In Too Deep?

The Economic Impacts of Deep Trade Agreements

Peter R. Herman

ECONOMICS WORKING PAPER SERIES
Working Paper 2024–12–D

U.S. INTERNATIONAL TRADE COMMISSION
500 E Street SW
Washington, DC 20436

December 2024

The author would like to thank Alan Fox, Yoto Yotov, and the participants at the Spring 2024 Midwest Trade Conference for their helpful comments.

In Too Deep? The Economic Impacts of Deep Trade Agreements
Peter R. Herman
Economics Working Paper 2024–12–D
December 2024

Abstract

As preferential trade agreements (PTAs) have grown in their scope and complexity, so too has the need to capture this heterogeneity in assessments of their effects. This paper demonstrates an approach for estimating the effects of "deep" PTAs that allows for non-linear impacts from increased depth. It finds that deeper PTAs can increase trade but that there are diminishing—and eventually negative—marginal returns from adding additional policy provisions. This finding fits the observation that certain deep policies may represent new frictions to trade rather than facilitation efforts. To illustrate the potential trade and welfare gains that can be attained by increasing the depth of shallow PTAs, a series of counterfactual simulations are undertaken using the Agadir agreement between Egypt, Jordan, Morocco, and Tunisia as an example. The counterfactual analysis suggests that increasing the depth of the relatively shallow Agadir agreement could increase trade between its members by about 13 percent and the value of their real manufacturing outputs by up to 0.03 percent. Notably, the exercise demonstrates that the optimal version of an agreement is not necessarily the deepest.

Keywords: International trade, trade agreements, non-tariff measures, gravity, general equilibrium

Peter R. Herman
Research Division, Office of Economics
U.S. International Trade Commission
peter.herman@usitc.gov

1 Introduction

The contents of trade agreements have changed considerably over time. While initially focused on lowering tariffs, modern agreements have expanded dramatically and now frequently address a broad range of other policy areas, such as health and safety regulations, intellectual property rights (IPR), investment, labor, and the environment. These modern "deep" trade agreements can both achieve greater integration between members as well as erect new frictions to trade. In light of these developments, some policy makers have sought to update existing shallow trade agreements and replace them with deeper ones. The USMCA agreement (also referred to as CUSMA and T-MEC) between Canada, Mexico, and the United States is one such example. USMCA, which entered into force in 2020, replaced the NAFTA agreement from 1994 and included a large collection of new policy provisions addressing issues such as digital trade, labor conditions, and services. These new policies are expected to further integrate the economies and generate large economic gains (USITC, 2019).

This paper examines the impacts of deep trade agreements and the potential benefits of deepening older agreements. New information on the contents of preferential trade agreements from Mattoo et al. (2020) is used to produce continuous measures of PTA depth. These new measures are incorporated into a structural gravity model of trade and estimated using recent advancements in the gravity literature, such as three-way fixed effects and domestic trade flows. The gravity estimates indicate that adding additional types of provisions to a PTA generally increases its positive impacts on trade flows. However, additional provisions exhibit diminishing and—eventually—negative returns, suggesting that the deepest agreements may include provisions that inhibit rather that facilitate trade. The estimates also identify spillovers from PTA depth to non-members, indicating that some policies adopted in PTAs can impact trade with all partners. However, the spillover effects are not a complete substitute for the effects of direct membership in deep PTA.

The empirical estimates, in conjunction with a general equilibrium (GE) quantitative model of trade, are used to assess potential economic gains from modernizing existing agreements. As an illustrative example, the deepening of the Agadir agreement between Egypt, Jordan, Morocco, and Tunisia enacted in 2004 is considered. Agadir is one of the shallowest agreements in effect and has not resulted in the types of trade increases that were expected. Based on the counterfactual analysis, increasing the number of provisions in the Agadir agreement to a slightly above average number (63 provisions) would have the largest positive effects on the 4 members. Notably, the optimal trade agreement is not necessarily the deepest agreement. Under the optimal scenario, the reductions to trade costs are estimated to increase trade between the members by 12.5 to 13.5 percent. Similarly, the value of their real output is estimated to increase by up to 0.03 percent, implying welfare gains. The counterfactual analyses also use information about the variances and covariances of the econometric estimates to produce measures of standard errors of the GE model outcomes that reflect the underlying statistical precision of the estimates. Ultimately, the counterfactual demonstrates the potential trade and welfare gains that can be attained by revisiting older agreements and introducing modern, deep PTA policy provisions.

The remainder of the paper proceeds as follows. Section 2 discusses the existing literature on PTA depth and outlines the contributions of this paper. Section 3 describes the data used in the analysis and the econometric gravity approach. Section 4 presents the econometric estimates. Section 5 presents the GE impacts of deepening the Agadir agreement. Finally, section 6 concludes.

2 Background

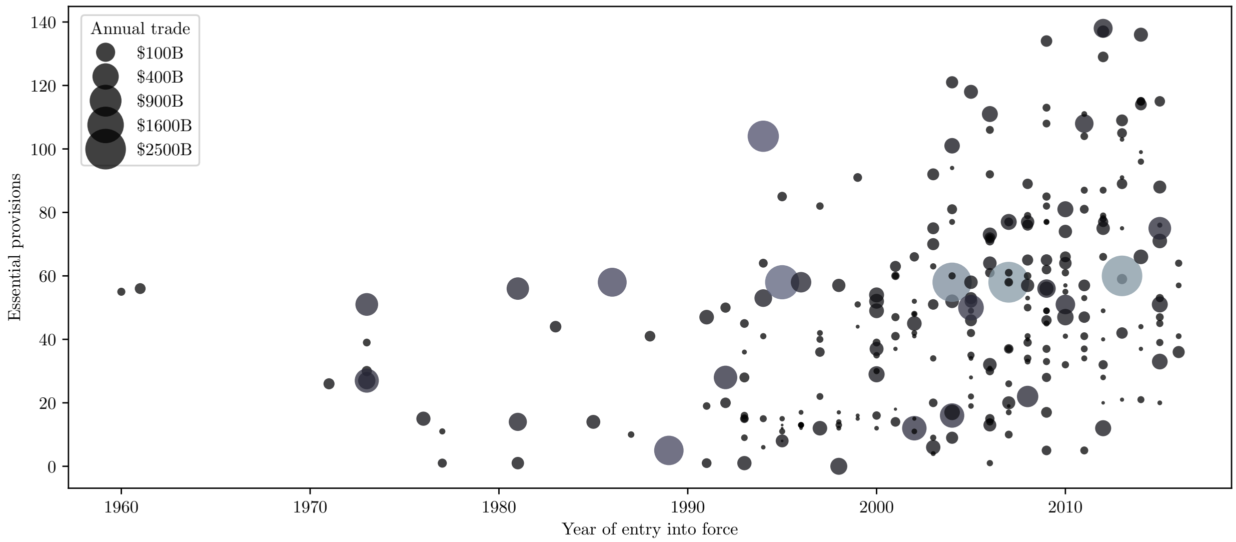
As the number of trade agreements in effect has grown over time, so too has their contents. In addition to lowering tariffs, PTAs now also frequently address a wide range of other policy areas. Based on data from Mattoo et al. (2020), Figure 1 plots the depth of 274 PTAs as they entered into force. As is clear from the plot, the scope of many PTAs has expanded considerably over the last 50 years. Prior to 1994, the deepest agreements featured fewer than 60 "essential" provisions.[1] However, in more recent years, most agreements have featured more than 60 provisions and some have up to 138. Despite this general upward trend, there also remains significant differences in depth across the PTAs enacted within each year. Given this heterogeneity, it is increasingly important to account for the wide ranging differences between agreements.

The impacts of trade agreements have been studied extensively throughout the literature, utilizing a variety of strategies to identify their effects (see Larch and Yotov, 2024).[2] PTAs have most often been examined using 0/1 indicators that reflect the presence of an agreement (Baier and Bergstrand, 2007; Bergstrand et al., 2015). While these types of studies have consistently identified robust positive effects of trade agreements, the estimates have generally represented average impacts across all PTAs in effect and have not considered heterogeneity across PTAs. Some studies have found ways to estimate heterogeneous effects from indicator variables in a few different ways. For example, some work has used categorizations of PTAs into several broad types based on their scope, such as free trade agreements, nonreciprocal agreements, or customs unions. These approaches have generally found that the deeper categories of PTAs have larger trade effects (Magee, 2008; Roy, 2010; Baier et al., 2014). In a similar vein, Anderson and Yotov (2016) categorized PTAs depending on whether they were formed between countries with high or low MFN tariffs and found that

[1] "Essential" provisions, which are described in more detail in section 3.1, refer to a subset of the most important provisions in trade agreements as defined by Mattoo et al. (2020).

