalized only in goods and later extended to include services. We can then investigate the formation of preferential agreements using the Heckman-Probit approach following Van de Ven and Van Praag (1981) and recently applied by Facchini et al. (2021) in a related context. This approach calls for estimating a latent equation (expression 5) with the decision about EIA formation for the countries that self-select into a PTA (expression 4). As such, each specification displayed in Table 3 offers the results for the EIA decision on the top panel, while, at the bottom panel, we have the results for the selection equation.

[**Table 3 here**]

Columns (1) and (2) report the results for the estimation of equations (4) and (5) for the same variants of the share of trade in differentiated and differentiated-manufactured inputs on bilateral trade used in Table 2, respectively. In this case, our dependent variable in the latent regression is EIA (see top panel), while PTA represents the dependent variable on the selection side (see bottom panel). Marginal effects are shown in column (5) to quantify the economic magnitudes of changes in our preferred measure $RatioINP\_diffman1$. The Wald test provided at the bottom of the table indicates that the Probit model with sample selection performs better than estimating equations (4) and (5) separately. Furthermore, the results found in columns (1) and (2) support our Hypotheses 1 and 2. Concentrating on the factors that determine the formation of a PTA (lower panel), we find that an increase in trade in differentiated and differentiated-manufactured inputs (columns (2) and (3)) will positively affect the probability that two countries establish a PTA, respectively. Moreover, these results are statistically significant at the 1% level, and they confirm the results outlined in Hypothesis 1.

Notably, the top panel results suggest that an increase in bilateral trade of differentiated and differentiated-manufactured inputs tends to increase the likelihood that an EIA will be formed. The results for the bilateral share of differentiated-manufactured inputs in column (2) are statistically significant at the 5 percent level. These results then confirm the predictions outlined in Hypothesis 2. It is important to highlight that these results are economically meaningful. As for the control variables related to PTA formation, our findings are in line with the existing literature (Baier and Bergstrand (2004) and Egger and Larch (2008)) and are similar to the ones reported in Table 2.

We then run similar regressions using different variants of the definition of an EIA. As explained in Ruta (2017) and Hofmann et al. (2019), preferential agreements have varied widely in the policy areas covered and have become deeper in the last two decades, covering many areas beyond goods and services. Thus, as explained in Section 2.1, we alternatively consider the effects of bilateral trade in differentiated inputs on the policy coverage of EIAs. In particular, columns (3)-(4) use as the dependent variable EIA_WTO_deep, in which an EIA covers all major WTO programs (GATS, TRIPS, and TRIMS) as well as investment policy and additional intellectual property rights not addressed under TRIPS. See Section 2.1 for details. Instead, columns (5)-(6) use the dependent variable EIA_WTO_leg, which also controls for enforceable rules in all major WTO programs, investment policies, and intellectual property rights. Finally, columns (7)-(8) control for EIA interdependence using the policy coverage defined in the dependent variable EIA_WTO_leg.

The results found on columns (3)-(4), (5)-(6), and (7)-(8) show strong support for our Hypothesis 1 and 2. Focusing on the decision to form an EIA (top panel), the results found in columns (3)-(8) show that the share of bilateral trade in differentiated and differentiated-manufactured inputs has a positive and statistically significant effect on the decision to form an EIA with additional protections to investors and property rights, with exception of column (7). The foundations of our hypothesis 2 are then confirmed, i.e., the higher the level of trade in differentiated-intermediate products, the greater the support to form an EIA with varying provisions that also protect the member countries' bilateral-specific economic goals.

We can then use statistical and economic analysis to select the policy coverage of EIAs that are mostly affected by the share of trade in differentiated inputs on bilateral trade. In this case, we rely on a comparison using both the Akaike (AIC) and Schwartz (BIC) criteria (see bottom of Table 3) while focusing on our preferred measure of the share of trade in inputs in selecting the best specification, $RatioINP\_diffman1$. Besides, we consider the success rates of the models in explaining 'true positives' and 'true negatives" in EIA formation (see bottom of Table 3). These two criteria unequivocally suggest that the model with the definition of EIA covering the major WTO programs (GATS, TRIPS, and TRIMS) and that additionally covers investment, property rules protections, and enforceability serves as our best model (EIA_WTO_leg).[44] In columns (7)-(8), we control for the interdependence among EIA ($INTERDEIA$), based on our definition of EIA comprising of this policy coverage (EIA_WTO_leg). As expected, the coefficient of interdependence is positive and strongly statistically significant, confirming the importance of EIA interdependence in promoting the formation of EIAs. These results provide ample support to Hypothesis 2. Moreover, the AIC and BIC criteria selects these specifications over the specifications used in columns (5)-(6). Likewise, the success rates shown at the bottom of Table 3 corroborates the selection of the specifications controlling interdependence (columns (7) and (8)) as well.

More importantly, the results for the specification used in column (8) show that our preferred measure for trade on inputs and EIA interdependence are economically important. In this case, we find that a one-standard-deviation increase in the variable $RatioINP\_diffman1$ (see column (9) for marginal effects) leads to an increase in the probability of EIA formation of 1.08 percentage points (0.064*0.168*100). This result is important since the conditional probability (given the decision to form a PTA) of forming an EIA with the policy coverage included in variable EIA_WTO_leg is 5.78 percent. Thus, it increases the probability of this type of EIA formation by 19 percent relative to the average conditional probability. Likewise, the economic importance of interdependence in forming EIA is similarly relevant: a one-standard-deviation increase in EIA interdependence ($INTERDEIA$) leads to an increase of 3.29 percentage points (1.173*0.028*100), or, equivalently, a 57 percent increase relative to the conditional average that an EIA is formed. Thus, the bilateral share of differentiated inputs and EIA interdependence are statistically and economically

[44]The lower the values of the AIC and SIC criteria the better the statistical performance.

important in determining EIA formation.

Likewise, our model specified in column (8) successfully explains the formation of PTAs and EIAs. As already mentioned, the occurrence of a PTA amounts to only 12 percent in our dataset. Whenever the predicted probability is greater than or equal to that threshold (0.12), we assume that the model forecasts the existence of a PTA membership. In this case, our model correctly predicts about 84 percent of the country pairs with a PTA. Likewise, it can accurately predict the absence of a PTA about 83 percent of the time. These results confirm the findings outlined in Table 2.

Most importantly, among those country pairs with a PTA, about 60 percent have an EIA in place. However, only 5.78 of PTAs also cover WTO+ and WTO-X areas along side with their legal enforceability. Therefore, the conditional cutoff probability for an EIA with this deep coverage, given that a PTA exists, is 5.78 percent. The results from specification (8) suggest that our model successfully predicts 78 percent of the observations involving country pairs belonging to an EIA with this coverage (EIA_WTO_leg) when a PTA exists. Likewise, our model can correctly predict about 64 percent of the observations involving country pairs that do not belong to a PTA.

We dedicate Table D in the appendix to show that controlling for matrix X elements in the latent equation (EIA formation) does not alter our results.[45] The specifications across columns (1)-(8) add one element of matrix X at a time, but do not affect the positive effect that the share of differentiated-manufactured inputs and EIA interdependence ($RatioINP\_diffman1$ and $INTERDEIA$) have on EIA formation. The specification used in column (9) controls for all elements of matrix X concurrently. Still, the coefficient for key variables ($RatioINP\_diffman1$ and $INTERDEIA$) remain positive and statistically significant. Instead, Table E in the appendix estimates the equations for PTA and EIA formation jointly using a bivariate Probit specification. We use the different EIA variables controlling for different policy areas. Our key results are robust to estimating both equations jointly. However, we notice that the Wald Test shown at the bottom of Table 3 suggest the Probit model with sample selection is the appropriate model in line with the presence of agreements that extend to cover service-related policies after an agreement in goods is in place.[46]

[45] We appreciate the suggestion made by a referee to include this extension of our results.
[46] We appreciate the suggestion made by a referee to include this extension of our results.

5 Robustness Checks

This section considers various extensions to our benchmark analysis discussed in Tables 2 and 3. First, we consider our results through a linear probability model where we control for multiple fixed effects. Next, we investigate the robustness of the results by considering alternative datasets and definitions for our main variables related to interdependence and differentiated inputs. Lastly, we turn our attention to different econometric and modeling strategies.

5.1 Liner Probability Model

The literature points out that using fixed effects with non-linear estimation strategies such as Probit models may lead to biased estimates. Following Chen and Joshi (2010), we test our results by estimating expression (4) to study the formation of PTAs under the presence of a large set of year and country-pair fixed effects. A motivation to use country-pair fixed effects are the findings in Egger and Shingal (2021) that asymmetry in SRTI affect the probability of preferential agreements in services. These measures usually do not vary much by year, or the data available do not vary yearly. Using year and country-pair fixed effects means the inclusion of more than 16000 binary variables, which forces us to rely on a linear probability model. As discussed in the paragraph following expression (6), the introduction of these fixed effects means that some elements of the matrix X are dropped ($NATURAL, REMOTE, and DCONT$) since these variables do not vary (much or at all) by time.

