Measuring Heterogeneity Across Preferential Trade Agreements

(Preliminary and Incomplete)

Maria Ptashkina[*] [†]

University Pompeu Fabra, PhD Candidate

March 13, 2020

Abstract

Most studies of preferential trade agreements (PTAs) use gravity models with PTAs as binary variables. This approach leaves out substantial amount of metadata and semantic information embodied within the texts. Over the last several decades the PTA texts have evolved, creating a considerable variation in the content and structure across agreements. I construct a novel dataset with individual characteristics of PTAs and a document-term matrix. Using text-mining techniques and predetermined intended semantic meanings, I build a continuous index measuring heterogeneity in legal enforceability across PTAs. Using the index in a structural gravity estimation indicates that once heterogeneity is taken into account, the effect of most PTAs is considerably dampened.

Key Words: International Trade, Gravity Model, Text Mining.

JEL Codes: *F13, F14, F15*

[*] E-mail: maria.ptashkina@barcelonagse.eu.

[†] I thank my advisors Jaume Ventura and Manuel García-Santana; the participants of International Lunch Seminar Series at UPF Fernando Broner, Giacomo Ponzetto, David Nagy, and Marta Santamaria; and the participants of the seminars at ECARES, Brussels.

1 Introduction

Virtually all countries in the world are a member of at least one preferential trade agreement (PTA). As of 2019, a total of 302 trade agreements were in force, corresponding to 481 notifications from the World Trade Organization members, counting goods, services and accessions. The provisions of these agreements cover 80-90% of bilateral trade between signatories and nearly 50% of the total world trade. Over the past several decades, trade agreements have increased in number, in depth, and in length, reflecting the rising importance of non-tariff measures as an instrument to regulate international trade. Moreover, PTAs nowadays govern a plethora of economic and social policies, establishing a link to trade issues. As a result of the evolution of trade texts, there is now considerable variation in structure, coverage and content across agreements.

Due to a lack of measurement indicators, most studies in the international trade literature use empirical specifications of a gravity model, where PTAs are coded as binary variables, indicating an agreement coming into force in a particular year for a given country pair. This approach leaves out any heterogeneity related to coverage, depth of the provisions, and implied levels of commitment under these agreements. In addition, it assumes away the importance of the semantic structures which underpin the regulations themselves. At the same time, these agreements are a product of years of intense negotiations carried out by countries' trained diplomats and lawyers, who shape the exact wording, and are ultimately responsible for the application and enforcement.

In this paper I suggest to investigate the importance of heterogeneity across trade agreements by explicitly taking into account their characteristics and semantic information embodied in the texts. While using text mining techniques to analyze legal texts has potentially wide applications, in this paper I focus only on measuring the heterogeneity associated with implied legal enforceability. It has long been recognized that the depth of PTAs, as measured by the credibility of the provisions and the presence of enforcement mechanisms is critical in generating investment and trade effects (Limão (2016)). However, due to measurement issues, there is lack of empirical evidence regarding the alleged association between contract enforcement mechanisms and trade values.

To measure the heterogeneity in legal enforceability, I construct a novel dataset containing individual characteristics of trade agreements and a document-term matrix. Using text-mining techniques and predetermined semantic intended word meanings, I identify words that correspond to "high" and "low" levels of enforceability within an agreement. These normalized word counts represent a proxy to measure the levels of enforceability of a given trade agreement. In addition, combining text-based measures with metadata, I construct a composite index representing the level of signatories commitment. These indexes, along with the other measures, are then used in a structural gravity estimation to measure the associated variation in bilateral trade flows. The baseline empirical specification covers the period from 1948 to 2015, and includes almost all ever-concluded trade deals.

The results indicate that representing the PTAs as homogeneous using dummy variables in a gravity model might lead to an overestimation of the magnitude of the average association between these agreements and trade volumes. Using the novel measurement presented in this paper, once the heterogeneity related to legal enforceability is accounted for, the estimate for the average association is dampened by 10 percentage points for the mean agreement. Focusing on an agreement with median legal enforceability levels, the estimates suggest that country pairs with such an agreement trade 30% more on average. However, the association can be as little as 6,5% increase in trade flows due to the presence of a PTA. Thus, there is a distribution of potential gains associated with the legal strength of agreements, which reconciles the widely controversial estimates in the literature.

Hundreds of empirical papers focused on measuring the effects of trade agreements. Head and Mayer (2014) do a meta-analysis of 159 papers with more than 2,500 usable estimates from the top-5 journals spanning publications from 2006 through 2012. Their findings on PTA effects suggest that in most studies PTAs are found to have large effects on trade, albeit with large standard deviations. The average coefficient across different studies is 0.59, which corresponds to around 80% of increase in bilateral trade as a result of PTA formation. Critically, all of these papers model PTAs as binary variables, not being able to account for heterogeneity across agreements.

Large differences in the estimates of PTA effects on trade volumes have been under scrutiny in the recent empirical explorations of heterogeneous effects of trade agreements. Kohl (2014) finds indicative evidence that deeper and more extensive agreements tend to be more effective. Baier et al. (2014) show that the type of a trade agreement matters for the extent to which PTAs may be related to trade flows. Kohl (2015) distinguish 17 trade-related policy domains, indicate whether the agreements contain legally enforceable commitments, and show that that trade agreement heterogeneity matters for international trade. Some studies attempt to account for heterogeneity in the design of particular agreements, including provisions on dispute settlement, investment, services or trade remedies (Lesher and Miroudot (2006); Houde et al. (2007); Fink and Molinuevo (2008); Kucik (2012)). Other studies investigate the use of trade agreements with respect to international cooperation (Estevadeordal and Suominen (2008)) or regional integration (Horn et al. (2010); Hicks and Kim (2012); Haftel (2013)). These studies largely rely on tedious and lengthy process of human collection of semantic information. I contribute to this literature by systematically exploiting semantic data of trade agreements using text mining techniques, to produce an easy-to-replicate and convenient measurement. In addition, most of the measures in the existing studies are still based on binary or count data, while the indicators proposed in this paper are continuous, allowing to unveil deeper heterogeneity across trade agreements.

The concept of legal enforceability of international contracts has been attracting attention in contract theory and political economy literature. In the recent paper Battaglini and Harstad (2020) identify the so-called "paradox of weak agreements," arguing that international environmental agreements generally do not include effective enforcement or monitoring mechanisms, against the background that treaty negotiations are expensive and laborious. Their theoretical model rationalizes this phenomenon with political economy considerations, but lacks a convenient empirical measurement of legal strength. More broadly, the measure presented in this paper can be applied to various types of legal contracts, and can help to empirically test theoretical predictions regarding various international treaties. In addition, the flexibility of the index construction methodology may allow to test theoretical predictions beyond the concept of legal enforceability.