[2] Throughout the literature, the term "preferential trade agreement" has regularly taken on different meanings. In some cases it is used to denote a specific type of trade agreement (typically an agreement cutting fewer tariffs than a more comprehensive FTA). In others, it is use as a general descriptor to refer to all different categories of trade agreements that grant preferential treatment. In this paper, I follow the latter meaning.

Figure 1: Number of provisions, date of entry into force, and volume of trade covered by each trade agreement



Note: This figure depicts the number of essential provisions in and date of entry for 274 preferential trade agreements. The size and shade of each marker indicates the volume of trade under the agreement in 2016 (or 2015 when 2016 data was unavailable).

the impacts of PTAs were much higher for countries with high tariffs. Other research has considered PTA heterogeneity by treating each PTA separately and producing an individual estimate for each one, thereby inferring the differences across PTAs based on their impacts (Magee, 2008, Baier et al., 2019).

Many recent studies have looked more directly at the contents of PTAs to determine their depth. Much of this work has been supported by the release of extensive databases cataloging the content of trade agreements, such as the work of Horn et al. (2010), the Horizontal Depth database (Claudia et al., 2017), the World Bank Deep Trade Agreement database (Mattoo et al., 2020), and the DESTA database (Dür et al., 2014). These databases consist of a list of many different types of policy areas that may be covered by an agreement and indicate whether country-pairs belong to a trade agreement with that coverage. Common deep policy areas include non-tariff measures (NTM) like sanitary and phytosanitary measures (SPS) or technical barriers to trade (TBT); intellectual property rights (IPR); investment; rules of origin (RoOs); and services trade.

Using these types of trade agreement data, numerous studies have examined the impacts of PTAs and their varying levels of depth. This work has overwhelmingly found that deeper trade agreements have a larger positive impact on trade than shallower agreements (c.f. Kohl et al., 2016; Mattoo et al., 2022; Lee et al., 2023). This amplified effect of deeper PTAs appears to also hold for a variety of different specific types of trade, including supply chains (Laget et al., 2020; Orefice and Rocha, 2014), foreign direct investment (Osnago et al., 2017), and exports of state owned enterprises (Lefebvre et al., 2023). In most of this work, indexes reflecting the number of policy areas or provisions present in a PTA are estimated using a gravity (or gravity-inspired) empirical framework. The findings suggest that the impact of an agreement on trade flows is generally increasing in the number of policy areas covered.

The recent wealth of data on the contents of PTAs has solved some empirical challenges but introduced others. One significant challenge is how best to represent depth in an empirical model. The number of provisions cataloged in trade agreements range from dozens in some data sources to hundreds in others, making it difficult to consider each provision individually. Additionally, many of the provisions appearing in trade agreements are highly correlated, further complicating the identification of the impacts of individual provisions. For example, trade agreements covering SPS issues tend to also cover TBT issues, making it difficult to separate the effects of each category of provisions. Given these challenges, it is generally necessary to simplify the data into a more econometrically feasible collection of variables. As discussed before, the most common approach has been to construct indices of depth reflecting a count of the number of provisions appearing in each agreement. However, alternative other approaches have been arising as well.

To overcome some of thee empirical issues, statistical and machine learning methods have been proposed as a means to generate new measures of PTA depth for use in econometric models. One such method is principal component analysis (PCA), which is used to identify and summarize the most impactful provisions appearing in PTAs (Laget et al., 2020; Orefice and Rocha, 2014). Other studies have turned to machine learning techniques to similarly identify the set of provisions that are most important for explaining the impacts of deep PTAs (Breinlich et al., 2021; Baier and Regmi, 2023; Fontagné et al., 2023). In both cases, the methods allow the analysis to focus on the aspects of PTAs that appear to matter most and ignore the components that aren't very impactful. A key advantage of doing so is that it allows for the possibility of heterogeneous impacts across different provisions. A potential drawback, however, is that these methods do not necessarily identify the specific provisions having this impact. They may simply be identifying provisions that are closely correlated with the ones that matter most and offer the best fit of the data. Depending on the circumstance, this drawback could be limiting if the machine-identified provisions do not reflect those of interest in a particular application, such as the ex ante evaluation of a newly proposed PTA.

The effects of provisions can differ not only in the magnitude of their effects but also in their direction. Some provisions, such as the streamlining of customs procedures or the harmonization of regulations, ostensibly seek to reduce barriers and facilitate trade. However, other provisions may represent new frictions, such as SPS and TBT certification requirements, labor standards, or rules of origin. In these latter cases, new requirements or standards may increase the costs of trade or bar it entirely. In many cases, there may simultaneously be both trade facilitating and trade deterring effects. For example, SPS provisions may result in new requirements and higher trade costs but also increased demand if food imports are viewed as safer or higher quality as a result of the new policies (Xiong and Beghin, 2014). The effects of provisions may also depend on the specific products being traded. Provisions may reduce barriers or increase demand for certain products at the expense of others. For example, IPR-strengthening provisions likely enhance trade in the products covered by the IPR provisions and diminish it among the products not covered (e.g. patented vs generic pharmaceuticals). For these reasons, the usual treatment of depth as monotonically trade facilitating may overlook these more complex impacts of deep PTAs.

Another important feature of many deep PTA provisions is that they can be (explicitly or de facto) nondiscriminatory. In these cases, the provisions may apply not only to PTA members but also nonmembers. For example, digital trade provisions guaranteeing the free flow of data across borders generally apply universally rather than exclusively to data transfers between members.[3] The commitments made under the USMCA agreement to not introduce any data localization policies (and eliminate any that existed), which effectively apply to data from any country, are one such example (USITC, 2019). Some recent studies have attempted to identify these types of spillover effects from deep PTAs. The findings have generally supported the notion that deep PTA agreements do tend to benefit and increase trade with nonmembers in addition to members. Interestingly, the specific sources of these benefits appear to be multifaceted, including nondiscriminatory policies (Mattoo et al., 2022; Lee et al., 2023), regulatory harmonization Lee et al. (2023), supply chain strengthening (Laget et al., 2020), and impacts on relative firm competitiveness (Lefebvre et al., 2023). In each case, deep PTAs often make it easier for firms in nonmember countries to trade with the members of the PTA.

This paper contributes to the literature in several ways. First, it proposes a continuous measure of PTA membership that is intended to capture novel types of heterogeneity across PTAs. Specifically, it considers explicitly the possibility of diminishing—and even negative—returns to increased depth, relaxing the common assumption that the marginal impacts of deep provisions are constant and positive. Second, it uses domestic trade flows to identify third-party spillovers from deep agreements for both importers and exports, providing new evidence that deep PTAs can be nondiscriminatory. Notably, the estimates indicate that the spillover effects on exports may not be the same as on imports. Third, it uses the econometric estimates in conjunction with a GE extension of the gravity model to evaluate the potential trade and welfare gains from replacing a shallow agreement with a deeper version. This analysis highlights the novel non-monotonic effects of depth and demonstrates a way of producing econometrically-derived standard errors as a part of counterfactual, ex ante policy simulations.

The paper is most closely related to the work of Fontagné et al. (2023), who use a similar gravity approach to empirically estimate the effects of PTA depth and simulate the GE impacts of hypothetical counterfactual

[3] This is likely due at least in part to the ease of routing data and the challenge of tracking its origins, which would make it difficult to enforce a discriminatory version of the provision.

scenarios. The present work differs in several important ways. First, it estimates non-linear effects of PTA depth rather than using machine learning methods to classify PTAs into one of three abstract clusters. In doing so, the present work allows for increased depth to deter trade. Second, the ability to capture diminishing returns to depth allows for the identification of optimal vs sub-optimal levels of depth in the GE simulations. Third, I carry through the statistical imprecision from the econometric model to the GE model and produce counterfactual outcomes with computed standard errors. While the ultimate outcomes of the paper are consistent with most of the previous literature, the approach presented here offers several useful new contributions. In particular, it presents novel ways of conducting ex ante analysis of hypothetical new PTAs that may be less readily examined via other established approaches.