The results of the estimation of expression (6) confirm our major prediction outlined in Hypothesis 1. Our variables measuring the share of bilateral trade on differentiated and differentiated-manufactured inputs have the expected positive effects on PTA formation, and all of them are statistically significant. Columns (1)-(4) report findings without the interdependence variable, whereas columns (5)-(8) include the interdependence variable in the regression equation. Focusing on the specification used in column (7), we conclude that the probability of a country pair forming a PTA increases by 0.25 percent (0.0126*0.202), given a one-standard-deviation increase in the share of bilateral trade on differentiated-manufactured inputs. Likewise, we find qualitatively similar success rates of the model in predicting PTA formation and their absence relative to Tables 2 and 3.

[Table 4 here]

5.2 Robustness using different datasets and variables definition

Table 5 focuses on the robustness of our results to different data on PTA information, the definition of how PTAs are interdependent, and the definition of trade in differentiated inputs. Columns (1) through (4) in Table 5 contain our baseline results (see column (8) of Table 2), consisting of the estimation of equation (4) using a Probit model. In this case, we investigate the PTA formation's drivers, where PTA_{abt} represents our dependent variable. Columns (1) through (4) test hypothesis 1 controlling for the interdependence ($INTERDPTA_{abt}$) variable. Instead, Columns (5) through (8) investigate the formation of preferential agreements using the Heckman-Probit approach, where expression (5) explains the decision to form an EIA for the country pairs that self-select into forming a PTA according to expression (4). The specifications used in columns (5)-(8) display the results for the EIA decision on the bottom panel, while, at the top panel, we have the results for the selection equation.

[Table 5 here]

As mentioned earlier, preferential trading agreements have rapidly increased over time. Over the years, efforts have been made to collect information on the nature of various agreements in force. The datased provided by Baier et al. (2014) focuses on the de facto presence of PTAs. Instead, the PTA formation provided by Egger and Larch (2008) focuses on the de jure statements about PTA formation made by the WTO. Therefore, to assess the robustness of our analysis, the specifications used in columns (1) and (5)[47] rely on the PTA information used in Baier et al. (2014). Instead, columns (3) and (7) focus on the definition of PTAs and EIAs following Egger and Larch (2008). As such, these robustness tests focus on the definition of our variables PTA_{abt} and $INTERDPTA_{abt}$ and their counterparts for EIAs, EIA_{abt}, and $INTERDEIA_{abt}$.

The coefficient of the variable of focus, $RatioINP\_diffman1$, in column (1), where we use the information on PTAs provided by Baier et al. (2014) is positive and statistically significant at the 1% level, providing strong support for Hypothesis 1. This implies that the share of bilateral trade on differentiated-manufactured inputs does increase the probability of countries in a pair to form a PTA. Column (5) reports the results for the estimation of equations (4) and (5) for $RatioINP\_diffman1$. In this case, our dependent variable

[47] Facchini et al. (2021) follow a similar approach.

in the latent regression is EIA (see bottom panel), while PTA represents the dependent variable on the selection side (see top panel). The Wald test provided at the bottom of the table indicates that the Probit model with sample selection performs better than estimating equations (4) and (5) separately.

Furthermore, the results found in column (5) support our Hypotheses 1 and 2 while using information from Baier et al. (2014). Concentrating on the factors that determine the formation of a PTA (top panel), we find that an increase in trade of differentiated-manufactured inputs will positively affect the probability that two countries will establish a PTA. This result is positive and statistically significant at the 1% level. Furthermore, the bottom panel result suggests that an increase in bilateral trade of differentiated-manufactured inputs tends to increase the likelihood that an EIA will be formed. Likewise, the interdependence among EIAs has a positive and statistically significant effect in promoting the formation or enlargement of EIAs. All these results are highly significant and confirm the predictions outlined in Hypothesis 2. As for the control variables related to PTA formation, our findings are in line with the existing literature (Baier and Bergstrand (2004), Egger and Larch (2008), Facchini et al. (2021)) and are similar to the ones reported in Table 2.

As mentioned earlier, Egger and Larch (2008) highlight the importance of interdependence among PTAs in the probability that a country joins an existing PTA or forms a new one. However, in their paper, Baldwin and Jaimovich (2012) offer a theory-driven measure of PTA interdependence, where the country-pair decision to form a PTA depends on their direct and relative trade exposure to partner countries involved with other PTAs. To verify our previous results, where we use the approach specified by Egger and Larch (2008) to create the interdependence variable, the specifications used in columns (2) and (6) define interdependence among PTAs ($INTERDPTA_{abt}$) and EIAs ($INTERDEIA_{abt}$) following Baldwin and Jaimovich (2012).

We find that the coefficient of the variable of focus, $RatioINP\_diffman1$, in columns (2) and (5), is positive and statistically significant at the 1% level, providing strong support for both Hypothesis 1 and 2. This implies that the share of bilateral trade on differentiated-manufactured inputs does increase the probability of countries in a pair to form a PTA even using an alternative definition of PTA interdependence. Furthermore, the bottom panel result in column (6) suggests that an increase in bilateral trade of differentiated-manufactured inputs tends to increase the likelihood that an EIA will be formed. Moreover, the greater the number of EIA country-pairs, the more likely countries a and b join or form an EIA ($\alpha_2 > 0$).

We dedicate columns (3) and (7) to test the robustness of our results to using the information on PTAs and EIAs used by Egger and Larch (2008). In line with previous results, the specification used in column (3) confirms the robustness of our results in support of Hypothesis 1. Likewise, the results in column (6) show that trade in differentiated-manufactured inputs and EIA interdependence positively affect the EIA decision, providing further support to Hypothesis 2. Since Egger and Larch (2008) dataset uses the de jure information provided to the WTO, our results are robust to using the de facto or the de jure information on PTAs and EIAs. In terms of success rates, the specification used in column (7) correctly predicts the presence of PTAs in 82.6 percent of the cases and correctly predicts the presence of EIAs in 61.5 percent of the cases. Instead, these results allow us to predict the absence of EIA (true negatives) in 73.6 percent of the cases. These results confirm the ability of our empirical model in predicting the presence of PTAs and EIAs.

Next, we look at alternative approaches to measure the bilateral trade share on differentiated inputs to verify our results. As previously mentioned, we built measures of the bilateral trade share on inputs using the information on bilateral trade flows at the industry level. Then, we merged this dataset with James Rauch's (1999) Classification to identify differentiated products. His dataset offers a conservative and liberal definition of differentiated products. In our baseline results, we use the conservative approach to define our main variables represented by expressions (2) and (3). To test our results' robustness, we rebuilt these measures of the share of inputs in the bilateral trade flows using his liberal approach and used them in the specifications shown in columns (4) and (8). The results shown in these two specifications indicate that our benchmark analysis' qualitative conclusions are not affected by switching to the liberal definition in calculating the measure of bilateral trade.

5.3 Robustness using different econometric strategies

Table 6 presents several robustness tests related to the econometric model, concerns with endogeneity and institutional details related to the timeline used to implement preferential tariffs, robustness to economically relevant subsamples, and possible dependence of some results on the E.U. case. The specifications used in columns (1) through (6) represent expression (4), which is based on our use of a Probit model, and it controls for the interdependence ($INTERDPTA_{abt}$) variable. Columns (7) through (12) investigate the formation of preferential agreements using the Heckman-Probit approach following Van de Ven and

Van Praag (1981). We show the estimated results for the latent equation (expression 5) on the bottom panel of Table 6, while the results about the PTA decision (expression 4) are placed on the top panel.

[Table 6 here]

The formation of a PTA is rarely withdrawn, which can generate error terms that are correlated over time. We then test our results by estimating expression (4) while dropping the country pairs that have formed a PTA in the previous years. In this case, a country pair only stays in the dataset until a PTA is signed and is dropped after that. Moreover, notice that this strategy decreases simultaneity problems between the PTA variable and its lagged-explanatory variables since a PTA formed in the past could affect elements of the X matrix (e.g., country-pair size ($GDPSUM$)), which is used to explain the decision to form a PTA in the future.[48] The results are shown in (1). Likewise, we extend this approach to explain the EIA formation for country-pairs that have formed a PTA in column (7).