Another related strand of literature is a recent empirical investigation on applying automated means of content analysis to social science questions (Grimmer and Stewart (2013); Gentzkow et al. (2017); Alschner et al. (2017a)). Baker et al. (2016) develop an index-based measure of uncertainty using newspaper article data, and a similar approach is employed in this paper. Texts of trade agreements are a relatively new source of data, and to my knowledge, at the time of writing, the two papers that systematically exploit PTA semantic structures are Alschner et al. (2017b) and Seiermann (2018). Both use the semantic similarities across texts to highlight some stylized facts or investigate the negotiating power of parties. In general, the applications of text mining to legal contracts, and, in particular, to international agreements, remains very limited. This paper thus aims to provide a stepping stone to a systematic exploration of legal texts.

The paper is organized as follows. Section 2 describes the dataset construction, while section 3 introduces the text-mining methodology. Section 4 presents the results of an empirical exercise using a structural gravity model. Section 5 concludes.

2 Data

To construct the PTA dataset I extract a collection of variables representing the characteristics of trade agreements into a readable easy-to-use format from a collection of PTA texts coded in unbalanced `XML` format provided by UNCTAD Text of Trade Agreements (ToTA) database. The database contains 450 texts with metadata on participating signatories, the type of agreement, date of signature, date of entry into force, status of the agreement as of year 2017, its composition, region, language and text source. In addition, I extract information on the number of chapters, number of articles and number of signatories of a given agreement. A cleaning procedure described in detail in the appendix, results in 410 agreements spanning the period from 1949 to 2017.

The number of trade agreements has been increasing over the years, particularly since the beginning of the 1990s, when more than 10 agreements entered into force yearly (see Figure 1 in the appendix). Some of the characteristics of the final PTA dataset are presented in Table 1. There are several types of deals: around 250 agreements are Free Trade Areas (FTAs), followed by FTAs with elements of Economic Integration Agreements (EIAs), Customs Unions (CUs), and Partial Scope Agreements (PSAs)[1]. Most of the agreements (around 62%) are notified under Article XXIV of the General Agreement on Trade in Goods (GATT). There are also around 30% of deals that cover both goods and services, while the rest of the agreements are concluded under the Enabling Clause[2]. More than half

[1]Free Trade Areas involve a reduction in trade barriers - import quotas and tariffs - but limited harmonization of trade policy instruments towards to third countries. Economic Integration Agreements are usually treaties among countriess in a geographic region to reduce and ultimately remove tariff and non-tariff barriers to the free flow of goods, services and factors of production. Customs Unions are type of a trade bloc which is composed of a free trade area with a common external tariff. The participant countries set up common external trade policy, but in some cases they use different import quotas. Partial Scope Agreements cover only certain products and are usually notified under the Enabling Clause.

[2]The Enabling Clause is the WTO legal basis for the Generalized System of Preferences (GSP), under which developed countries offer non-reciprocal preferential treatment to products originating in developing countries

of the agreements in the final dataset are bilateral, and the remaining 43% include more than two countries. The majority of the trade deals are in force as of 2017. Roughly half of the treaties is cross-regional, while the rest are concluded by countries within the same region.

The text corpus of trade agreements is cleaned using text mining techniques, stripped of punctuation and the so-called "stopwords" that do not carry any semantic meaning, and transformed into a Document-Term Matrix (DTM) with each cell representing a particular word count per agreement. There are on average around 10 chapters and 82 articles in a given agreement (see Table 2); and the average amount of both chapters and articles per agreement has been growing over the years (see Figures 2 and 3). Naturally, the total number of words per agreement has also been increasing (see Figure 4): the length of agreement texts has increased from a mean of about 2,800 words before 1990s to around 8,500 words after 1990s[3]. This increase has come from both the rise in the number of chapters, articles within each chapter, and from growing number of words within each article (see Figures 5 and 6).

Data on trade volumes (value of exports) between country pairs is extracted from from IMF Direction of Trade Statistics (DOTS) database (see details in the appendix). Gravity variables come from CEPII, and include physical distances between country pairs, their GDPs, sizes of the population, area, and indicator variables for the presence of common border, common language, common colonial history, currency unions, free trade agreements (coded as dummy variables), etc. Applied bilateral time-varying tariffs are extracted from World Bank's World Intergrated Trade Solutions (WITS) database.

The dataset for the baseline contains observations for 47,272 country-pairs, spanning the period from 1948 to 2015. It is further refined in several ways (see Table 3), and the results using the refinements are presented in the appendix. First, controlling for bilateral time-varying average applied tariffs reduces the number of observations to around 10% of the original

[3] Note that the total word count here is lower than a raw count due to the text cleaning procedure.

dataset. Second, a panel dataset is constructed[4], whereby every country pair is represented by no more than one agreement (selected randomly). Third, the dataset is refined only to observations of country-pairs having an agreement in place.

3 Text Mining

The goal of the text mining exercise in this paper is to provide a convenient and easily replicable example of a measurement of heterogeneity across trade agreements. Although there are potentially many applications of the similar bag-of-words techniques for other aspects of heterogeneity, this paper will focus on the concept of legal enforceability. Trade texts contain words (mostly verbs) that can proxy for a degree to which signatories are obliged to implement the measures and policies according to a given agreement. These words can indicate "high" and "low" levels of legal enforceability.

In order to identify which words have can be classified into the two commitment categories, information on the intended semantic meanings was drawn from trade law practice. A trade lawyer identified which words, generally defined, can potentially lead to a litigation or a dispute using international arbitrage, or trigger the process of suspension of parties' concessions. The words that correspond to "strong" enforcement (those which have a higher binding power in trade law) include: "shall," "commit," "require," "compliance" (often represented in a bi-gram "non-compliance"), "penalty," "accord," "adopt," "ensure." The second group of words, representing "weak" enforcement includes words such as "may," "endeavor," "aim," "cooperate," "dialogue," "possible" (often used as a part of the world collocation "to the extent possible")[5].

[4] The baseline specification, as explained in the appendix, is not a panel, since many country pairs have several active agreements at the same time.

[5] A widely used technique in Natural Language Processing known as 'stemming' has been used to reduce the words to their stem (root), thus capturing a wider use of the same semantic meaning, even with different word forms.

The following example illustrates the differences transmitted by semantic structures. In the chapters of the agreement related to early release of goods from customs, the comprehensive agreement between US and Korea (KORUS) states: "Each Party *shall* adopt and maintain procedures providing for the expeditious release of goods admitted under this Article." On the same issue, Korea's agreement with Vietnam reads: "Each Party shall *endeavor* to adopt and maintain procedures providing for the expeditious release of goods admitted under this Article." In the former agreement, the obligation to release goods is prescriptive, while in the latter it is indicative. In other words, a signatory that "endeavors" to adopt and maintain a trade measure does not necessarily has to actually implement it.