3 Measuring PTA Depth and Identifying its Impacts

A gravity model of trade is used to econometrically estimate the impacts of deep trade agreements on international trade flows. The gravity model is a workhorse structural model that has been used for decades to estimate the determinants of trade (Yotov et al., 2016). The econometric version of model takes the following general form.[4]

$$X_{ijt} = \exp\left\{\alpha DTA_{ijt} + \sum_{k} \beta_k z_{k,ijt} + \mu_{it} + \nu_{jt} + \rho_{ij} + FRGN_{ij} \times \psi_t \right\} + \epsilon_{ijt} \tag{1}$$

X_{ijt} denotes the value of exports from country i to country j in year t. DTA_{ijt} is a measure of trade agreement depth between the two countries. z_{ij} represents a collection of other time-varying bilateral controls indexed by k that are included to aid in the identification of the PTA depth effects. μ_{it}, ν_{jt}, and ρ_{ij} are exporter-year, importer-year, and exporter-importer fixed effects. Importantly, the exporter-year and importer-year fixed effects control for country-level multilateral resistances (Anderson and van Wincoop, 2003; Feenstra, 2002). The exporter-importer fixed effects capture all non time-varying country-pair factors, which helps to reduce omitted variable biases and mitigate endogeneity concerns (Baier and Bergstrand, 2007). Finally, $FRGN_{ij} \times \psi_t$ denotes the interaction between a set of year fixed effects (ψ_t) and an indicator for foreign trade ($FRGN_{ij}$), which takes the value of one if the trade flow is international ($i \neq j$) and zero if it is domestic ($i = j$). These interaction terms capture changes in average foreign trade costs each year and help disentangle the effects of individual PTAs from general global trends in trade openness (Bergstrand et al., 2015).

[4] The theoretical foundations of the model are discussed in more detail in section 5.1.

Table 1: Policy areas of the 305 essential provisions in trade agreements

Policy Area	Provisions	Policy Area	Provisions
Anti-dumping	11	Movement of Capital	8
Competition Policy	14	Public Procurement	5
Countervailing	0	Rules of Origin	19
Environment	27	Sanitary and Phytosanitary	24
Export Taxes	23	State Owned Enterprises	13
Intellectual Property Rights	67	Services	21
Investment	15	Subsidies	13
Labor Market	12	Technical Barriers to Trade	19
Migration	3	Trade Facilitation	11

3.1 Data

To measure the depth of trade agreements, data describing the content of PTAs was sourced from the *Deep Trade Agreements Database 2.0* of Mattoo et al. (2020). The database catalogs the contents of 283 agreements across 18 commonly addressed policy areas, such as competition policy, intellectual property rights, rules of origin, and trade facilitation. In total it contains information on more than 900 potential features of a PTA. For the analysis, the full set of provisions was narrowed down to a core set of 305 "essential" provisions that are considered integral in achieving the objectives of the agreement, as characterized by the database authors.[5] These essential provisions, which are categorized in Table 1, reflect a diverse range of topics covering all but one of the 18 broad policy areas. Narrowing the set of provisions helps to reduce the dimensionality of the data and mitigate concerns about high correlation across the full set of provisions.

Each individual measure reflects a potential attribute or feature of a trade agreement. The following are illustrative examples of some of the most common essential provisions across several of the broad categories:[6]

1. **Rules of origin**: Does the agreement contain a transhipment rule?
2. **Trade facilitation**: Does the agreement require proof of origin?
3. **State owned enterprises**: Does the agreement prohibit anti-competitive behaviour of state enterprises?
4. **SPS**: Does the agreement refer to the World Trade Organization (WTO) SPS Agreement?
5. **Competition policy**: Does the agreement regulate monopolies?

[5] As described by the authors, essential provisions comprise "the set of substantive provisions plus the disciplines among procedures, transparency, enforcement, or objectives which are viewed as indispensable and complementary to achieving the substantive commitments. Non-essential provisions are referred to as 'corollary.' A caveat is that this exercise is based on the experts' knowledge and, hence, is subjective. However, this approach has the advantage of limiting the dimensionality of the data in an informed way." (Mattoo et al., 2020, p. 12)

[6] The descriptions are lightly edited in some cases to increase clarity. These measures correspond to the following measure codes in the database, respectively: "*Rules of Origin - roo_trs*", "*Trade Facilitation - prov_44*", "*STE - prov_35*", "*SPS - prov_02*", "*Competition Policy - prov_18*", and "*Services - dis_nt*".

6. **Services**: Is there a national treatment obligation?

Importantly, PTA provisions reflect a combination of both trade facilitation efforts and regulations that may restrict certain types of trade. Provisions that introduce rules of origin or establish new regulatory requirements, for example, may—intentionally or unintentionally—limit certain types of trade. Others, such as national treatment obligations, may be primarily trade facilitating. Many types of provisions may have mixed effects. For example, protections for IPRs may be trade facilitating for the firms or products with protected patents and trade decreasing for those without. Similarly, SPS regulations on food safety may raise the cost of trade and production while simultaneously increasing demand for imported food products. Thus, it may not necessarily be the case that increasing PTA depth by expanding its policy coverage will enhance trade between members.

To estimate the effects of trade agreement depth on international trade, the essential policy measures were used to define depth measures. In general, the essential measures are binary indicators, taking a value of 1 if the agreement addresses the policy measure in question and a 0 if it does not. In a few select cases, the measures are not binary, taking on either a categorical value (e.g. a multiple choice question about how a policy is administered) or a continuous value (e.g. the length of time a certification is valid). In these cases, the measures were converted to binary indicators reflecting whether the agreement addressed the issue at all and ignoring the specifics of how it was addressed.

The 305 indicator variables were used to produce several consolidated measures of agreement depth. The first measure is a count of the number of essential provisions in each agreement. In principle, this measure ranges between 0 and 305. In practice, the deepest agreement by this metric, the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP), has 167 essential provisions. Across all 283 agreements, the mean (median) number of provisions is 49.7 (47.0) with a standard deviation of 32.0 provisions.

As an alternative measure of agreement depth, the count of essential provisions was used to divide agreements into three equal groups: high, medium, and low depth. The high depth group includes provisions in the top 33rd percentile (59–167 provisions) while the low and medium depth groups reflect the lowest (0–33) and middle (34–58) 33rd percentiles, respectively. Compared to the continuous count measures, which estimates the marginal impact of each additional provisions, this grouping allows for the identification of more general trends in overall depth.

The measures of PTA depth were combined with additional data from multiple sources. Bilateral trade flows were sourced from *Structural Gravity Manufacturing Database* (SGMD) of Monteiro (2020). This data contains both positive and zero-valued trade flows (country pairs that do not trade in a given year) for 186

Table 2: Description of the data used in the gravity estimations

Variable	Type	Mean	Std. Dev.	Minimum	25%	Median	75%	Maximum
Trade ($M)	Continuous	743.83	40,724.04	0.00	0.00	0.20	6.65	13,758,384.57
PTA	Indicator	0.15	0.35	0	0	0	0	1
Essential provisions	Count	4.69	15.50	0	0	0	0	138
Low depth	Indicator	0.09	0.28	0	0	0	0	1
Medium depth	Indicator	0.03	0.18	0	0	0	0	1
High depth	Indicator	0.02	0.15	0	0	0	0	1
GATT membership	Indicator	0.07	0.26	0	0	0	0	1
WTO membership	Indicator	0.41	0.49	0	0	0	1	1
EU membership	Indicator	0.01	0.11	0	0	0	0	1
Foreign	Indicator	1.00	0.06	0	1	1	1	1

Note: This table presents summary information about the data sample used in the gravity model estimations. "Std. Dev." denotes the standard deviation. "25%" and "75%" denote the twenty-fifth and seventy-fifth percentiles, respectively.

trading partners between 1980 and 2016. Importantly, it also contains information on both international and domestic (intra-national) trade. While only covering manufacturing trade, this database was chosen for its relatively long time coverage compared to other sources with domestic flows. In addition to the trade data, other bilateral controls for European Union (EU), WTO, and General Agreement on Tariffs and Trade (GATT) membership were sourced from the *Dynamic Gravity Dataset* of Gurevich and Herman (2018).[7] Table 2 provides a brief summary of the compiled data.[8]

The bilateralization of the PTA data presents several complications. First, some pairs of countries are members of more than one shared active agreement in a given year. In these cases, the deepest active agreement between the countries in each year was used. Second, the *Deep Trade Agreements Database 2.0* used to produce the depth measures is not completely comprehensive and lacks information on some agreements, especially if they are no longer in force. Information from the *Dynamic Gravity Dataset* was used to identify these cases and all observations corresponding to members of a PTA missing from the PTA depth data were omitted from the analysis.