The results shown in columns (1) suggest that our key variable, $RatioINP\_diffman1$, is positive and statistically significant at the 5% level, providing strong support for Hypothesis 1. This result implies that the share of bilateral trade on differentiated-manufactured inputs does increase the probability of countries in a pair forming a PTA even when dropping observations for country pairs with PTAs formed in the past. Furthermore, the bottom panel result in column (7) suggests that an increase in bilateral trade of differentiated-manufactured inputs tends to increase the likelihood that an EIA will be formed even when dropping observations for country pairs that have formed an EIA in the past and is statistically significant at the 1% level. Importantly, EIA interdependence remains positively and statistically affecting the formation of EIA, confirming our baseline results.

Our baseline specifications employed a five-year lag for our explanatory variables to address the concern related to reverse causality. Instead, in columns (2) and (8), we check the robustness of our results using a ten-year lag for the variables explaining the decision to form a PTA and the variables explaining the decision to form an EIA for the country-pairs that have decided to form a PTA, respectively. This strategy assists us in capturing the longer-term determinants of the PTA formation process. Once again, our results align with the previous findings and confirm predictions outlined in Hypotheses 1 and 2. All our

[48] Baldwin and Jaimovich (2012) follow a similar approach throughout their paper.

coefficients of interest are positive and significant at the 1% level.

Baier and Bergstrand (2004) and Bergstrand and Egger (2013) argue that some predictions derived from static trade models are better empirically investigated as cross-section outcomes. As such, we estimate expressions (4) by focusing on the PTA formation in the year 2015 while using a cross-section of our variables for the year 2010. Likewise, we consider its extension to the formation of EIAs as well. The results can be found in columns (3) and (9). In both cases, our variable $RatioINP\_diffman1$ is positive and statistically significant in column (3) and the selection part of column (9). However, it is not statistically significant in the latent part of column (9).

It is a fact that most PTAs have been formed since the end of the 20th century. We then wish to test the robustness of our results for the year 2000 and preceding years, and after the year 2000. We achieve this objective by estimating the decision to form a PTA up to 2000 in columns (4) while dedicating column (5) to study the PTA formation decision from 2005 and onwards. Equivalently, we explain the decision to form an EIA for country pairs that have decided to form a PTA in columns (10) and (11). The coefficient of the variable of focus ($RatioINP\_diffman1$) is positive in all these specifications. Likewise, the interdependence of PTAs and EIA continues to show it positively and statistically affect PTA and EIA formation. These results confirm the robustness of our findings for different time samples of our data.

The introduction of this paper points out that one of the most popular PTAs is the European Union (E.U.). It is also the first comprehensive post-WWII PTA. The E.U. has gone through several rounds of expansion since its inception in 1957, with just six original members that later expanded to a total of 28 member countries by the year 2015.[49] In columns (6) and (12), we investigate whether our main predictions are robust to dropping the significant number of PTA country-pairs related to the E.U. from our sample. The results suggest that our predictions are robust to dropping CU country-pairs associated with this critical PTA since our primary variable of focus ($RatioINP\_diffman1$) is positive and significant at least at the 1 percent level, except for the statistical significance in the latent equation in column (12). Again, the EIA interdependence plays an important and significant role in fostering EIA formation. The bottom of Table 6 shows the success rates for each specification. The results suggest that success rates in terms of 'true positive' and

[49]More information on E.U. is available at https://europa.eu/european-union/index_en.

‘true negatives’ are mostly above 80 percent, reflecting positively on our model’s ability to explain the geography of PTAs and EIA.

6 Conclusion

In this paper, we have discussed two hypotheses that provide innovative insights into the process of PTA formation and then looked further into the formation of an EIA with varying degrees of coverage. Our analysis suggests that trade in differentiated inputs plays an essential part in the country-pair decision to form a PTA and in their decision to expand it to trade in services with varying degrees of coverage, which transforms a preferential agreement into an EIA. We then assess our hypotheses empirically, using a sample of 189 countries from 1980 to 2015. Our empirical analysis provides strong support for our predictions. In particular, the greater the trade of manufactured-differentiated inputs, the more likely a country-pair will have a PTA in place, which will later turn into an EIA with varying degrees of coverage.

Since the number of PTAs keeps increasing, it is vital to understand the PTA formation process. It’s then essential to test new theories that help explain this phenomenon. Also, the fact that the WTO’s multilateral approach has stagnated shows that PTAs represent the best chances to liberalize different economies further. PTAs play an increasingly important role in the multilateral trade system not only by liberalizing trade in goods beyond multilateral negotiations. These agreements tend to foster additional rules that diminish the uncertainty regarding trade policy in goods and, depending on coverage, by providing assurances for services and investment flows, including FDI. This reason makes it vital to dig further into understanding the PTA formation process.

References

Antràs, P. and R. W. Staiger (2012). Offshoring and the role of trade agreements. *American Economic Review 102*(7), 3140–3183.

Bagwell, K., C. P. Bown, and R. W. Staiger (2016). Is the wto passé? *Journal of Economic Literature 54*(4), 1125–1231.

Bagwell, K. and R. W. Staiger (1999). An economic theory of gatt. *American Economic Review 89*(1), 215–248.

Baier, S. L. and J. H. Bergstrand (2004). Economic determinants of free trade agreements. *Journal of international Economics 64*(1), 29–63.

Baier, S. L. and J. H. Bergstrand (2007). Do free trade agreements actually increase members' international trade? *Journal of international Economics 71*(1), 72–95.

Baier, S. L., J. H. Bergstrand, and M. Feng (2014). Economic integration agreements and the margins of international trade. *Journal of International Economics 93*(2), 339–350.

Baldwin, R. and D. Jaimovich (2012). Are free trade agreements contagious? *Journal of international Economics 88*(1), 1–16.

Bergstrand, J. H. and P. Egger (2013). What determines bits? *Journal of International Economics 90*(1), 107–122.

Blanchard, E. J. (2010). Reevaluating the role of trade agreements: Does investment globalization make the wto obsolete? *Journal of International Economics 82*(1), 63–72.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. Technical report, National Bureau of Economic Research.

Chen, M. X. and S. Joshi (2010). Third-country effects on the formation of free trade agreements. *Journal of International Economics 82*(2), 238–248.

Cole, M. T. and A. Guillin (2015). The determinants of trade agreements in services vs. goods. *International Economics 144*, 66–82.

Egger, H., P. Egger, and D. Greenaway (2008). The trade structure effects of endogenous regional trade agreements. *Journal of international Economics 74*(2), 278–298.

Egger, P. and M. Larch (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics 76*(2), 384–399.

Egger, P. and A. Shingal (2021). Determinants of services trade agreement membership. *Review of world economics 157*(1), 21–64.

Egger, P. H. and A. Shingal (2017). Granting preferential market access in services sequentially versus jointly with goods. *The World Economy 40*(12), 2901–2936.

Estevadeordal, A., C. Freund, and E. Ornelas (2008). Does regionalism affect trade liberalization toward nonmembers? *The Quarterly Journal of Economics 123*(4), 1531–1575.

Facchini, G., P. Silva, and G. Willmann (2013). The customs union issue: Why do we observe so few of them? *Journal of international economics 90*(1), 136–147.

Facchini, G., P. Silva, and G. Willmann (2021). The political economy of preferential trade agreements: an empirical investigation. *The Economic Journal 131*(640), 3207–3240.

Fontagné, L. and G. Santoni (2021). Gvcs and the endogenous geography of rtas. *European Economic Review 132*, 103656.

Hofmann, C., A. Osnago, and M. Ruta (2017). Horizontal depth: a new database on the content of preferential trade agreements. *World Bank Policy Research Working Paper* (7981).

Hofmann, C., A. Osnago, and M. Ruta (2019). The content of preferential trade agreements. *World Trade Review 18*(3), 365–398.

Horn, H., P. C. Mavroidis, and A. Sapir (2010). Beyond the wto? an anatomy of eu and us preferential trade agreements. *The World Economy 33*(11), 1565–1588.

Johnson, R. C. and G. Noguera (2017). A portrait of trade in value-added over four decades. *Review of Economics and Statistics 99*(5), 896–911.

Laget, E., A. Osnago, N. Rocha, and M. Ruta (2020). Deep trade agreements and global value chains. *Review of Industrial Organization 57*, 379–410.

Lake, J. and H. M. Yildiz (2016). On the different geographic characteristics of free trade agreements and customs unions. *Journal of International Economics 103*, 213–233.

Limão, N. (2006). Preferential trade agreements as stumbling blocks for multilateral trade liberalization: Evidence for the united states. *American Economic Review 96*(3), 896–914.

Liu, X. and E. Ornelas (2014). Free trade agreements and the consolidation of democracy. *American Economic Journal: Macroeconomics 6*(2), 29–70.

Mattoo, A., A. Mulabdic, and M. Ruta (2022). Trade creation and trade diversion in deep agreements. *Canadian Journal of Economics/Revue canadienne d'économique 55*(3), 1598–1637.