Following the initial identification of word groups, I construct an index which serves as a proxy for the legal enforceability of a given agreement. The simple version of the index takes the ratio of the total sum of "strong" words indexed from $\{1, ..., i\}$ to the sum of "weak" words indexed from $\{1, ..., j\}$ in a given document d:

$$\text{Simple Index}_d = \frac{\sum_i \text{Strong}_{i,d}}{\sum_j \text{Weak}_{j,d}} \tag{1}$$

Considering the variation in the total wordcount across agreements, and following Baker et al. (2016), I normalize the index by the total number of words in an agreement, indexed from $\{1, ..., w\}$, to get:

$$\text{Normalized}_d = \frac{\text{Simple Index}_d}{\sum_w \text{Words}_{w,d}} \tag{2}$$

There are other available characteristics of trade agreements which can be used for the construction of a composite index. In particular, the correlations among word-based measures and the characteristics of agreements suggest a number of associations (see Figure 8). First, there is a strong positive correlation between different word counts (total words, 'strong' and 'weak' words), and the number of chapters or articles. Second, there is positive correlation between the type of the agreement[6] and word count measures, naturally suggesting that more comprehensive agreements cover more issues and contain more words. Third, there is virtually no correlation between the constructed index, and the measurements of the agreements' size, such as the number of chapters or articles, or wordcount. Thus, these characteristics can potentially proxy for some additional information regarding the design of PTAs. The composite index is constructed as follows:

$$\text{Composite}_d = \omega_1 \text{Normalized}_d + \omega_2 \text{Chapters}_d + \omega_3 \text{Type}_d \tag{3}$$

where ω_1, ω_2 and ω_3 are the weights attributed to different components of the index[7].

Other specifications use measurements such as the ratio of the number of "strong" words over the total word count, or a "difference" measure constructed as follows:

$$\text{Difference}_d = \frac{\sum_i \text{Strong}_{i,d} - \sum_j \text{Weak}_j, d}{\sum_w \text{Words}_{w,d}} \tag{4}$$

Some descriptive statistics of the word counts and constructed text-based indexes are presented in Table 2. The distribution of the benchmark measure is presented in Figure 7, with a mean around 3.22, and with a longer right tail.

4 Gravity Model

To demonstrate the importance of taking into account the heterogeneity across trade agreements, I use empirical specifications of the structural gravity model, with the text-based measures as variables of interest. The main question of interest in the empirical estimation is the following: given

[6]Here the higher number represents a higher level of integration, ranging from PSAs (1) to FTAs (2), FTAs&EIAs (3), CUs (4), and CUs&EIAs (5).

[7]Different weights were used, and the results do not change qualitatively.

that a trade agreement enters into force for a given country pair, how taking account of the heterogeneity in legal enforceability is associated with the value trade flows.

Following the rapid methodological advances in the gravity literature, as summarized by Yotov et al. (2016) and Head and Mayer (2014), I employ multiple specifications for the baseline dataset, as well as for the regressions using data refinements presented in the appendix. Column (1) in each of the regression results tables employs the following simple version of the gravity model, in order to show consistency with the earlier estimates for the standard gravity variables:

$$X_{ijt} = \exp\left(\alpha_1 \boldsymbol{\pi}_{it} + \alpha_2 \boldsymbol{\chi}_{jt} + \alpha_3 \boldsymbol{\eta}_{ij} + \delta_i + \gamma_j + \phi_t + \beta_1 \text{Index}^I_{ijt}\right) \times \varepsilon_{ijt} \quad (5)$$

where X_{ijt} denotes the exports from origin i to destination j. Vector $\boldsymbol{\pi_{it}}$ captures time-varying source country characteristics (such as GDP and population to proxy for market size). Similarly, $\boldsymbol{\chi_{jt}}$ represents a set of destination country time-varying characteristics. Vector $\boldsymbol{\eta_{ij}}$ contains time invariant country-pair characteristics, such as distance, contiguity, and common official language. Origin, destination, and time fixed effects are captured by δ_i, γ_j and ϕ_t respectively. $\text{Index}^I_{ij,t}$ is the main variable of interest, and indicates variables related to preferential trade agreements. The superscript I denotes the set containing the usual dummy variable indicating the entry of a PTA into force, the different versions of the indexes presented in equations (1)-(4), the ratio of strong words to total, and a categorical variable taking value zero if there is no agreement, 1 or 2 if the value of the simple index is below or above average, respectively:

$$I = \begin{cases} \text{PTA Dummy} \in \{0,1\}; \\ \text{Simple Index}_d, \text{ defined in equation 1}; \\ \text{Normalized Index}_d, \text{ defined in equation 2}; \\ \text{Composite Index}_d, \text{ defined in equation 3}; \\ \sum_i \text{Strong}_{i,d} / \sum_w \text{Words}_{w,d}; \\ \text{Difference Index}_d, \text{ defined in equation 4}; \\ \text{Categotrical}_d \in \{0,1,2\} \end{cases}$$

Column (2) in each regression results table adds origin-time and destination -time fixed effects, to control for the unobservable multilateral resistances, and potentially for any other characteristics that vary over time for each exporter and importer (Anderson and van Wincoop (2003)). Following Baier and Bergstrand (2007), to account for endogeneity of trade policy, column (3) of each regression results table adds country-pair fixed effects (denoted by μ_{ij}), to account for the unobservable linkages between the endogenous trade policy covariates and the error term in gravity regressions, and thus estimates the following specification:

$$X_{ij,t} = \exp\left[\pi_{it} + \chi_{jt} + \phi_t + \mu_{ij} + \beta_1 \text{Index}^I_{ijt}\right] \varepsilon_{ijt} \tag{6}$$

where π_{it} and χ_{jt} are origin-time and destination-time fixed effects, respectively, and ϕ_t is the time fixed effect.

All the specifications, as well as all dataset refinements are estimated using both Ordinary Least Squares (OLS) and Poisson Pseudo Maximum Likelihood (PPML) estimators[8]. The latter has been shown to perform better in gravity estimation (Silva and Tenreyro (2006)), accounting for heteroskedasticity, taking advantage of the information contained in the zero trade flows, and ensuring that the gravity fixed effects are identical to their corresponding structural terms (Arvis and Shepherd (2013), Fally (2015)).

[8]The OLS specification uses log-log representation of all continuous variables, whereas PPML estimator is based on level-log regression.

In all estimations standard errors are clustered by trading pair in order to account for any intra-cluster correlations at the country-pair level.

The results of the estimation of the baseline sample are provided in Table 1. Each gravity specification includes only one element of the set I. The coefficients on GDPs, bilateral distance, contiguity, and common language are presented for the regressions including the PTA dummy, although neither the coefficient levels, nor the robust standard errors (presented in parenthesis below the estimates) vary substantially when running the model with other elements of the set I.