Using this data, the model described by Equation 1 was estimated using Poisson Psuedo Maximum Likelihood (PPML). As demonstrated by Santos Silva and Tenreyro (2006), PPML offers several advantages for gravity analysis. First, it provides superior treatment of heteroskedasticity, which is known to be a concern with this type of trade analysis. Second, it allows for the inclusion of zero-valued trade flows. In the present sample, zero trade flows are included and represent about 21 percent of trade flows.

[7] GATT membership covers the sample period through 1994. WTO membership covers the years 1995 and after. Outside of these respective time periods, both indicators are always equal to zero.

[8] The CPTPP agreement was enacted in 2017 and is therefore outside the sample period. Thus, the maximum number of essential provisions in the estimating sample is 138, reflecting the second deepest agreement (Korea–United States).

Table 3: Estimated impacts of PTA depth on international trade

	(1)	(2)	(3)	(4)	(5)
PTA	0.117***		0.0951	0.00219	0.00316
	(0.0396)		(0.0598)	(0.0765)	(0.0741)
EU	0.279***	0.270***	0.280***	0.256***	0.253***
	(0.0725)	(0.0743)	(0.0726)	(0.0702)	(0.0692)
WTO	0.155***	0.188***	0.157***	0.156***	0.148***
	(0.0374)	(0.0367)	(0.0377)	(0.0375)	(0.0379)
GATT	0.601***	0.632***	0.603***	0.601***	0.597***
	(0.0713)	(0.0704)	(0.0715)	(0.0718)	(0.0723)
High depth		0.185***			
		(0.0432)			
Medium depth		0.0308			
		(0.0400)			
Low depth		0.0446			
		(0.0582)			
Essential provisions			0.000283	0.00521**	0.00499**
			(0.000734)	(0.00205)	(0.00201)
Essential provisions^2				-0.0000386***	-0.0000389***
				(0.0000140)	(0.0000137)
Exporter maximum provisions					-0.0350***
					(0.00797)
Importer maximum provisions					0.0361***
					(0.00802)
N	875151	875151	875151	875151	875151
Pseudo R^2	0.998	0.998	0.998	0.998	0.998
AIC	18465305.1	18341827.4	18464517.2	18432450.2	18412438.7

This table presents the results of a series of gravity models of trade for the period 1980–2016. Each specification also included exporter-year, importer-year, exporter-importer, and border-year fixed effects, which are omitted for brevity. Standard errors are clustered at the exporter-importer level and reported in parentheses. p<0.10 *, p<0.05 **, and p<0.01 ***.

4 Estimated Impacts of PTA Depth on Trade Flows

The gravity estimates indicate that PTA depth has a significant impact on trade. Table 3 presents the results from a series of specifications using the data described above. All specifications in this table included exporter-year, importer-year, exporter-importer, and foreign-year fixed effects. Column (1) reflects a traditional model containing a simple indicator for the presence of a PTA along with controls for EU, WTO, and GATT membership. The estimated coefficient (0.117) is reflective of other estimates in the literature and provides some validation of the data sample. Similarly, the EU, WTO, and GATT controls exhibit positive effects as expected.

Column (2) introduces the first measures of PTA depth: the three indicators for a high, medium, or low depth PTA. Of the three categories of PTAs, only high depth agreements appear to have a significant positive impact on trade. This suggests that only the deepest agreements have consistently increased bilateral trade during the sample period.

The next specifications introduce the continuous measure of PTA depth that counts the number of essential provisions in each agreement. Column (3) replaces the high, medium, and low depth indicators with the continuous measure. While the estimate is positive, it is not significant at conventional levels. The depth estimates in columns (2) and (3) suggest that the impacts of provisions may be more complex than can be captured by a simple count that implicitly assumes a constant marginal effect for each additional provision.

Based on this observation, column (4) adds an additional layer of complexity to PTA depth in the form of a quadratic term for the essential provisions (i.e. the count of provisions squared). This term allows for non-linearity in the effects of provisions, such as increasing or diminishing marginal effects of additional provisions. The corresponding estimates, which are statistically significant, suggest there is merit to this approach.[9] The inclusion of the quadratic term results in a positive and statistically significant impact of additional provisions. The quadratic term is negative and significant, suggesting that the marginal benefit of additional provisions is diminishing as PTAs become deeper. As part of their recent survey, Larch and Yotov (2024) briefly test a comparable quadratic approach for assessing PTA depth and find similar effects. The estimates for the three other policy controls—EU, WTO, and GATT membership—are largely unaffected by the inclusion of the depth measures, suggesting that PTA depth represents different dimensions of economic integration than are captured by those terms. Finally, the indicator for PTA is not significant when the continuous depth measures are included. This implies that the depth measures fully capture the influences of PTAs on trade and dominate the traditional indicator.

In many cases, policies enacted in trade agreements can have "nondiscriminatory" impacts, effectively applying to both members and non-members of a trade agreement. To attempt to capture these types of spillovers from deep agreements, the specification in column (5) adds two additional terms. These terms represent the level of depth of the deepest PTA that the exporter and importer are members of in each year, respectively. To avoid collinearity with the country-year fixed effects and identify the effects of depth spillovers on foreign trade, both measures are set equal to zero for domestic trade flows. The estimates suggest that there are spillovers but that they are not the same for exporters and importers. Belonging to a deep agreement in general tends to decrease bilateral exports but increase bilateral imports. A possible explanation is that deep trade agreements tend to narrow where a country sends its exports, taking advantage of deep preferential market access. Meanwhile, import patterns suggest evidence of increased non-discriminatory openness, as hypothesized and observed throughout the literature. These findings are largely consistent with those of Herman (2021), which found that countries generally tend to export narrowly and import widely, and perhaps offers an explanation for why this is the case. Importantly, the estimated effects of the

[9] It is worth noting this treatment of depth is atheoretical and this specific quadratic form may not be the optimal one. Nonetheless, it produces useful empirical insights.

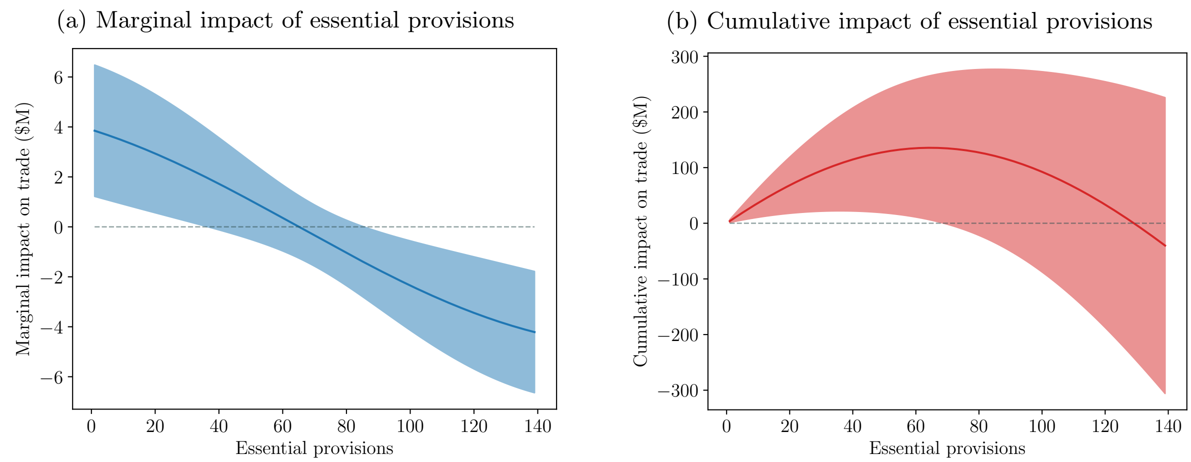


Figure 2: Impact of essential provisions in agreements on trade

bilateral essential provisions, both linear and quadratic, remain significant in column (5) and are of similar magnitudes with the addition of the spillover effects. This suggests that spillover depth from other PTAs is not a substitute for the provisions in an agreement to which both parties are members.