Mulabdic, A., A. Osnago, and M. Ruta (2017). *Deep integration and UK–EU trade relations*. Springer.

Nicita, A., M. Olarreaga, and P. Silva (2018). Cooperation in wto's tariff waters? *Journal of Political Economy 126*(3), 1302–1338.

Ornelas, E. and P. Tovar (2022). Intra-bloc tariffs and preferential margins in trade agreements. *Journal of International Economics 138*, 103643.

Osnago, A., N. Rocha, and M. Ruta (2019). Deep trade agreements and vertical fdi: The devil is in the details. *Canadian Journal of Economics/Revue canadienne d'économique 52*(4), 1558–1599.

Prusa, T. J. and R. Teh (2010). Protection reduction and diversion: Ptas and the incidence of antidumping disputes. Technical report, National Bureau of Economic Research.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of international Economics 48*(1), 7–35.

Ruta, M. (2017). Preferential trade agreements and global value chains: Theory, evidence, and open questions. *World Bank Policy Research Working Paper* (8190).

Subramanian, A. and S.-J. Wei (2007). The wto promotes trade, strongly but unevenly. *Journal of international Economics 72*(1), 151–175.

Tabakis, C. and M. Zanardi (2019). Preferential trade agreements and antidumping protection. *Journal of International Economics 121*, 103246.

Tsirekidze, D. (2021). Global supply chains, trade agreements and rules of origin. *The World Economy 44*(11), 3111–3140.

Van de Ven, W. P. and B. M. Van Praag (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of econometrics 17*(2), 229–252.

WTO (2011). The wto and preferential trade agreements: From co-existence to coherence.

WTO and IDE-JETRO (2011). *Trade Patterns and Global Value Chains in East Asia: From Trade in Goods to Trade in Tasks.* WTO.

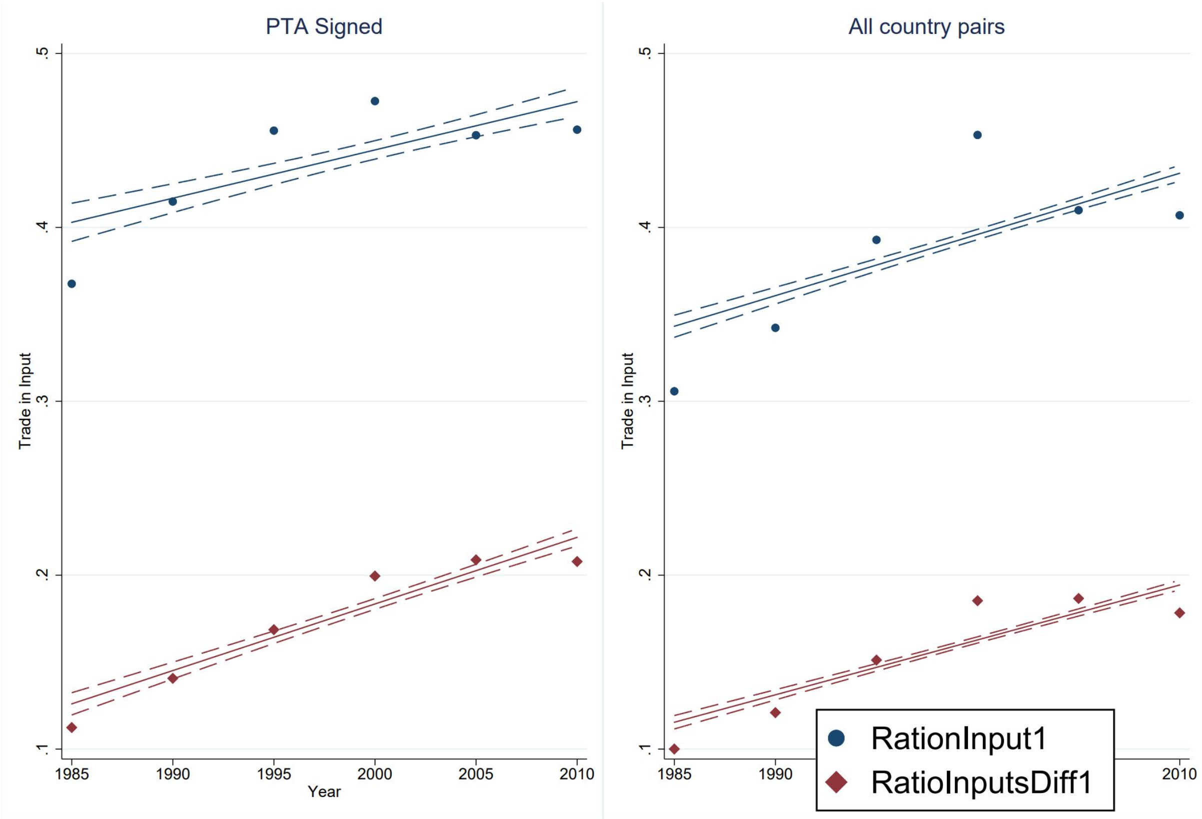

Figure 1: Evolution of trade over time

Table 1: Descriptive Statistics

Main Variables	Entire Sample	PTA	EIA
RatioINP_1	0.3605	0.4565	0.452
	(0.3577)	(0.2864)	(0.288)
RatioINP_2	0.4552	0.5846	0.588
	(0.3964)	(0.3121)	(0.315)
RatioINP_diff1	0.1529	0.2026	0.21
	(0.2234)	(0.1808)	(0.191)
RatioINP_diff2	0.2476	0.3305	0.346
	(0.2957)	(0.2473)	(0.26)
RatioINP_diffman1	0.1278	0.1745	0.183
	(0.2020)	(0.1675)	(0.179)
RatioINP_diffman2	0.2218	0.3020	0.319
	(0.2811)	(0.2408)	(0.255)
INTERDEIA	0.0020	0.009	0.01
	(0.0100)	(0.028)	(0.032)
Matrix X Elements			
INTERDPTA	0.2072	0.4135	0.423
	(0.1485)	(0.1858)	(0.184)
NATURAL	-8.6798	-7.6812	-7.767
	(0.7966)	(0.9800)	(1.029)
RGDPSUM	11.0702	11.1479	11.123
	(0.7720)	(0.8651)	(0.947)
RGDPSIM	-2.5275	-1.90	-1.952
	(1.8152)	(1.3510)	(1.422)
DKL	1.8810	1.3110	1.185
	(1.3180)	(0.9910)	(0.933)
SQDKL	5.2751	2.7010	2.274
	(6.2314)	(3.5620)	(3.178)
DCONT	0.2455	0.6550	0.657
	(0.4304)	(0.4750)	(0.475)
REMOTE	8.9277	8.7580	8.772
	(0.2143)	(0.2890)	(0.284)
DROWKL	1.2512	1.3050	1.386
	(0.6010)	(0.6110)	(0.672)
Number of Observations	69,824	8,364	5,013