The coefficients on continuous variables show the export elasticity with respect to bilateral distances and GDPs (for example, $\hat{\beta}^{PPML}_{Distance} = -1.397$ indicates that a 10% increase in distance is associated with 13.97% decrease in bilateral trade). The corresponding coefficients for the presence of a PTA dummy suggest that having a trade deal in place is associated with approximately 93% higher trade value in the case of OLS estimator, and around 40% higher value in the case of PPML estimator. The goal of the specifications in Table 1 is to demonstrate that the estimates for the standard gravity variables are well in line with the traditional results, summarized in Head and Mayer (2014).

The estimated coefficients in the text-based measurements essentially represent the effects of the interaction between a PTA dummy and the continuous legal enforceability indicators. Once the heterogeneity of the trade treaties in this dimension has been taken into account, the association between trade and an average agreement drops by 30 percentage points for the OLS estimator and by 10 percentage points for the PPML estimator. In particular, in the case of the OLS estimator, the index indicates that having in place a PTA with an average level of legal enforceability would result in approximately 63% higher trade flows compared to no PTA in place. The figure drops to 30% using the results of the PPML estimation. Given high symmetry in the distribution of the normal index, the association between trade and an agreement with median legal enforceability level is also around 30%. For the agreements with legal enforceability levels in the bottom quantile (the least enforceable agreements) the association with trade is as low as 6.5-10.%.

Naturally, the interpretation of the index variables is not trivial. The coefficient of interest on the simple index shows that an increase by 10% in the index, which would mean a relative increase of the number of “strong” words relative to “weak” words, is associated with an additional 1.9-3.5% increase in export volumes. Intuitively, consider a hypothetical trade agreement text with 300 “strong” words and 100 “weak” words, with a corresponding value of simple index equal to 3. If negotiators decide to replace 7 “weak” words across the entire agreement (regardless of a particular chapter or measure) with 7 “strong” words, which means the increase of the index for this particular agreement by 10%, this would correspond to 1.9-3.5% higher bilateral trade flows for the countries which have this agreement in place. While this may seem as a big number, consider, for example, an agreement between Armenia and Kazakhstan, where the word count is 38 over 9, and changing several words can substantially alter the meaning and the intention of a trade deal. The normalized index and the composite index are hard to interpret, but they show the same direction of the coefficient, and are statistically significant at 1%.

Table 1: Estimation Results: Baseline Sample

	OLS (1)	OLS (2)	PPML (1)	PPML (2)
Origin GDP	0.823***		0.748***	
	(0.019)		(0.033)	
Destination GDP	0.611***		0.617***	
	(0.017)		(0.041)	
Distance	-1.397***	-1.391***	-0.782***	-0.787***
	(0.021)	(0.020)	(0.035)	(0.034)
Contiguity	0.648***	0.595***	0.584***	0.557***
	(0.107)	(0.101)	(0.075)	(0.078)
Language	0.774***	0.756***	0.159**	0.171**
	(0.041)	(0.040)	(0.068)	(0.068)
PTA Dummy	0.621***	0.658***	0.334***	0.355***
	(0.031)	(0.033)	(0.054)	(0.055)
Simple Index	0.353***	0.375***	0.193***	0.197***
	(0.021)	(0.022)	(0.033)	(0.034)
Normalized Index	0.268***	0.282***	0.095***	0.086***
	(0.014)	(0.015)	(0.022)	(0.023)
Composite Index	0.273***	0.291***	0.124***	0.124***
	(0.014)	(0.015)	(0.023)	(0.023)
Strong/Total	9.547***	9.957***	6.243***	6.195***
	(0.666)	(0.706)	(1.098)	(1.125)
Difference Index	10.214***	10.739***	7.058***	6.810***
	(0.910)	(0.949)	(1.369)	(1.417)
Categorical Index				
Below Average	0.795***	0.845***	0.350***	0.396***
	(0.034)	(0.037)	(0.060)	(0.063)
Above Average	0.171***	0.204***	0.303***	0.285***
	(0.053)	(0.053)	(0.062)	(0.064)
Fixed Effects				
Time	Yes	Yes	Yes	Yes
Origin + Destination	Yes	Yes	Yes	Yes
Origin-time + Destination-time	No	Yes	No	Yes
Adjusted/Pseudo R-squared	0.697	0.727	0.939	0.947
Number of Observations	741,343	806,645	1,625,104	1,704,299

Notes: statistically significant estimates are indicated with *** for P-value$<$ 0.01, with ** for P-value$<$ 0.05, and with * for P-value$<$ 0.1.

The results of estimating equation 5 are presented in table 2. This estimation includes all types of fixed effects, and, although still cannot be interpreted as causal, is viewed in the literature as the one accounting for endogeneity of trade agreements (Baier and Bergstrand (2007)). Similarly to the previous results, the magnitude of the coeffcient drops when the legal enforceability heterogeneity is being taken into account. The coefficient on the PTA dummy can be interpreted as around 40% increase in bilateral trade for the OLS estimation, and approximately 3.6% for the PPML estimator. In the case of the OLS estimator for the simple index, the coefficient implies that having in place a PTA with an average level of legal enforceability would result in approximately 33% higher trade flows compared to no PTA in place. The figure drops to 4.6% for the PPML estimator. For the weakest agreements in the case of PPML estimator, however, the associations are around 1,1%. Again, all the coefficients are very precisely statistically estimated, despite the presence of country-pair fixed effects.

Table 2: Estimation results: Policy Endogeneity

	OLS (1)	OLS (2)	PPML (1)	PPML (2)
PTA Dummy	0.342***		0.033***	
	(0.008)		(0.006)	
Simple Index		0.224***		0.032***
		(0.005)		(0.003)
Fixed Effects				
Time	Yes	Yes	Yes	Yes
Origin-time + Destination-time	Yes	Yes	Yes	Yes
Country-pair	Yes	Yes	Yes	Yes
Adjusted/Pseudo R-squared	0.864	0.864	0.989	0.989
Number of Observations	856,863	856,863	1,617,800	1,617,800

Notes: statistically significant estimates are indicated with *** for P-value$<$ 0.01, with ** for P-value$<$ 0.05, and with * for P-value$<$ 0.1.

The main takeaway from the gravity exercise is that representing the trade agreements as homogeneous in the form of a PTA dummy likely leads to an overestimation of the average effects of these agreements on trade flows. In addition, the dummy specification does not take into account the whole distribution of the effects, with least legally enforceable agreements having very modest association with the country-pairs' trade flows. This result reconciles the widely diverging estimates of the effects of PTAs in the previous literature.