To better understand the impacts of PTA depth, Figure 2 plots the estimated marginal effects of the 1st essential provision through the 138th based on the estimates in column (5) of Table 3. The shaded areas reflect standard errors. The marginal effects are calculated at the mean values of all the other regression variables. Panel (a) depicts the marginal effects of each provision. The first provision increases trade by about $3.84 million dollars but decreases with each additional provision. By the 66th provision, the marginal impacts are negative, suggesting that PTAs with 65 provisions provide the highest level of trade facilitation, on average. Panel (b) of Figure 2 presents the cumulative impact on trade. At 65 provisions, the cumulative impact is a total average increase in bilateral trade of about $135.6 million. Interestingly, the estimates suggest that agreements with 130 or more provisions tend to actually decrease trade on average.[10] It is important to note, however, that these marginal impact estimates do not necessarily capture the full impacts of increasing PTA depth. Increasing the number of provisions in an agreement would also affect the multilateral resistances of the importer and exporter in a way that is not identified in the model. While the multilateral resistances are controlled for with the importer-year and exporter-year fixed effects, the relationship between them and PTA depth is not identified. As such, these marginal effects should be considered "partial" effects, representing a large share of the impact but not all of it. The full effects of PTA depth are considered in section 5.

[10]It should be noted that 130 provisions is near the end of the distribution of PTA depths in the sample. These high-provision estimates should be considered with some caution as there are relatively few agreements with this level of depth to infer from.

Table 4: Robustness tests of main specification

	(1) OLS	(2) 3-yr interval	(3) 5-yr interval	(4) Pair trend	(5) Bias correction
PTA	0.264***	0.00222	0.0635	0.00165	0.00294
	(0.0472)	(0.0785)	(0.0819)	(0.0742)	(0.0826)
EU	0.892***	0.169***	0.493***	0.224***	0.239**
	(0.0925)	(0.0507)	(0.131)	(0.0680)	(0.103)
WTO	0.0846**	0.150***	0.00414	0.150***	0.128***
	(0.0406)	(0.0402)	(0.0520)	(0.0379)	(0.0379)
GATT	0.0493	0.552***	0.641***	0.604***	0.573***
	(0.0476)	(0.0669)	(0.0881)	(0.0715)	(0.0728)
Essential provisions	-0.00292	0.00478**	0.00290	0.00498**	0.00503**
	(0.00179)	(0.00210)	(0.00227)	(0.00201)	(0.00227)
Essential provisions^2	0.00000284	-0.0000365**	-0.0000259*	-0.0000389***	-0.0000402**
	(0.0000148)	(0.0000145)	(0.0000150)	(0.0000137)	(0.0000171)
Exporter maximum provisions	-0.0132*	0.000371	0.00178**	-0.0351***	-0.0345***
	(0.00713)	(0.000726)	(0.000834)	(0.00800)	(0.0113)
Importer maximum provisions	0.0218***	[omitted]	[omitted]	0.0362***	0.0360***
	(0.00752)			(0.00805)	(0.0114)
N	729286	277633	181258	875151	875151
Adjusted R^2	0.846				
Pseudo R^2		0.998	0.998	0.998	0.998
AIC	2678097.3	5676220.2	3832472.6	18370372.0	18412438.7

Note: This table presents the results of a series of gravity models of trade for the period 1980–2016. Each specification also included exporter-year, importer-year, exporter-importer, and border-year fixed effects, unless otherwise specified. Standard errors were clustered at the exporter-importer level and are reported in parentheses. Column (1) was estimated using OLS and logged trade values. Column (2) was estimated using 3-year intervals. Column (3) was estimated using 5-year intervals. Column (4) included linear exporter-importer trends instead of exporter-importer fixed effects. Column (5) employs the PPML bias correction of Weidner and Zylkin (2021). The "importer maximum provisions" term was omitted in columns (2) and (3) due to collinearities. Standard errors clustered at the country pair level reported in parentheses. $p<0.10$ *, $p<0.05$ **, and $p<0.01$ ***.

4.1 Robustness tests

To test the robustness of the main findings, several additional specifications were considered. These tests, which are presented in Table 4, examine alternative estimators and specifications. In all cases, the tests are based on specification (5) from Table 3 and use the same selection of variables and controls unless otherwise specified. The first test presented in column (1) of Table 4 uses an ordinary least squares (OLS) estimator and logged trade flows as an alternative to PPML. Many of the estimates are similar in sign under OLS and often in magnitude as well. The biggest difference is that the effects of PTAs appear to be captured by the 0/1 indicator variable and not the terms based on essential provisions, which are both statistically insignificant. However, there are concerns about biases under OLS that are the likely root of these differences so the PPML estimates should be preferred.

Columns (2) and (3) estimate the model using interval data, reflecting 3- and 5-year intervals, respectively. Some past literature has suggested that the effects of trade policies like those present in PTAs take time to go into effect and that it is best to estimate their impacts with a delay (c.f. Treffler, 2004). With interval data, the estimated effects of essential provisions are consistent with the main specification. Additional provisions increase trade but exhibit a diminishing marginal impact. The use of 3-yr interval data appears to slightly reduce the magnitude of these effects while the use of 5-yr data more significantly reduces them, even resulting in insignificant effects for the linear term in the latter case. The use of interval data presents some challenges for the estimation of the spillover terms. In both cases, the importer maximum provision term had to be dropped due to insufficient variation in the interval data. The exporter maximum provision term, which was able to be included, reverses sign but is only significant when using the much more limited 5-yr interval data. As with the OLS specification, it is not clear that the interval data represents a better approach than the preferred specification using all available years.

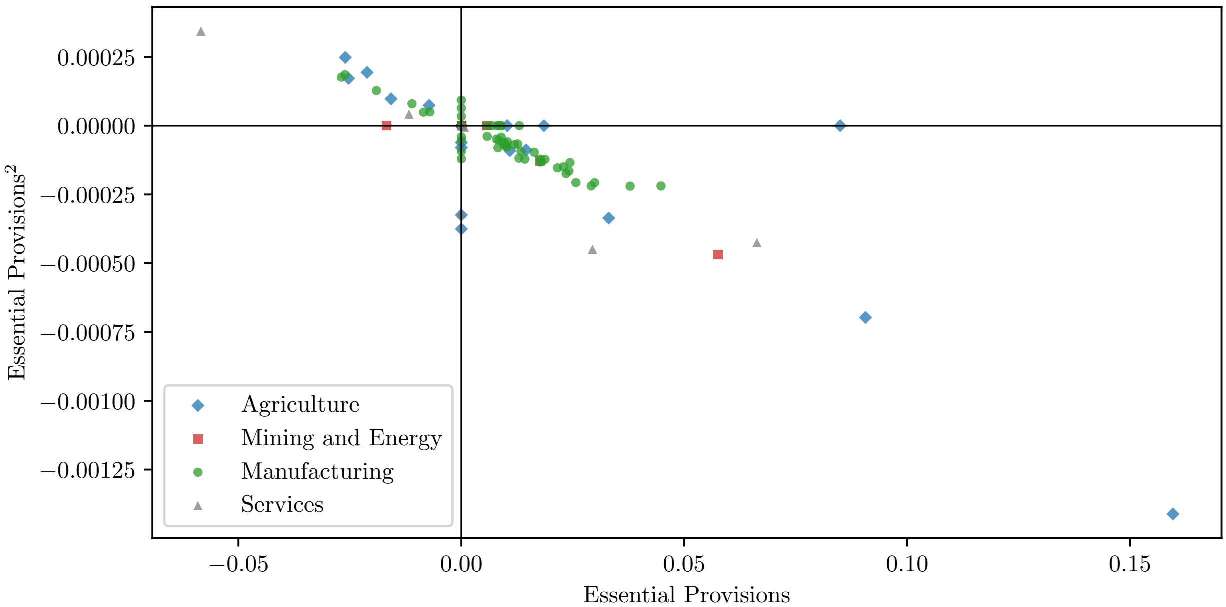
Columns (4) and (5) consider two alternative technical approaches from the recent gravity literature. Following the work of Bergstrand et al. (2015), column (4) replaces the exporter-importer fixed effects with exporter-importer linear trends. This change allows for additional variation in these country-pair controls over time and may further mitigate endogeneity concerns. The estimates with the linear trends are generally consistent with those without. In the case of the two essential provision terms and two spillover terms of interest, the estimates are almost identical. Column (5) introduces a bias correction for gravity models developed by Weidner and Zylkin (2021). There is a concern that "three-way" gravity models with exporter-year, importer-year, and exporter-importer fixed effects may be prone to biases arising from the incidental parameter problem. Weidner and Zylkin's approach corrects for these issues. The resulting estimates in column (5) are again largely consistent with the original estimates. The two essential provision terms are slightly larger in magnitude while the spillover terms are slightly smaller, but the differences are very minor. Meanwhile, the standard errors of these estimates are larger with the bias correction, as expected, but all remain significant at conventional levels.