Table 2

VARIABLES	(1) PTA	(2) PTA	(3) PTA	(4) PTA	(5) Margins	(6) PTA	(7) PTA	(8) PTA	(9) PTA	(10) Margins
RatioINP_1	0.0368					0.00740				
	(0.0257)					(0.0276)				
RatioINPdiff_1		0.197***					0.137***			
		(0.0351)					(0.0380)			
RatioINP_diffman1			0.2477***	0.248***	0.0315***			0.180***	0.1799***	0.0202***
			(0.0390)	(0.0391)	(0.00496)			(0.0423)	(0.0423)	(0.00474)
RatioINP_diffnonman1				-0.0177					-0.0484	
				(0.0716)					(0.0775)	
INTERDPTA						3.305***	3.292***	3.289***	3.289***	0.369***
						(0.111)	(0.112)	(0.112)	(0.112)	(0.0128)
NATURAL	0.683***	0.696***	0.697***	0.697***	0.0887***	0.410***	0.422***	0.423***	0.423***	0.0474***
	(0.0253)	(0.0257)	(0.0258)	(0.0258)	(0.00310)	(0.0252)	(0.0254)	(0.0255)	(0.0255)	(0.00286)
RGDPSUM	1.823***	1.823***	1.820***	1.821***	0.232***	0.932***	0.928***	0.926***	0.927***	0.104***
	(0.0325)	(0.0325)	(0.0325)	(0.0325)	(0.00464)	(0.0457)	(0.0459)	(0.0459)	(0.0459)	(0.00526)
RGDPSIM	0.444***	0.444***	0.446***	0.446***	0.0568***	0.257***	0.253***	0.255***	0.255***	0.0285***
	(0.0318)	(0.0318)	(0.0318)	(0.0318)	(0.00406)	(0.0339)	(0.0336)	(0.0336)	(0.0336)	(0.00377)
DKL	0.208***	0.208***	0.209***	0.209***	0.0266***	0.237***	0.236***	0.237***	0.237***	0.0265***
	(0.0433)	(0.0435)	(0.0435)	(0.0435)	(0.00554)	(0.0466)	(0.0465)	(0.0465)	(0.0465)	(0.00521)
SQDKL	-0.0224**	-0.0207*	-0.0195*	-0.0199*	-0.00248*	-0.0184	-0.0173	-0.0167	-0.0168	-0.00187
	(0.0109)	(0.0109)	(0.0109)	(0.0110)	(0.00139)	(0.0119)	(0.0119)	(0.0119)	(0.0119)	(0.00133)
DCONT	0.400***	0.395***	0.391***	0.393***	0.0498***	0.551***	0.543***	0.541***	0.541***	0.0606***
	(0.0343)	(0.0343)	(0.0344)	(0.0344)	(0.00445)	(0.0360)	(0.0360)	(0.0361)	(0.0361)	(0.00406)
REMOTE	-0.836***	-0.804***	-0.805***	-0.804***	-0.102***	-0.175**	-0.162**	-0.162**	-0.162**	-0.0182**
	(0.0738)	(0.0744)	(0.0745)	(0.0745)	(0.00947)	(0.0797)	(0.0795)	(0.0796)	(0.0796)	(0.00891)
DROWKL	0.611***	0.609***	0.609***	0.609***	0.0775***	0.544***	0.543***	0.544***	0.543***	0.0609***
	(0.0367)	(0.0370)	(0.0370)	(0.0370)	(0.00475)	(0.0401)	(0.0401)	(0.0401)	(0.0401)	(0.00453)
Constant	10.32***	9.848***	9.962***	9.912***		0.919	0.686	0.749	0.742	
	(0.657)	(0.655)	(0.656)	(0.656)		(0.779)	(0.774)	(0.775)	(0.775)	
Observations	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824
AIC										
BIC										
Success rate										
PTA==1	82%	81.80%	81.90%	81.90%		83.81%	83.99%	83.93%	83.94%	
PTA==0	80%	79.88%	79.89%	79.87%		83.30%	83.25%	83.26%	83.26%	

We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	EIA_{abt}		$EIA\_WTO-deep_{abt}$		$EIA\_WTO\_leg_{abt}$		$INTERDEIA$		Margins
RatioINP_diff1	0.137		0.297***		0.198*		0.170		
	(0.0906)		(0.114)		(0.110)		(0.112)		
RatioINP_diffman1		0.239**		0.485***		0.368***		0.345***	0.0639***
		(0.101)		(0.119)		(0.114)		(0.115)	(0.0217)
INTERDEIA							6.129***	6.334***	1.173***
							(1.228)	(1.257)	(0.230)
PTA									
RatioINP_diff1	0.138***		0.130***		0.129***		0.130***		
	(0.0381)		(0.0383)		(0.0382)		(0.0382)		
RatioINP_diffman1		0.182***		0.170***		0.168***		0.170***	0.0191***
		(0.0424)		(0.0426)		(0.0425)		(0.0425)	(0.00478)
INTERDPTA	3.276***	3.272***	3.303***	3.301***	3.304***	3.301***	3.303***	3.300***	0.371***
	(0.111)	(0.111)	(0.111)	(0.111)	(0.110)	(0.110)	(0.110)	(0.111)	(0.0127)
NATURAL	0.426***	0.427***	0.421***	0.423***	0.420***	0.422***	0.420***	0.422***	0.0474***
	(0.0255)	(0.0256)	(0.0253)	(0.0253)	(0.0253)	(0.0253)	(0.0253)	(0.0254)	(0.00286)
RGDPSUM	0.927***	0.926***	0.942***	0.940***	0.941***	0.940***	0.938***	0.937***	0.105***
	(0.0458)	(0.0458)	(0.0456)	(0.0456)	(0.0455)	(0.0456)	(0.0456)	(0.0456)	(0.00525)
RGDPSIM	0.253***	0.254***	0.256***	0.257***	0.256***	0.257***	0.255***	0.257***	0.0288***
	(0.0337)	(0.0337)	(0.0338)	(0.0338)	(0.0338)	(0.0338)	(0.0338)	(0.0338)	(0.00380)
DKL	0.239***	0.240***	0.224***	0.225***	0.223***	0.225***	0.225***	0.226***	0.0254***
	(0.0465)	(0.0465)	(0.0465)	(0.0465)	(0.0465)	(0.0465)	(0.0465)	(0.0465)	(0.00523)
SQDKL	-0.0167	-0.0161	-0.0167	-0.0160	-0.0172	-0.0165	-0.0174	-0.0167	-0.00188
	(0.0119)	(0.0119)	(0.0120)	(0.0120)	(0.0121)	(0.0121)	(0.0121)	(0.0121)	(0.00136)
DCONT	0.540***	0.538***	0.555***	0.552***	0.553***	0.550***	0.553***	0.550***	0.0618***
	(0.0360)	(0.0360)	(0.0360)	(0.0361)	(0.0360)	(0.0360)	(0.0360)	(0.0360)	(0.00407)

Table 3: (continued)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	EIA_{abt}		$EIA\_WTO-deep_{abt}$		$EIA\_WTO\_leg_{abt}$		$INTERDEIA$		Margins
REMOTE	-0.164**	-0.165**	-0.108	-0.109	-0.108	-0.109	-0.113	-0.114	-0.0128
	(0.0795)	(0.0795)	(0.0785)	(0.0786)	(0.0784)	(0.0785)	(0.0786)	(0.0786)	(0.00883)
DROWKL	0.537***	0.537***	0.562***	0.562***	0.566***	0.566***	0.565***	0.565***	0.0635***
	(0.0402)	(0.0402)	(0.0394)	(0.0395)	(0.0392)	(0.0393)	(0.0393)	(0.0394)	(0.00447)
No of Observations	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824
Selected	8,364	8,364	8,364	8,364	8,364	8,364	8,364	8,364	8,364
Not Selected	61,460	61,460	61,460	61,460	61,460	61,460	61,460	61,460	61,460
Wald test of independent Equations	2.21	2.35	56.11	54.03	66.44	64.04	53.01	50.96	
Prob > chi2	0.137	0.126	0.000	0.000	0.000	0.000	0.000	0.000	
Wald chi2	14.560	21.510	41.120	60.860	36.18	53.79	61.770	80.780	
Prob > chi2	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Log pseudolikelihood	-20009.96	-19999.04	-16213.93	-16200.01	-16108.89	-16096.71	-16097.47	-16085.04	
AIC	40063.93	40042.08	32471.85	32444.02	32261.77	32237.41	32242.94	32218.08	
BIC	40265.31	40243.46	32673.24	32645.4	32463.15	32438.79	32462.63	32437.77	
Success rate									
PTA==1	83.98%	83.99%	83.89%	84.03%	83.91%	84.05%	83.90%	84.04%	
PTA==0	83.27%	83.29%	83.23%	83.26%	83.23%	83.27%	83.24%	83.26%	
EIA==1	48.87%	48.61%	74.70%	75.68%	77.43%	78.26%	77.43%	78.05%	
EIA==0	56.94%	57.54%	59.55%	61.03%	59.98%	61.19%	62.52%	63.63%	

We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Linear Probability Model

Variables	(1) PTA	(2) PTA	(3) PTA	(4) PTA	(5) PTA	(6) PTA	(7) PTA	(8) PTA
RatioINP_diff1	0.0168***				0.00945**			
	(0.00397)				(0.00378)			
RatioINP_diff2		0.0152***				0.00604**		
		(0.00316)				(0.00303)		
RatioINP_diffman1			0.0233***				0.0126***	
			(0.00449)				(0.00429)	
RatioINP_diffman2				0.0185***				0.00738**
				(0.00339)				(0.00324)
INTERDPTA					1.010***	1.009***	1.009***	1.009***
					(0.0226)	(0.0226)	(0.0226)	(0.0226)
RGDPSUM	-0.0959***	-0.0964***	-0.0962***	-0.0967***	-0.0590***	-0.0592***	-0.0592***	-0.0594***
	(0.0105)	(0.0105)	(0.0105)	(0.0105)	(0.0100)	(0.0100)	(0.0100)	(0.0100)
RGDPSIM	-0.00466	-0.00476	-0.00464	-0.00478	-0.00567*	-0.00571*	-0.00566*	-0.00571*
	(0.00350)	(0.00349)	(0.00350)	(0.00349)	(0.00328)	(0.00328)	(0.00328)	(0.00328)
DKL	0.0337***	0.0338***	0.0338***	0.0339***	0.0456***	0.0457***	0.0457***	0.0457***
	(0.00413)	(0.00413)	(0.00413)	(0.00413)	(0.00391)	(0.00391)	(0.00391)	(0.00391)
SQDKL	-0.0113***	-0.0113***	-0.0113***	-0.0113***	-0.0125***	-0.0125***	-0.0125***	-0.0125***
	(0.000769)	(0.000769)	(0.000769)	(0.000769)	(0.000737)	(0.000736)	(0.000737)	(0.000737)
DROWKL	0.0370***	0.0371***	0.0370***	0.0371***	0.0312***	0.0312***	0.0312***	0.0312***
	(0.00468)	(0.00468)	(0.00468)	(0.00468)	(0.00456)	(0.00456)	(0.00456)	(0.00456)
Constant	0.961***	0.966***	0.964***	0.969***	0.474***	0.475***	0.475***	0.477***
	(0.108)	(0.108)	(0.108)	(0.108)	(0.104)	(0.104)	(0.104)	(0.104)
Observations	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824
R-squared	0.711	0.711	0.711	0.711	0.735	0.735	0.735	0.735