The result is robust across a myriad of specifications, besides the 42 regression results presented in Tables 1 and 2. In fact, the result is robust for all six different specifications of the text-based measures, and for all the refinements of the data. In the appendix I present the estimates for the regressions with bilateral time-varying tariffs (see Table 4) and for unbalanced panel data (see Table 5). In addition, the appendix contains the specification built only for country-pairs which have an active agreement (see Table 6). The coefficients on those imply, in line with the categorical variable estimations, that more legally enforceable agreements are associated with higher trade flows.

5 Conclusion

In this paper I propose a new set of text-based measures that help capture the heterogeneity in legal enforceability of trade texts. The measures are robust to multiple specifications of a gravity model, and indicate that ignoring the across-PTA heterogeneity can lead to an overestimation of the average effects that these agreements have on trade volumes. In addition, the underlying distribution of potential associations is ignored when using a dummy variable. In other words, the large variation in the estimates found in the previous studies may be attributed to a pervasive measurement error. While there are many ways in which texts can help us uncover the heterogeneity across legal texts, including other types of contracts or other aspects of heterogeneity, this paper is a first approximation in doing so.

References

Wolfgang Alschner, Joost Pauwelyn, and Sergio Puig. The Data-Driven Future of International Economic Law. *Journal of International Economic Law*, 20(2):217–231, 2017a.

Wolfgang Alschner, Julia Seiermann, and Dmitriy Skougarevskiy. The Impact of the TPP on Trade Between Member Countries – A Text-as-Data Approach. *Asian Development Bank*, 2017b.

James E. Anderson and Eric van Wincoop. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192, March 2003.

Jean-François Arvis and Ben Shepherd. The poisson quasi-maximum likelihood estimator: a solution to the 'adding up' problem in gravity models. *Applied Economics Letters*, 20(6):515–519, 2013.

Scott L. Baier and Jeffrey H. Bergstrand. Do Free Trade Agreements Actually Increase Members' International Trade? *Journal of International Economics*, 71(1):72–95, March 2007.

Scott L. Baier, Jeffrey H. Bergstrand, and Michael Feng. Economic Integration Agreements and the Margins of International Trade. *Journal of International Economics*, 93(2):339–350, 2014. doi: 10.1016/j.jinteco.2014.03.

Scott Baker, Nicholas Bloom, and Steven Davis. Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016.

Marco Battaglini and Bård Harstad. The political economy of weak treaties. *Journal of Political Economy*, 128(2):544–590, 2020. doi: 10.1086/704610.

Antoni Estevadeordal and Kati Suominen. Sequencing Regional Trade Integration and Cooperation Agreements. *The World Economy*, 31(1):112–140, 2008.

Thibault Fally. Structural gravity and fixed effects. *Journal of International Economics*, 97(1):76–85, 2015.

Carsten Fink and Martín Molinuevo. East Asian Preferential Trade Agreements in Services: Liberalization Content and WTO Rules. *World Trade Review*, 7(4):641–673, 2008.

Matthew Gentzkow, Bryan T. Kelly, and Matt Taddy. Text as Data. NBER Working Papers 23276, National Bureau of Economic Research, March 2017.

Justin Grimmer and Brandon M. Stewart. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297, 2013.

Yoram Haftel. Commerce and Institutions: Trade, Scope, and the Design of Regional Economic Organizations. *The Review of International Organizations*, 8(3):389–414, September 2013.

Keith Head and Thierry Mayer. Gravity Equations: Workhorse,Toolkit, and Cookbook. volume 4, chapter 3, pages 131–195. Elsevier, 2014.

Raymond Hicks and Soo Yeon Kim. Reciprocal Trade Agreements in Asia: Credible Commitment to Trade Liberalization or Paper Tigers? *Journal of East Asian Studies*, 12(1):1–29, 2012.

Henrik Horn, Petros C. Mavroidis, and Andre Sapir. Beyond the WTO? An Anatomy of EU and US Preferential Trade Agreements. *The World Economy*, 33(11):1565–1588, 2010.

Marie-France Houde, Akshay Kolse-Patil, and Sébastien Miroudot. OECD Trade Policy Papers 55, OECD Publishing, June 2007.

Tristan Kohl. Do We Really Know that Trade Agreements Increase Trade? *Review of World Economics*, 150(3):443–469, August 2014. doi: 10.1007/s10290-014-0188-3.

Tristan Kohl. Do Trade Agreements Stimulate International Trade Differently? Evidence from 296 Trade Agreements. *The World Economy*, 39(1): 97–131, June 2015.

Jeffrey Kucik. The Domestic Politics of Institutional Design: Producer Preferences over Trade Agreement Rules. *Economics and Politics*, 24(2): 95–118, 2012.

Molly Lesher and Sébastien Miroudot. Analysis of the Economic Impact of Investment Provisions in Regional Trade Agreements. OECD Trade Policy Papers 36, OECD Publishing, July 2006.

Nuno Limão. Chapter 6 - Preferential Trade Agreements. In *Handbook of Commercial Policy*, pages 279–367. Elsevier, 2016.

Julia Seiermann. Only words? How Power in Trade Agreement Texts Affects International Trade Flows. UNCTAD Research Paper 2, United Nations Conference on Trade and Development, August 2018.

J. M. C. Santos Silva and Silvana Tenreyro. The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658, November 2006.

Yoto V. Yotov, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model. Technical report, 2016.

Appendix

Data Cleaning Notes

PTA Dataset

- Raw data contains 450 files in unbalanced `XML` format, covering agreements from 1948 to 2017, provided by UNCTAD Text of Trade Agreements (ToTA) database.
- The cleaning procedure results in 410 agreements, deleting, among others:
  - Agreements in languages other than English: 24 Spanish and 2 French;
  - Agreements under non-reciprocal concessions, with coding mistakes in the `XML` files, and amendments: 1st Convention of Lome, 2nd Convention of Lome, 3rd Convention of Lome, Generalized System of Preferences, Yaoundé Convention I, Yaoundé Convention II, EC-Syria, Arusha Agreement, EU-Overseas Territories, and Croatia-Serbia-Montenegro Agreement;
  - Agreements not yet in force as of 2017 (early announcement): EFTA-Gulf Cooperation Council, Cross-Straights Agreement between China and Taiwan, EFTA-Philippines Agreement, Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP).
- Plurilateral agreements are coded as follows:
  - As all country-pair combinations for purely plurilateral agreements (for example, for agreements such as Southern African Customs Union);
  - As a country combination with each member of a pre-existing PTA, for agreements where one party is a PTA (for example, such as ASEAN-China agreement);

- As country combinations of members of one pre-exisiting PTA with each member of another pre-existing PTA for agreements where both parties are PTAs (for example, Southern Common Market (MERCOSUR) - Southern African Customs Union Agreement).