Taken together, these tests demonstrate that the main findings presented in section 4 are generally robust to common alternative specifications. In the cases where some of the main variables of interest differ from those in the main specification, there are good theoretical and empirical reasons to prefer the main specification.

4.2 Industry-specific Impacts

In this section, I consider the impacts of PTA depth on individual industries. Data on trade in individual industries was sourced from the *International Trade and Production Database for Estimation* (ITPD-E) of Borchert et al. (2021, 2022). The ITPD-E data consists of administrative trade statistics grouped into 170

Figure 3: PTA depth effects by industry



Note: This figure depicts the coefficient estimates for the linear and quadratic essential provisions measures across 170 industries. Statistically insignificant estimates are plotted at zero.

different industries. These groupings include 28 agriculture industries (1986–2016), 7 mining and energy industries (1988–2016), 118 manufacturing industries (1988–2016), and 17 services industries (2000-2016).[11] Like the data from the SGMD, the ITPD-E data include both domestic and international trade flows.

The empirical model was estimated individually for each of the 170 industries. For the sake of parsimony and because of difficulties in deriving estimates for the exporter and importer max provision term for many of the industries (due to collinearities), the specification used for each industry was that from column (4) of Table 3. This specification includes the PTA indicator, both the linear and quadratic essential provision count terms, and the three international organization controls for the EU, GATT, and WTO. The estimated impacts of deep PTA policy provisions on each sector are depicted in Figure 3. The value of the linear term is plotted along the horizontal axis while the value of the squared term is plotted along the vertical axis for each industry. In cases where an estimate was not statistically significant at the 10 percent level, that estimate is plotted at 0.

The industry-level estimates are largely consistent with the main estimates for aggregate manufacturing trade but do exhibit heterogeneous effects. For the linear term, 49 industries exhibit a positive estimate, 108 exhibit an insignificant estimate, and only 13 exhibit a negative estimate ($p < 0.10$). For the quadratic

[11] The ITPD-E data also includes data for the years 2017–19, which are not used here in order to better match the time coverage of the SGMD used in the main analysis. The ITPD-E data does not cover the earliest years that are present in the SGMD data.

term, 47 industries exhibit a negative estimate, 108 exhibit an insignificant estimate, and only 15 exhibit a positive estimate. Taken together, 38 industries exhibit PTA depth effects that are initially positive but diminish (bottom right quadrant of Figure 3), matching the estimates using aggregate manufacturing trade. The largest of these industries are electronic valve tubes, plastic products, fresh vegetables, and wearing apparel (excluding fur). Meanwhile, 12 industries exhibit the opposite effects; negative linear effects and positive quadratic effects (top left quadrant). For these industries, shallow agreements appear to reduce trade but additional provisions offset these frictions and increasingly promote trade. The largest industries are transport services, insurance and pension services, motor vehicles, and pharmaceuticals and medicinal chemicals. Finally, for the vast majority of industries (96), additional provisions appear to have no systematic, statistically significant effect on trade—either linearly or quadratically. The remaining 24 industries have only one statistically significant term but they fall primarily along the quadrant of positive but diminishing effects (20) rather than negative but increasing (4). No industries exhibit positive and increasing effects or negative and decreasing effects.

5 General Trade and Welfare Effects of Increasing PTA Depth

The econometric estimates in the preceding section provide a good indication of how PTA depth can impact trade costs and the direct (partial) effects these have on bilateral trade. To build on that insight and examine the effects of these changes in trade costs more extensively, we turn to a general equilibrium version of the gravity model. Doing so allows us to better account for the indirect impacts of PTA depth and connect the effects on trade costs to measures of welfare.

To illustrate the effects of deepening a PTA, I consider the impacts of increasing the depth of the Agadir agreement between Egypt, Jordan, Morocco, and Tunisia, which entered into effect in 2004. Increasing trade among the Arab Mediterranean countries has long been a point of interest. In 1998, Egypt, Jordan, Morocco, Tunisia, and 12 other countries in the region signed the Pan-Arab Free Trade agreement, which lowered tariffs between the 16 members but featured zero essential provisions. In an effort to further encourage trade, the four members adopted the Agadir agreement several years later, which further lowered tariffs and featured 10 essential provisions.

However, despite these multiple efforts to increase trade between the Arab Mediterranean countries, trade between them has remained relatively low. Kourtelis (2021) argues that there are several reasons for the limited impact of the agreement, many of which stem from structural weaknesses of the economies like a lack of regulatory harmonization between them. Many of these weaknesses are the types of issues addressed by modern deep PTAs. Given this background, the Agadir agreement represents an insightful case in point for this analysis. The introduction of additional types of provisions targeting noted structural weaknesses could have large positive impacts on the four member countries.

5.1 GE Structural Gravity

The GE extension of the gravity model is one of the most frequently used quantitative trade models in recent years.[12] Importantly, as noted by Arkolakis et al. (2012), the framework is representative of a wide range of traditional trade models. Drawing on the work of Anderson et al. (2018) and Yotov et al. (2016), the model system takes the following form:

$$X_{ijt} = \frac{Y_{it}E_{jt}}{Y_t}\left(\frac{\tau_{ijt}}{\Pi_{it}P_{jt}}\right)^{1-\sigma}, \tag{2}$$

$$\Pi_{it}^{1-\sigma} = \sum_j \left(\frac{\tau_{ijt}}{P_{jt}}\right)^{1-\sigma}\frac{E_{jt}}{Y_t}, \tag{3}$$

$$P_{jt}^{1-\sigma} = \sum_i \left(\frac{\tau_{ijt}}{\Pi_{it}}\right)^{1-\sigma}\frac{Y_{it}}{Y_t}, \tag{4}$$

$$p_{it} = \left(\frac{Y_{it}}{Y_t}\right)^{\frac{1}{1-\sigma}}\frac{1}{\gamma_{it}\Pi_{it}}, \tag{5}$$

$$E_{it} = \phi_{it}Y_{it} = \phi_{it}p_{it}Q_{it}. \tag{6}$$

Equations 2, 3, and 4 reflect the canonical gravity model of Anderson (1979) and Anderson and van Wincoop (2003). In equation 2, bilateral trade (X_{ijt}) between exporter i and importer j in year t is a function of economic size and trade frictions. Economic size is given by the first component, consisting of the exporter's output (Y_{it}) and the importer's expenditures (E_{jt}), divided by global output (Y_t). Trade frictions are determined by the second component, which consists of bilateral trade costs (τ_{ijt}), outward multilateral resistance (OMR) of the exporter (Π_{it}), inward multilateral resistance (IMR) of the importer (P_{jt}), and the elasticity of substitution (σ). The OMR and IMR terms are defined by equations 3 and 4, respectively. Both terms can be thought of as aggregate trade cost or price indices for both countries, reflecting weighted bilateral trade costs not only between i and j but also between each country and all other countries.

[12] Similar models have been used to examine the impacts of a variety of different types of trade determinants, including the signing of free trade agreements (Anderson and Yotov, 2016; Baier et al., 2019); Brexit (Brakman et al., 2018), and internet connectivity and digital trade policies (Herman and Oliver, 2023), for example.

Equations 5 and 6 add endogenous producer prices (p_{it}) and real expenditures (E_{it}). Producer prices (i.e. factory gate prices received by producers) are a function their global output share, their outward multilateral resistance, and a constant elasticity of substitution (CES) share parameter (γ_{jt}). Meanwhile, expenditures are determined by a market clearing condition in which expenditures are a fixed proportion (ϕ_{it}) of real output. In what follows, real output is equal to the product of producer prices and a fixed quantity of output (Q_{it}). Thus, all changes in output and expenditure are determined by changes in prices rather than adjustments to production quantities.