We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Robustness Checks On Datasets And Variables' Construction

	(1) BBF-2014	(2) JB-2012	(3) ELPTA	(4) Liberal	(5) BBF-2014	(6) JB-2012	(7) ELPTA	(8) EIA-Liberal
Variables	PTA				PTA Decision (Selection)			
RationINP_diffman1	0.200***	0.224***	0.122***	0.153***	0.193***	0.217***	0.110***	0.143***
	(0.0601)	(0.0482)	(0.0392)	(0.0534)	(0.0482)	(0.0396)	(0.0395)	(0.0438)
INTERDPTA	2.436***	17.01***	2.832***	3.301***	2.266***	16.65***	2.859***	3.311***
	(0.133)	(1.869)	(0.104)	(0.131)	(0.129)	(2.397)	(0.103)	(0.111)
NATURAL	0.502***	0.544***	0.390***	0.425***	0.465***	0.546***	0.390***	0.423***
	(0.0160)	(0.0166)	(0.0253)	(0.0146)	(0.0301)	(0.0296)	(0.0252)	(0.0254)
RGDPSUM	1.060***	1.721***	0.658***	0.930***	1.162***	1.744***	0.668***	0.940***
	(0.0620)	(0.0335)	(0.0423)	(0.0525)	(0.0552)	(0.0339)	(0.0421)	(0.0457)
RGDPSIM	0.362***	0.428***	0.159***	0.255***	0.373***	0.432***	0.160***	0.257***
	(0.0419)	(0.0305)	(0.0296)	(0.0338)	(0.0432)	(0.0318)	(0.0299)	(0.0339)
DKL	0.275***	0.212***	0.208***	0.236***	0.268***	0.202***	0.200***	0.225***
	(0.0539)	(0.0456)	(0.0412)	(0.0481)	(0.0537)	(0.0441)	(0.0415)	(0.0466)
SQDKL	-0.0292**	-0.0214*	-0.0139	-0.0163	-0.0293**	-0.0214*	-0.0145	-0.0163
	(0.0147)	(0.0114)	(0.0109)	(0.0125)	(0.0147)	(0.0111)	(0.0112)	(0.0121)
DCONT	0.623***	0.423***	0.364***	0.544***	0.660***	0.435***	0.377***	0.553***
	(0.0224)	(0.0194)	(0.0374)	(0.0205)	(0.0399)	(0.0349)	(0.0373)	(0.0360)
REMOTE	0.256***	-0.919***	-0.219***	-0.163***	0.258***	-0.880***	-0.177**	-0.114
	(0.0492)	(0.0454)	(0.0761)	(0.048)	(0.0820)	(0.0767)	(0.0752)	(0.0786)
DROWKL	0.344***	0.591***	0.361***	0.545***	0.353***	0.624***	0.386***	0.566***
	(0.0472)	(0.0399)	(0.0355)	(0.0428)	(0.0434)	(0.0372)	(0.0351)	(0.0394)
					EIA decision (Latent)			
RatioINP_diffman1					0.557***	0.336***	0.435***	0.328***
					(0.144)	(0.116)	(0.109)	(0.116)
INTEDREIA					23.23***	6.509***	7.149***	6.262***
					(3.551)	(1.149)	(1.341)	(1.25)
No of Observations	69,824	69,759	69,824	69,824	69,824	69,759	69,824	69,824
No Of observations-Selected					6,303	8,326	8,903	8,364
Wald test of independent Equations					37.74	43.64	56.79	52.38
Prob > chi2					0.000	0.000	0.000	0.000
Success rate								
PTA==1	0.861	0.81	0.825	0.84	0.86	0.811	0.826	0.84
PTA==0	0.855	0.81	0.83	0.83	0.855	0.81	0.83	0.83
EIA==1					0.66	0.49	0.615	0.45
EIA==0					0.72	0.63	0.736	0.63

We use $EIA\_WTO\_leg_{abt}$ specification for EIA in columns 6,7 and 8. We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Robustness Checks on Econometric Strategy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	DropPTA==1	Ten year lag	Cross Section 2010	Sample<=2000	Sample>2000	ex-EU	EIAdropPTA==1	EIA-Ten year	Cross Section 2010	EIA-Sample<=2000	EIA-Sample>2000	EIA-exEU
VARIABLES	PTA						PTA decision (Selection)					
RatioINP_diffman1	0.297***	0.192***	0.212***	0.187	0.164**	0.181***	0.299***	0.179***	0.236***	0.187	0.162***	0.184***
	(0.0607)	(0.0641)	(0.0611)	(0.147)	(0.0647)	(0.0669)	(0.0601)	(0.0604)	(0.0693)	(0.118)	(0.0427)	(0.0589)
MATRIX X ELEMENTS												
INTERDPTA	11.66***	1.208***	3.454***	2.937***	3.546***	2.277***	12.14***	1.379***	3.981***	2.937***	3.569***	2.280***
	(3.293)	(0.149)	(0.132)	(0.311)	(0.238)	(0.173)	(4.236)	(0.121)	(0.142)	(0.186)	(0.151)	(0.137)
NATURAL	0.420***	0.550***	0.206***	0.606***	0.310***	0.545***	0.413***	0.545***	0.280***	0.606***	0.309***	0.543***
	(0.0204)	(0.0173)	(0.0298)	(0.0261)	(0.0181)	(0.0175)	(0.0230)	(0.0266)	(0.0325)	(0.0384)	(0.0277)	(0.028)
RGDPSUM	2.226***	2.332***	0.176***	1.367***	0.434***	1.662***	2.204***	2.321***	0.119***	1.367***	0.462***	1.647***
	(0.0439)	(0.0720)	(0.0226)	(0.148)	(0.0827)	(0.0837)	(0.0455)	(0.0713)	(0.0261)	(0.0485)	(0.118)	(0.0727)
RGDPSIM	0.497***	0.736***	0.0455***	0.402***	-0.0481	0.345***	0.492***	0.731***	0.00446	0.402***	-0.0420	0.342***
	(0.0385)	(0.0435)	(0.0100)	(0.0879)	(0.0510)	(0.0439)	(0.0400)	(0.0426)	(0.0116)	(0.0817)	(0.0314)	(0.0421)
DKL	0.347***	0.0925*	0.159***	0.462***	0.298***	0.164***	0.354***	0.0880*	0.172***	0.462***	0.292***	0.170***
	(0.0628)	(0.0544)	(0.0375)	(0.0993)	(0.0727)	(0.0613)	(0.0626)	(0.0505)	(0.0424)	(0.0851)	(0.0466)	(0.0549)
SQDKL	-0.0428***	-0.000865	-0.0601***	-0.0393	-0.0368**	-0.0281*	-0.0443***	-0.00237	-0.0806***	-0.0393	-0.0360***	-0.0294**
	(0.0153)	(0.0144)	(0.00845)	(0.0290)	(0.0180)	(0.0162)	(0.0151)	(0.014)	(0.0101)	(0.0251)	(0.0107)	(0.0146)
DCONT	0.143***	0.494***	0.291***	0.821***	0.492***	0.392***	0.137***	0.512***	0.435***	0.821***	0.506***	0.387***
	(0.0273)	(0.0239)	(0.0382)	(0.0433)	(0.0243)	(0.0255)	(0.0288)	(0.0386)	(0.0399)	(0.0645)	(0.0377)	(0.0414)
REMOTE	-0.738***	-0.425***	0.105	-0.578***	0.179***	0.00623	-0.798***	-0.388***	0.0283	-0.577***	0.225***	-0.0191
	(0.0605)	(0.0591)	(0.0798)	(0.0948)	(0.0543)	(0.0593)	(0.0610)	(0.0891)	(0.0903)	(0.123)	(0.0811)	(0.0908)
DROWKL	0.888***	-0.0693	-0.173***	0.133	0.546***	0.518***	0.869***	-0.0836*	-0.130***	0.133*	0.609***	0.516***
	(0.0539)	(0.0489)	(0.0234)	(0.0813)	(0.0743)	(0.0560)	(0.0552)	(0.0492)	(0.0251)	(0.0684)	(0.0463)	(0.0527)
							EIA decision (Latent)					
RatioINP_diffman1							0.590***	0.491***	0.0292	2.771***	0.0157	0.0979
							(0.221)	(0.157)	(0.179)	(1.065)	(0.104)	(0.226)
INTERDEIA							197.2***	24.61***	5.412***	40.65***	7.581***	7.731***
							(67.35)	(3.321)	(1.078)	(9.986)	(1.904)	(1.728)
No of Observations	63,614	54,460	13,672	29,137	40,687	54,148	63,614	54,460	13,534	29,137	40,687	54,148
No Of observations-Selected							2,220	5,543	2,036	1,829	6,535	4,322
Wald test of independent Equations							23.67	69.15	42.02	0.01	55.17	6.02
Prob > chi2							0.000	0.000	0.000	0.93	0.000	0.014
Success rates												
PTA==1	0.87	0.85	0.48	0.79	0.9	0.84	0.84	0.84	0.43	0.89	0.789	0.85
PTA==0	0.72	0.84	0.87	0.83	0.87	0.83	0.77	0.84	0.9	0.86	0.84	0.83
EIA==1							0.85	0.81	0.61	0.9	0.78	0.78
EIA==0							0.66	0.69	0.63	0.75	0.62	0.61