- The country-pair coding procedure results into 4,397 country-pair observations:
  - 238 country pairs from 238 bilateral agreements, and 4,159 country pairs from 172 plurilateral agreements;
  - 3,688 country pairs from agreements in force as of 2017, and 733 country pairs from inactive agreements.
- Some country pairs overlap, since the same country-pair can have several different agreements in force (for example, Japan and Brunei are participants of ASEAN-Japan agreement, and also have a bilateral agreement of their own):
  - Most of the country-pairs have only two active agreements at the same time (around 300 country pairs), but some have up to five simultaneously active agreements;
  - There are 77,490 repeated country-pair-years, corresponding to 1,107 country-pairs and 213 agreements involving at least one country pair that has another agreement;
  - There are 230,580 distinct country-pair-years, corresponding to 3,294 country-pairs and 197 agreements where a country-pair has only one agreement

Trade Volumes Dataset

- Raw data contains value of FOB exports (in USD) by 219 origin-destinations from 1948 to 2017 – a total of 885,903 country-pair-year observations (with missing country-pair and years), provided by IMF Direction of Trade Statistics (DOTS) database.

- The grid is expanded to include all the combinations of countries and years:
  - 219 countries make 47,742 country-pair combinations, not including intra-national trade;
  - 47,742 country-pairs over 70 years (1948-2017) make 3,341,940 country-pair-year observations.
- Countries that are newly formed or seize to exist are manually recoded.
- Following the literature, the non-reported trade flows are replaced by zeros, resulting in a dataset with 2,720,148 country-pair-year observations: 47,724 country-pairs (219 countries – both existing and not existing) over 70 years (from 1948 to 2017), with 31.67% of observations containing zero trade flows.

Tables and Figures

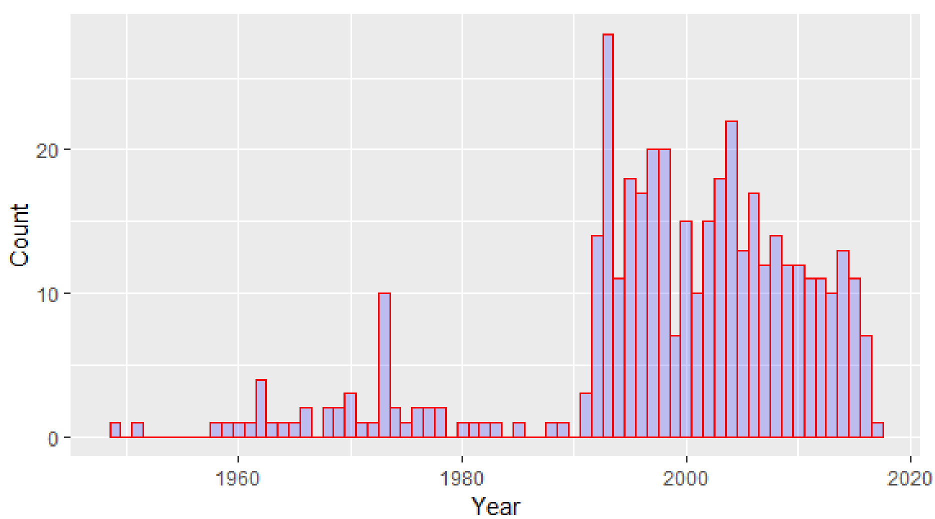

Figure 1: Number of PTAs per year, by the year of entry into force, 1948-2017

Table 1: PTAs by type, status, notification, party and regional composition

	Number of PTAs	Percentage
Type		
Economic Integration Agreement	2	0.48
Customs Union & Economic Integration	7	1.69
Customs Union	20	4.83
Free Trade Agreement & Economic Integration	122	29.47
Free Trade Agreement	248	59.90
Partial Scope Agreement	15	3.62
WTO Notification Type		
GATT Art. XXIV	252	62.07
GATS Art. V	2	0.49
GATT Art. XXIV & GATS Art. V	117	28.82
Enabling Clause	28	6.90
Enabling Clause & GATS Art. V	7	1.72
Composition		
Bilateral	236	57.28
Bilateral; All Parties are RTAs	3	0.73
Bilateral; One Party is an RTA	62	15.05
Plurilateral	102	24.76
Plurilateral; One Party is an RTA	9	2.18
Status		
In Force	247	60.24
Inactive	163	39.76
Region		
Crossregional	170	41.16
Within Region	243	58.84
Total	410	100

Table 2: Selected PTA Textual Characteristics

	Mean	Std. dev.	Min	Max
Number of chapters	10.19	8.64	1	51
Number of articles	82.5	83.91	1	489
Total words	7859.51	10070.41	448	46799
Total 'strong' words	333.36	407.76	2	1907
Total 'weak' words	109.16	131.09	3	706
Simple index	3.22	1.42	0.39	12.6
Normalized index	13.52	18.69	0.57	158.05
Composite index	11.38	12.64	2.95	111.38
Strong-to-total ratio	0.04	0.01	0.004	0.072
Difference-to-total ratio	0.03	0.01	-0.01	0.06

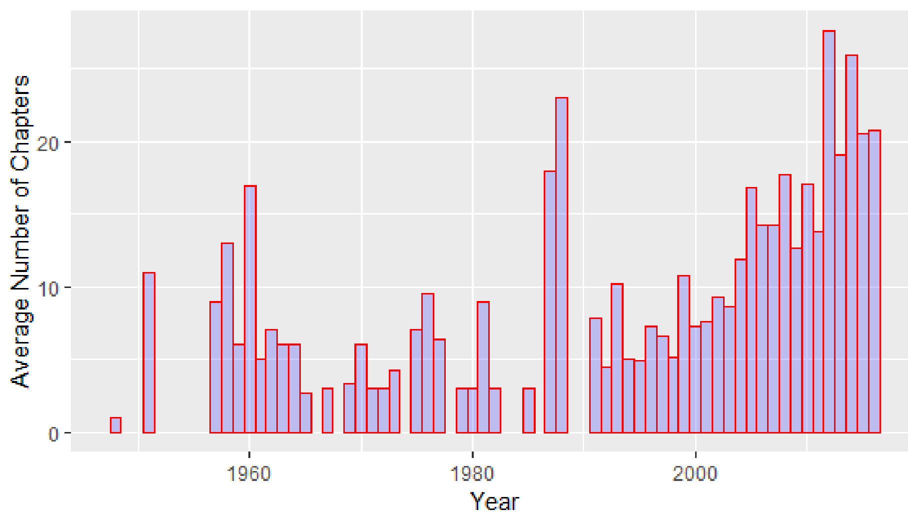

Figure 2: Average number of chapters per PTA, by the year of entry into force, 1948-2017