Compared to the empirical model presented in the previous sections, this GE version of the model allows for a more complete assessment of the impacts of a change in trade costs. The empirical model is able to identify the direct impact of changes in trade costs (τ_{ijt}) on bilateral trade but is generally unable to identify the indirect impacts stemming from the relationship between trade costs and multilateral resistances. When estimated using importer-year and exporter-year fixed effects, the model does not explicitly identify the ways in which a change in a trade cost variable like PTA depth influences these multilateral resistances. The GE model, on the other hand, does explicitly capture these relationships.

Within the model, increasing the depth of a PTA between Egypt and Jordan, for example, is expected to have several layers of effects. Assuming the change in depth lowers trade costs, the first order effects are the direct effects of a lower $\tau_{EGY,JOR}$ on trade between the two countries. This should directly increase bilateral trade between them. The second order effects are those arising from the change in each country's multilateral resistances. When costs between Egypt and Jordan decline, the relative costs of trading with all other countries rise. This will result in trade diversions away from other countries and towards the PTA members. Finally, the third order impacts are those arising from changes in income. Egypt and Jordan will see their incomes rise as a result of the change in PTA depth and will import more from all sources. Meanwhile, the trade diversion will reduce the incomes of all other countries, lowering their imports. The combined GE impact of the change, therefore, is the culmination of all three of these effects. In general, the countries lowering their trade costs are expected to experience the largest positive impacts while everyone else will generally experience small negative impacts.

5.2 Deepening the Agadir Agreement

The simulations were undertaken using most of the same data as the econometric estimations. The main difference is that the trade values were replaced with those from an alternative source in order to insure that the panel was fully balanced with values for imports, exports, and domestic trade for all countries in the model. The raw administrative used for the estimation is missing some values for certain countries and/or trading pairs due to data unavailability. While the gravity model can be econometrically estimated in the presence of missing values, GE structural simulations require a fully balanced panel. For this reason, the Eora input/output table was used for baseline trade flows instead (Lenzen et al., 2012, 2013).[13] The Eora data use imputation methods to fill in for unobserved flows, which makes the data poorly suited for estimation but well suited for simulation. To match the sectoral coverage of the data used for estimation, a series consisting only of manufacturing trade was derived from the full set of Eora sectors.[14]

Using the Eora data, a baseline model of manufacturing trade was defined based on the year 2016. For the sake of parsimony, the data was narrowed to a set of 53 countries. These 53 countries represent 50 of the largest trading countries in 2016—representing more than 95 percent of all international and domestic trade—as well as the three Agadir members not among the top 50 countries: Jordan, Morocco, and Tunisia.[15]

To simulate the effects of increasing the depth of Agadir, several steps were taken. First, the empirical estimates derived in the previous section were used to construct bilateral trade costs (τ_{ij}) between all 53 countries in 2016. The cost estimates from column (5) of Table 3 were chosen. Second, the bilateral costs were used to solve for baseline multilateral resistance terms.[16] Values for the exogenous parameters ϕ_{it} and γ_{it} were calibrated to the baseline model data and σ was set equal to 7, which is representative of estimates in the literature (Head and Mayer, 2014). With the baseline model defined, the third step introduces counterfactual trade costs. For the Agadir experiment, the depth of the PTA between the four members was increased while all other costs variables were held at their baseline values. Counterfactual trade costs (τ^*_{ijt}) were constructed based on the adjusted depth levels. In the present experiment, this change in depth affects trade costs through both the linear ($\hat{\alpha}_1 = 0.00499$) and quadratic depth terms ($\hat{\alpha}_2 = -0.00004$).[17] Fourth, the model was re-solved using the counterfactual trade costs, producing new counterfactual multilateral resistances, prices, trade flows, and real expenditures. Finally, the baseline and counterfactual values were compared to determine the impacts of the policy change on these values and other outcomes.

The cost estimates produced by the econometric model inherently feature some potential imprecision. While most of the model terms are highly significant and feature relatively small standard errors, all still feature some non-zero error. To capture this potential imprecision in the GE simulations, a Monte Carlo

[13]Specifically, the Eora26 global input/output table is used, which is available at https://worldmrio.com/eora26/.

[14]The selected manufacturing sectors reflect broad ISIC manufacturing categorizations and include Food & Beverages; Textiles and Wearing Apparel; Wood and Paper; Petroleum, Chemical and Non-Metallic Mineral Products; Metal Products; Electrical and Machinery; Transport Equipment; Other Manufacturing; and Recycling.

[15]These 53 countries are: ARE, ARG, AUS, AUT, BGD, BLX, BRA, CAN, CHE, CHL, CHN, COL, CZE, DEU, DNK, EGY, ESP, FIN, FRA, GBR, HKG, HUN, IDN, IND, IRL, IRN, ISR, ITA, JOR, JPN, KOR, MAR, MEX, MYS, NLD, NOR, PAK, PHL, POL, PRT, ROU, RUS, SAU, SGP, SWE, THA, TUN, TUR, UKR, USA, VEN, VNM, ZAF (ISO codes).

[16]The multilateral resistance terms are not uniquely identified so one term must be pinned down to a specific value. All other terms are solved for relative to that term. In the present example, Germany was selected to be that reference country and all terms were constructed relative to Germany's IMR term, which was set equal to 1.

[17]For the sake of focusing on a single channel of interest, the policy change modified only these two measures of PTA depth and not also the country-level exporter and importer maximum depth measures. Within the Agadir experiment, this had no impact for most counterfactual levels of depth as all 4 members belong to other relatively deep agreements with maximum values of 57 provisions (Tunisia), 64 (Egypt), and 66 (Jordan and Morocco).

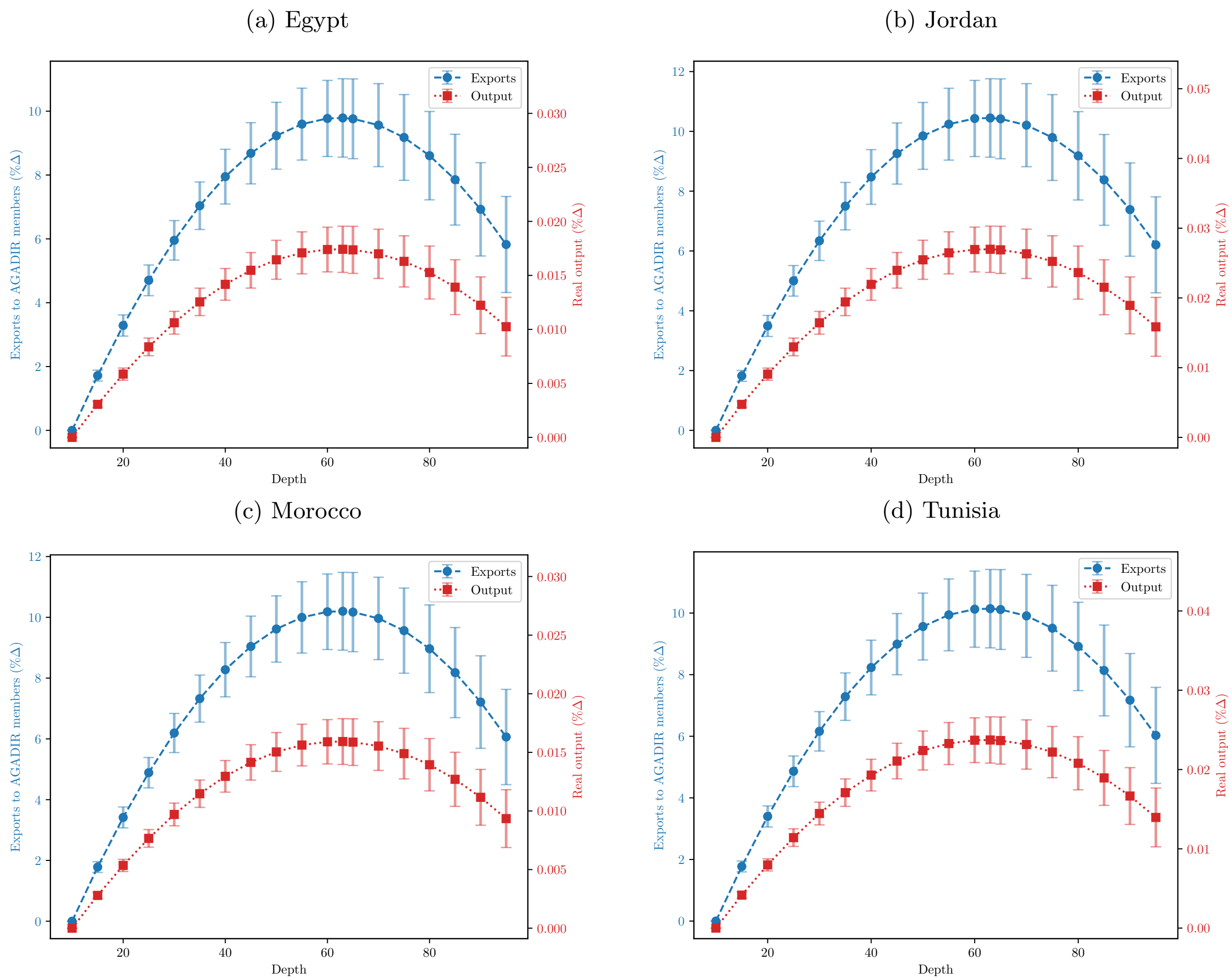
approach was taken. Instead of simply using the point estimates for each value from the regression, a sampling distribution was constructed from the estimates and variance-covariance matrix.[18] The GE model was then solved multiple times using different random draws from this distribution. This allows for the constructions of standard errors of the mean for the simulation results that reflect this natural empirical uncertainty.