We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix

Table A: Control Variables

Variables	Description	Expected Signs
Natural	measures the log of the inverse of the great circle distance between two trade partners' capitals. The closer the country pair are, the more likely they are to form a PTA	+
RGDPSUM	log$(RGDP_{it} + RGDP_{jt})$	+
RGDPSIM	log$\{1$ - $[RGDP_{it}/(RGDP_{it} + RGDP_{jt})]^2 - [RGDP_{jt}/(RGDP_{it} + RGDP_{jt})]^2\}$	+
DKL	$\lvert \frac{log(RGDP_{it}}{POP_{it})}$ - log$\frac{(RGDP_{jt}}{POP_{jt})} \rvert$	+
DCONT	DCONT is a dummy variable that takes the value one if two countries are located at the same continent and zero otherwise.	+
REMOTE	$0.5\frac{log\Sigma_{k\neq j}Distance_{ik}}{(n_t-1)}$ + $\frac{log\Sigma_{k\neq j}Distance_{ik}}{(n_t-1)}$ + $\frac{log\Sigma_{k\neq i}Distance_{jk}}{(n_t-1)}$ trading partners are from the rest of the world, the more likely they are to form an FTA	+
DROWKL	$0.5\lvert log(\Sigma_{kt\neq it}RGDP_{kt}/\Sigma_{kt\neq it}POP_{kt} - log(RGDP_{it}/POP_{it})\rvert + \lvert log(\Sigma_{kt\neq jt}RGDP_{kt}/\Sigma_{kt\neq jt}POP_{kt} - log(RGP_{jt}/POP_{jt})\rvert$	-

Table B: Logit

VARIABLES	(1) PTA	(2) PTA	(3) PTA	(4) PTA	(5) Margins	(6) PTA	(7) PTA	(8) PTA	(9) PTA	(10) Margins
RatioINP_1	0.104**					0.0426				
	(0.0477)					(0.0524)				
RatioINP_diff1		0.393***					0.270***			
		(0.0657)					(0.0722)			
RatioINP_diffman1			0.497***	0.498***	0.0340***			0.348***	0.347***	0.0207***
			(0.0733)	(0.0735)	(0.00501)			(0.0808)	(0.0809)	(0.00481)
RatioINP_diffnonman1				-0.0341					-0.0485	
				(0.131)					(0.140)	
INTERDPTA						6.287***	6.274***	6.270***	6.270***	0.373***
						(0.225)	(0.225)	(0.226)	(0.226)	(0.0136)
NATURAL	1.255***	1.282***	1.283***	1.283***	0.0878***	0.742***	0.766***	0.768***	0.768***	0.0457***
	(0.0427)	(0.0430)	(0.0431)	(0.0431)	(0.00277)	(0.0465)	(0.0468)	(0.0468)	(0.0468)	(0.00278)
RGDPSUM	3.247***	3.253***	3.247***	3.250***	0.222***	1.626***	1.618***	1.616***	1.617***	0.0961***
	(0.0595)	(0.0596)	(0.0595)	(0.0596)	(0.00474)	(0.0856)	(0.0857)	(0.0858)	(0.0858)	(0.00526)
RGDPSIM	0.826***	0.821***	0.825***	0.825***	0.0565***	0.496***	0.486***	0.489***	0.489***	0.0291***
	(0.0599)	(0.0600)	(0.0600)	(0.0600)	(0.00414)	(0.0647)	(0.0641)	(0.0642)	(0.0642)	(0.00381)
DKL	0.395***	0.396***	0.397***	0.397***	0.0272***	0.478***	0.479***	0.481***	0.481***	0.0286***
	(0.0813)	(0.0819)	(0.0818)	(0.0820)	(0.00560)	(0.0891)	(0.0890)	(0.0890)	(0.0890)	(0.00529)
SQDKL	-0.0361*	-0.0341	-0.0318	-0.0327	-0.00218	-0.0312	-0.0307	-0.0295	-0.0297	-0.00175
	(0.0211)	(0.0211)	(0.0212)	(0.0212)	(0.00145)	(0.0230)	(0.0229)	(0.0229)	(0.0229)	(0.00136)
DCONT	0.729***	0.717***	0.710***	0.714***	0.0486***	1.029***	1.014***	1.011***	1.012***	0.0601***
	(0.0599)	(0.0597)	(0.0599)	(0.0598)	(0.00414)	(0.0661)	(0.0662)	(0.0663)	(0.0663)	(0.00396)
REMOTE	-1.560***	-1.492***	-1.495***	-1.492***	-0.102***	-0.439***	-0.410***	-0.412***	-0.412***	-0.0245***
	(0.138)	(0.139)	(0.139)	(0.139)	(0.00947)	(0.157)	(0.156)	(0.156)	(0.156)	(0.00930)
DROWKL	1.147***	1.144***	1.144***	1.142***	0.0783***	0.986***	0.986***	0.987***	0.986***	0.0587***
	(0.0690)	(0.0696)	(0.0696)	(0.0697)	(0.00484)	(0.0771)	(0.0769)	(0.0770)	(0.0770)	(0.00464)
Constant	19.08***	18.20***	18.43***	18.32***		2.763*	2.318	2.454	2.432	
	(1.275)	(1.268)	(1.267)	(1.27)		(1.544)	(1.532)	(1.532)	(1.534)	
Observations	69,824	69,824	69,824	69,824	69,824	69,824	69,825	69,824	69,824	69,824