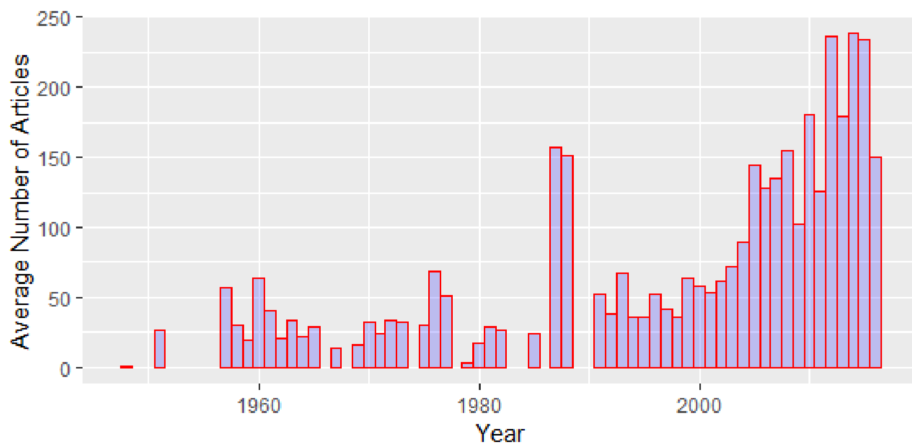

Figure 3: Average number of articles per PTA, by the year of entry into force, 1948-2017

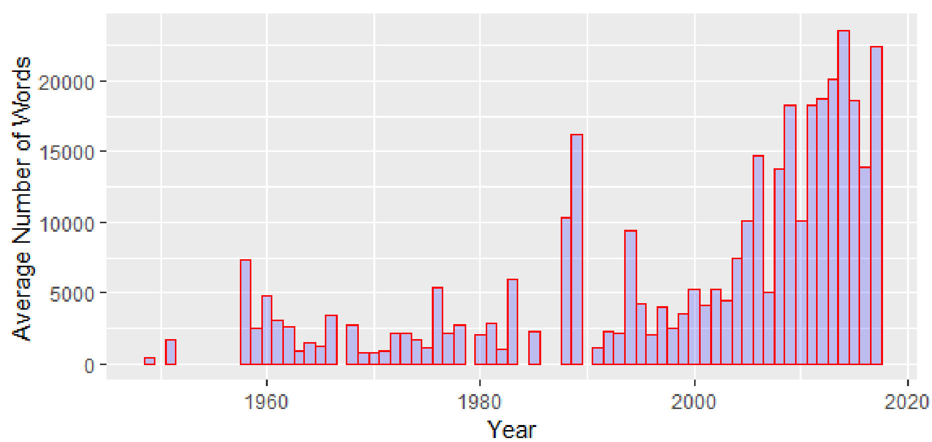

Figure 4: Average number of total words per PTA, by the year of entry into force, 1948-2017

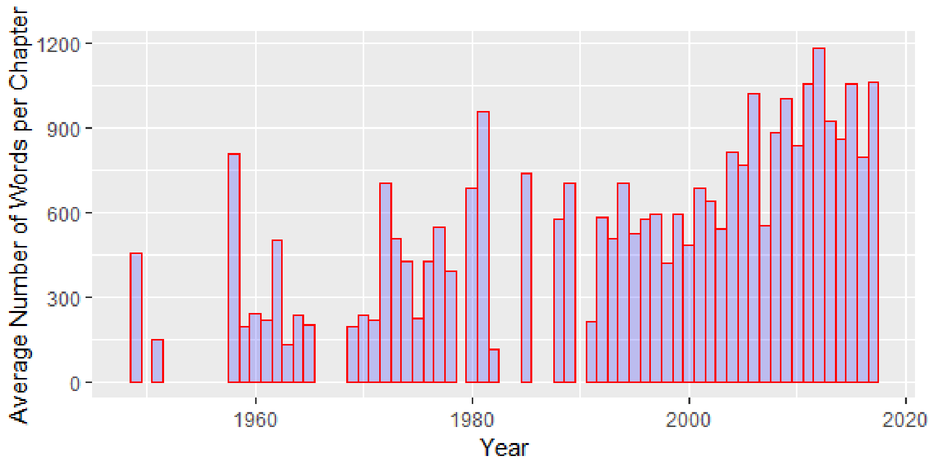

Figure 5: Average number of total words per chapter, by the year of entry into force, 1948-2017

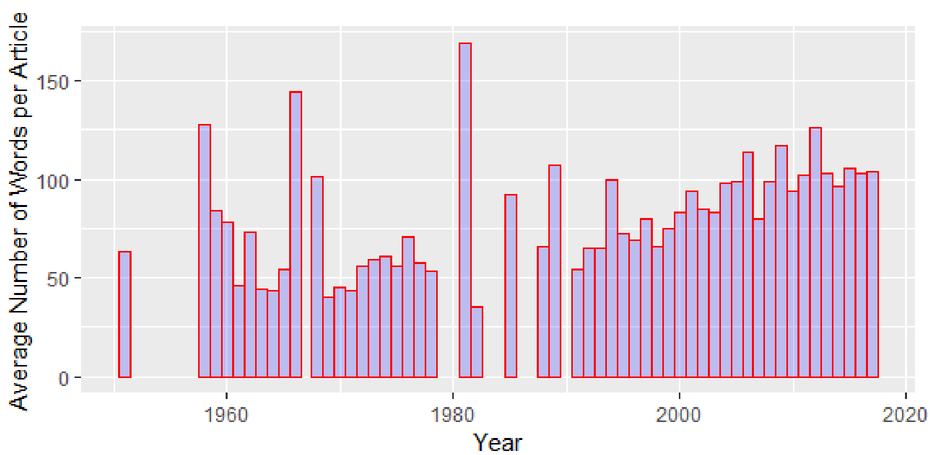

Figure 6: Average number of total words per article, by the year of entry into force, 1948-2017

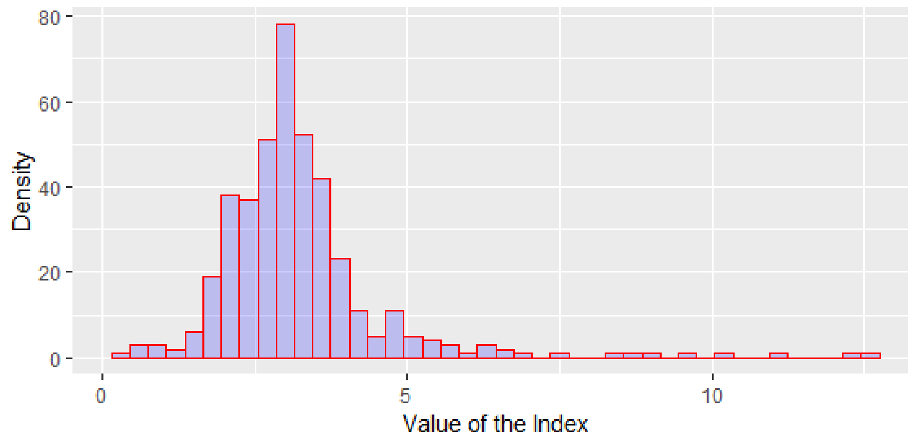

Figure 7: The distribution of simple index

Figure 8: The correlation martix of diferent textual characterristics of PTAs

Table 3: Dataset Characteristics

	Baseline Sample	Bilateral Tariffs	Unbalanced Panel	Active PTAs
Number of observations	2,840,444	316,309	2,714,196	163,944
Time span	1948-2015	1988-2015	1948-2015	1949-2015
Number of country-pairs	47,272	27,292	47,272	6,418
Number of active PTA years	163,944	34,747	125,562	163,944

Table 4: Estimation Results: Bilateral Time-Varying Tariffs

	OLS (1)	OLS (2)	PPML (1)	PPML (2)
Origin GDP	0.415***		0.754***	
	(0.034)		(0.040)	
Destination GDP	0.561***		0.674***	
	(0.030)		(0.055)	
Distance	-1.671***	-1.661***	-0.753***	-0.763***
	(0.026)	(0.026)	(0.056)	(0.055)
Contiguity	0.924***	0.905***	0.807***	0.765***
	(0.129)	(0.124)	(0.125)	(0.122)
Language	0.790***	0.787***	0.081	0.111
	(0.046)	(0.045)	(0.094)	(0.089)
Weighted Average Tariff	-0.070***	-0.067***	-0.022*	-0.024
	(0.007)	(0.007)	(0.013)	(0.016)
PTA Dummy	0.616***	0.598***	0.230**	0.249**
	(0.044)	(0.045)	(0.077)	(0.079)
Simple Index	0.432***	0.416***	0.131***	0.144**
	(0.032)	(0.032)	(0.050)	(0.051)
Normalized Index	0.287***	0.263***	0.027	0.019
	(0.021)	(0.021)	(0.031)	(0.029)
Composite Index	0.287***	0.273***	0.064*	0.064**
	(0.020)	(0.020)	(0.033)	(0.032)
Strong/Total	12.872***	12.339***	4.889**	5.497**
	(1.014)	(1.022)	(1.809)	(1.833)
Difference Index	17.730***	0.240**	5.670**	6.493***
	(1.441)	(0.081)	(2.386)	(2.413)
Fixed Effects				
Origin	Yes	Yes	Yes	Yes
Destination	Yes	Yes	Yes	Yes
Time	Yes	Yes	Yes	Yes
Origin-time	No	Yes	No	Yes
Destination-time	No	Yes	No	Yes
Adjusted R-squared	0.718	0.734	0.931	0.9398
Number of Observations	201,360	207,602	263,224	274,104

Notes: statistically significant estimates are indicated with *** for P-value< 0.01, with ** for P-value< 0.05, and with * for P-value< 0.1.

Table 5: Estimation Results: Panel Data

	OLS (1)	OLS (2)	PPML (1)	PPML (2)
Origin GDP	0.828***		0.769***	
	(0.019)		(0.021)	
Destination GDP	0.630***		0.640***	
	(0.016)		(0.041)	
Distance	-1.406***	-1.391***	-0.772***	-0.779***
	(0.021)	(0.020)	(0.038)	(0.038)
Contiguity	0.618***	0.567***	0.565***	0.539***
	(0.108)	(0.101)	(0.079)	(0.081)
Language	0.776***	0.761***	0.151**	0.161**
	(0.040)	(0.039)	(0.069)	(0.069)
PTA Dummy	0.509***	0.575***	0.310**	0.329***
	(0.035)	(0.037)	(0.064)	(0.068)
Simple Index	0.273***	0.317***	0.177***	0.186**
	(0.024)	(0.025)	(0.041)	(0.043)
Normalized Index	0.247***	0.272***	0.073**	0.063***
	(0.017)	(0.018)	(0.031)	(0.032)
Composite Index	0.243***	0.274***	0.108**	0.107**
	(0.016)	(0.017)	(0.029)	(0.031)
Strong/Total	6.710***	7.803***	5.990***	6.113**
	(0.770)	(0.797)	(1.319)	(1.371)
Difference Index	6.278***	7.820***	6.861**	6.959***
	(1.051)	(1.073)	(1.697)	(1.756)
Fixed Effects				
Origin	Yes	Yes	Yes	Yes
Destination	Yes	Yes	Yes	Yes
Time	Yes	Yes	Yes	Yes
Origin-time	No	Yes	No	Yes
Destination-time	No	Yes	No	Yes
Adjusted R-squared	0.686	0.734	0.9417	0.9487
Number of Observations	690,073	749,681	1,569,978	1,642,887

Notes: statistically significant estimates are indicated with *** for P-value$<$ 0.01, with ** for P-value$<$ 0.05, and with * for P-value$<$ 0.1.

Table 6: Estimation Results: Active PTAs

	OLS (1)	OLS (2)	PPML (1)	PPML (2)
Origin GDP	0.660***		0.676***	
	(0.046)		(0.044)	
Destination GDP	0.581***		0.631***	
	(0.043)		(0.068)	
Distance	-1.810***	-1.750***	-0.965***	-0.973***
	(0.045)	(0.047)	(0.039)	(0.040)
Contiguity	0.065	0.175	0.321***	0.306***
	(0.111)	(0.112)	(0.059)	(0.058)
Language	0.666***	0.648***	0.186**	0.190**
	(0.107)	(0.109)	(0.088)	(0.088)
Simple Index	0.950***	0.937***	0.650***	0.595***
	(0.137)	(0.139)	(0.186)	(0.184)
Normalized Index	0.498***	0.496***	0.037	-0.002
	(0.052)	(0.057)	(0.069)	(0.075)
Composite Index	0.588***	0.611***	0.072	0.046
	(0.079)	(0.086)	(0.123)	(0.132)
Strong/Total	38.611***	34.249***	17.606***	17.607***
	(4.178)	(4.742)	(5.503)	(5.503)
Difference Index	30.884***	28.839***	17.112***	15.460***
	(3.450)	(3.684)	(4.429)	(4.504)
Fixed Effects				
Origin	Yes	Yes	Yes	Yes
Destination	Yes	Yes	Yes	Yes
Time	Yes	Yes	Yes	Yes
Origin-time	No	Yes	No	Yes
Destination-time	No	Yes	No	Yes
Adjusted R-squared	0.827	0.858	0.9668	0.9675
Number of Observations	84,309	86,346	99,087	98,027

Notes: statistically significant estimates are indicated with *** for P-value < 0.01, with ** for P-value < 0.05, and with * for P-value < 0.1.