We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C: Ratio2

Variables	(1) PTA	(2) PTA	(3) PTA	(4) PTA	(5) Margins	(6) PTA	(7) PTA	(8) PTA	(9) PTA	(10) Margins
RatioINP_2	0.102***					0.0443*				
	(0.0243)					(0.0260)				
RatioINPdiff_2		0.214***					0.134***			
		(0.0288)					(0.0309)			
RatioINP_diffman2			0.239***	0.240***	0.0302***			0.155***	0.154***	0.0173***
			(0.0302)	(0.0304)	(0.00381)			(0.0325)	(0.0326)	(0.00363)
RatioINP_diffnonman2				0.0127					-0.0307	
				(0.0703)					(0.0756)	
INTERDPTA						3.301***	3.283***	3.279***	3.279***	0.367***
						(0.112)	(0.112)	(0.112)	(0.112)	(0.0128)
NATURAL	0.686***	0.708***	0.709***	0.709***	0.0897***	0.418***	0.433***=	0.434***	0.434***	0.0486***
	(0.0254)	(0.0259)	(0.0260)	(0.0260)	(0.00310)	(0.0254)	(0.0257)	(0.0257)	(0.0257)	(0.00288)
RGDPSUM	1.816***	1.817***	1.813***	1.815***	0.229***	0.926***	0.922***	0.920***	0.921***	0.103***
	(0.0329)	(0.0331)	(0.0331)	(0.0332)	(0.00466)	(0.0461)	(0.0462)	(0.0462)	(0.0462)	(0.00528)
RGDPSIM	0.447***	0.445***	0.446***	0.446***	0.0564***	0.256***	0.251***	0.251***	0.251***	0.0281***
	(0.0323)	(0.0322)	(0.0321)	(0.0322)	(0.00408)	(0.0343)	(0.0339)	(0.0338)	(0.0339)	(0.00378)
DKL	0.210***	0.212***	0.214***	0.213***	0.0270***	0.238***	0.238***	0.239***	0.239***	0.0267***
	(0.0441)	(0.0442)	(0.0441)	(0.0442)	(0.00558)	(0.0470)	(0.0468)	(0.0467)	(0.0468)	(0.00523)
SQDKL	-0.0229**	-0.0200*	-0.0188*	-0.0193*	-0.00237*	-0.0185	-0.0172	-0.0165	-0.0166	-0.00185
	(0.0112)	(0.0111)	(0.0111)	(0.0112)	(0.00141)	(0.0121)	(0.0119)	(0.0120)	(0.0120)	(0.00134)
DCONT	0.410***	0.401***	0.397***	0.399***	0.0502***	0.554***	0.543***	0.541***	0.542***	0.0605***
	(0.0344)	(0.0344)	(0.0345)	(0.0345)	(0.00443)	(0.0360)	(0.0361)	(0.0361)	(0.0361)	(0.00406)
REMOTE	-0.818***	-0.769***	-0.771***	-0.769***	-0.0974***	-0.173**	-0.156*	-0.156**	-0.156**	-0.0175**
	(0.0743)	(0.0749)	(0.0750)	(0.0750)	(0.00948)	(0.0798)	(0.0797)	(0.0797)	(0.0797)	(0.00891)
DROWKL	0.608***	0.604***	0.603***	0.603***	0.0763***	0.542***	0.539***	0.540***	0.539***	0.0604***
	(0.0371)	(0.0375)	(0.0375)	(0.0375)	(0.00478)	(0.0403)	(0.0402)	(0.0402)	(0.0402)	(0.00454)
Constant	10.52***	9.873***	10.00***	9.946***		1.160	0.862	0.931	0.922	
	(0.661)	(0.659)	(0.658)	(0.659)		(0.781)	(0.774)	(0.775)	(0.775)	
Observations	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824

We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table D: Matrix X In The Latent Equation

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
EIA_WTO_leg									
RatioINP_diffman1	0.175	0.105	0.286**	0.263**	0.248**	0.401***	0.275**	0.217*	0.305*
	(0.111)	(0.169)	(0.119)	(0.119)	(0.120)	(0.123)	(0.111)	(0.123)	(0.182)
INTERDEIA	7.150***	7.947***	6.229***	5.964***	6.140***	6.648***	5.183***	8.025***	8.743***
	(1.463)	(1.804)	(1.296)	(1.277)	(1.277)	(1.280)	(1.114)	(1.380)	(1.593)
NATURAL	-0.309***								0.142***
	(0.0437)								(0.0503)
RGDPSUM		1.161***							2.344***
		(0.102)							(0.106)
RGDPSIM			0.373***						0.451***
			(0.115)						(0.114)
DKL				-0.485***					-0.0584
				(0.0793)					(0.138)
SQDKL					-0.0822***				-0.0197
					(0.0282)				(0.0355)
DCONT						0.241***			1.014***
						(0.0923)			(0.110)
REMOTE							0.509***		0.680***
							(0.0993)		(0.137)
DROWKL								1.875***	2.216***
								(0.151)	(0.159)
PTA									
RatioINP_diffman1	0.168***	0.176***	0.171***	0.171***	0.171***	0.171***	0.168***	0.174***	0.190***
	(0.0427)	(0.0423)	(0.0425)	(0.0425)	(0.0426)	(0.0425)	(0.0426)	(0.0425)	(0.0422)
INTERDPTA	3.254***	3.281***	3.294***	3.293***	3.293***	3.305***	3.292***	3.295***	3.197***
	(0.110)	(0.111)	(0.111)	(0.111)	(0.111)	(0.111)	(0.110)	(0.111)	(0.110)

Table D: Matrix X In The Latent Equation (Continued)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NATURAL	0.433***	0.422***	0.422***	0.422***	0.421***	0.423***	0.418***	0.422***	0.417***
	(0.0256)	(0.0254)	(0.0254)	(0.0254)	(0.0254)	(0.0254)	(0.0253)	(0.0254)	(0.0254)
RGDPSUM	0.947***	0.926***	0.938***	0.938***	0.939***	0.934***	0.939***	0.934***	0.963***
	(0.0453)	(0.0456)	(0.0456)	(0.0456)	(0.0456)	(0.0457)	(0.0455)	(0.0456)	(0.0457)
RGDPSIM	0.257***	0.256***	0.249***	0.257***	0.257***	0.256***	0.257***	0.256***	0.260***
	(0.0339)	(0.0335)	(0.0338)	(0.0337)	(0.0337)	(0.0338)	(0.0338)	(0.0339)	(0.0335)
DKL	0.220***	0.228***	0.227***	0.235***	0.223***	0.229***	0.224***	0.230***	0.236***
	(0.0465)	(0.0465)	(0.0466)	(0.0468)	(0.0464)	(0.0466)	(0.0465)	(0.0465)	(0.0466)
SQDKL	-0.0169	-0.0162	-0.0168	-0.0164	-0.0140	-0.0167	-0.0168	-0.0173	-0.0162
	(0.0123)	(0.0120)	(0.0121)	(0.0120)	(0.0119)	(0.0120)	(0.0121)	(0.0120)	(0.0119)
DCONT	0.551***	0.548***	0.550***	0.550***	0.551***	0.545***	0.554***	0.550***	0.554***
	(0.0357)	(0.0360)	(0.0360)	(0.0360)	(0.0360)	(0.0362)	(0.0359)	(0.0361)	(0.0360)
REMOTE	-0.0744	-0.114	-0.112	-0.115	-0.113	-0.125	-0.129*	-0.115	-0.222***
	(0.0793)	(0.0789)	(0.0787)	(0.0787)	(0.0787)	(0.0796)	(0.0782)	(0.0788)	(0.0800)
DROWKL	0.576***	0.564***	0.565***	0.564***	0.566***	0.561***	0.570***	0.529***	0.574***
	(0.0388)	(0.0396)	(0.0394)	(0.0394)	(0.0394)	(0.0396)	(0.0391)	(0.0402)	(0.0401)
Constant	-0.262	0.282	0.109	0.133	0.102	0.277	0.179	0.135	1.269
	(0.774)	(0.769)	(0.769)	(0.768)	(0.769)	(0.781)	(0.762)	(0.770)	(0.777)
No of Observations	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824	69,824
No Of observations-Selected	8,364	8,364	8,364	8,364	8,364	8,364	8,364	8,364	8,364
Wald test of independent Equations	113.8700	57.1300	50.3100	53.6700	58.3500	20.5100	106.8400	43.5200	34.5800
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

We use robust standard errors clustered at country pair level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table E: Bivariate Probit

Variables	(1) $EIA\_WTO-deep_{abt}$	(2) PTA	(3) $EIA\_WTO\_leg_{abt}$	(4) PTA
RatioINPdiffman_1	0.426***	0.221***	0.389***	0.221***
	(0.0443)	(0.0397)	(0.0445)	(0.0399)
INTERDEIA	12.00***		12.07***	
	(2.123)		(2.148)	
INTERDPTA		3.063***		3.076***
		(0.105)		(0.106)
NATURAL		0.399***		0.402***
		(0.0243)		(0.0244)
RGDPSUM		0.859***		0.868***
		(0.0435)		(0.0436)
RGDPSIM		0.227***		0.230***
		(0.0314)		(0.0315)
DKL		0.234***		0.232***
		(0.0439)		(0.0439)
SQDKL		-0.0187*		-0.0177
		(0.0109)		(0.0109)
DCONT		0.492***		0.500***
		(0.0346)		(0.0346)
REMOTE		-0.246***		-0.228***
		(0.0749)		(0.0749)
DROWKL		0.475***		0.477***
		(0.0385)		(0.0386)
Constant	-2.717***	1.911***	-2.741***	1.662**
	(0.0256)	(0.724)	(0.0257)	(0.724)
Observations	69,824	69,824	69,824	69,824

Robust standard errors clustered at the country pair level shown in parentheses. ***p<0.10, ** p<0.05, * p<0.10.

Declaration

Funding

This paper has not benefited from any specific research grant or other extramural funding sources.

Conflicts of interest/Competing interests

We declare having NO ties with any institution that may have a policy interest in the topics addressed in this paper.

Availability of data and material

The dataset used in this paper will be available from the authors upon request.

Code availability

Our statistical analysis is based on a common software (Stata, version 16) used in the profession, and the code used to generate the results are available upon request.

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent for publication

We agree to publish this article in the Review of World Economics if its editorial board finds it suitable for publication