In considering the deepening of the Agadir agreement, there are many potential alternative levels of depth. Based on the econometric (partial) estimates, the impact of additional provisions is positive up until the 65th provision and negative afterwards. To examine this trend in the GE model, a series of simulations were completed for a variety of different levels of depth. The model was simulated using a range of potential counterfactual depth values and 10 randomly drawn simulations were completed for each. Figure 4 presents the estimated impacts on key outcomes for each of the four Agadir members. Specifically, it depicts the impacts on each country's exports to the other Agadir members as well as the impact on the value of their real manufacturing output. Notably, real output can be considered a useful measure of welfare in this manufacturing-specific context that is analogous to GDP in a model covering all industries. Each plot depicts the mean estimated impact across the 10 simulations with an error bar representing the standard error of that mean. As expected, the impact of deepening the Agadir agreement is initially positive, increasing trade between the four members and raising their real output. The impacts are generally similar across the four members but tend to be largest for Jordan. As expected, the marginal benefits of additional provisions diminish as their numbers increase. The most beneficial level of depth in the GE model is 63 provisions for all 4 members, which is lower than the 65 provisions suggested by the partial analysis presented in Section 4. This difference is potentially the result of two factors. First, the marginal impacts described in Section 4 were evaluated at the sample-wide mean values of all other variables, which may not be perfectly representative of the values for the four Agadir members. Second, the impacts presented in the previous section did not consider the indirect GE effects of PTA depth—further highlighting the insights that can be gleaned from moving from partial impacts to GE impacts. For levels of depth beyond the optimal 63 provisions, the positive impacts start declining.

Focusing more closely on the impacts of the optimal level of depth, Table 5 presents additional implications of transforming Agadir into a 63 provision agreement. The table reports a variety of outcomes, all reflecting percent changes: exports to and imports from both Agadir members and the rest of the world (ROW), domestic trade, real manufacturing output, producer prices, and inward and outward multilateral resistances (IMR and OMR, respectively). These results are based on an expanded sampling of 100 randomly drawn simulations, which reduces the relative size of the standard errors. Individual results for the other countries

[18] Specifically, the PPML estimates were assumed to be joint-normally distributed (Dobson, 2002).

Figure 4: Simulated effects of deepening the Agadir agreement on the value of exports and real manufacturing output



Note: These figures plot the estimated impacts on exports and real manufacturing output for a range of potential levels of Agadir's depth. For each level of depth, the mean effect is presented as a point with whiskers depicting the standard error of the mean. Numerical values for the export and output estimates are depicted on the left and right sides of each figure, respectively.

Table 5: Impacts of Increasing Agadir depth to optimal level

	Egypt		Jordan		Morocco		Tunisia	
Impact	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Exports to Agadir	12.650	[0.649]	13.515	[0.695]	13.196	[0.678]	13.124	[0.674]
Exports to ROW	-0.231	[0.013]	0.141	[0.007]	0.024	[0.005]	-0.025	[0.004]
Imports from Agadir	13.531	[0.696]	12.515	[0.641]	12.842	[0.659]	12.913	[0.663]
Imports from ROW	0.215	[0.013]	-0.387	[0.023]	-0.137	[0.009]	-0.135	[0.008]
Domestic trade	-0.029	[0.007]	-0.251	[0.016]	-0.120	[0.009]	-0.162	[0.011]
Real output	0.022	[0.001]	0.034	[0.002]	0.020	[0.001]	0.030	[0.002]
Producer prices	0.048	[0.002]	-0.028	[0.001]	-0.004	[0.001]	0.004	[0.001]
IMR	0.026	[0.001]	-0.062	[0.003]	-0.024	[0.001]	-0.025	[0.001]
OMR	-0.048	[0.002]	0.028	[0.001]	0.004	[0.001]	-0.004	[0.001]

Note: This table presents the results of 100 GE gravity simulations in which the number of essential provisions in the Agadir agreement was increased from 10 to 63. For each country and outcome, the mean percent change and standard error of the mean (SEM) are reported, with the latter appearing in brackets. IMR and OMR denote inward and outward multilateral resistances, respectively. ROW refers to "rest of world".

in the model are presented in Table 6 in the appendix.

The results suggest certain distinct patterns of changes. All members trade more extensively with one another, increasing within-region exports and imports by about 13 percent for all members. Jordan increases its exports to Agadir the most while Egypt increases it the least, although the differences between the magnitudes quite small. By comparison, Egypt increases it imports from Agadir the most and Jordan the least. The impacts on trade with the ROW are mixed. Exports to the ROW decrease for Egypt and Tunisia but increase for Jordan and Morocco. This is due largely to the fact that producer prices increase for the former two countries, prompting diversion away for most non-Agadir countries. The opposite is true for Jordan and Morocco. All countries except for Egypt reduce their imports from the ROW, favoring Agadir partners instead. Egypt, on the other hand, increases imports from both sources, driven in part by the relatively large increase in producer prices and, therefore, income. For all countries, the increase in international trade reduces domestic trade, as would be expected.

The changes in IMR and OMR give a sense of whether each country has become relatively more open or closed to trade, regardless of the specific source or destination. All countries except Egypt face lower IMRs and, therefore, lower import price indices. The OMR estimates indicate that Egypt and Tunisia both face lower average export barriers to all destinations, while the opposite is true for Jordan and Morocco. Egypt, Jordan, and Morocco all become relatively more open in one direction while Tunisia becomes relatively more open in both.

All four Agadir members countries experience increases in the real value of their manufacturing output, implying overall welfare gains from the policy change. For Egypt, this increase is largely the result of higher producer prices and lower OMR, which make it easier to export goods at higher prices. For Jordan and

Morocco, the increase is driven primarily by lower IMRs and consumption prices, increasing the relative purchasing power of their output. For Tunisia, it is a combination of higher prices, lower OMR, and lower IMR. While the magnitude of the impacts on real output appear small, it is important to note that this is generally to be expected. Trade with a small number of even a country's most important partners represents a small part of a country's economic activity. Thus even large changes to these individual trade relationships are expected to have a relatively small impacts on welfare overall. Nonetheless, even these small changes in percent terms can result in meaningful dollar-value gains. Based on 2016 values, the increase represents between $7.8 million (Jordan) to $26.0 million (Egypt) in manufacturing output.

In summary, the counterfactual exercise demonstrates that the deepening of trade agreements can have positive effects on trade integration and welfare. This finding is not unique to the Agadir agreement. While Agadir represents a case with significant room for deepening, other PTAs could achieve even larger monetary gains from smaller increases in depth if existing trade volumes are larger than those under Agadir. Based on the empirical estimates, any agreement with fewer than around 65 agreements could benefit from reevaluating their past PTAs. Within the sample considered here, 201 agreements—about 72 percent—have fewer than 65 provisions and could be good candidates for renegotiation.

6 Conclusion

Trade agreements are complex and the traditional ways of identifying their effects likely overlook important differences in their contents. This paper demonstrates that the use of non-linear approaches to examine PTA depth can usefully capture heterogeneous, non-monotonic effects of PTAs. Most notably, it finds evidence that the contents of trade agreements can have diminishing—and even negative—effects on bilateral trade. The most trade-enhancing agreements over the last decades have been moderately deep PTAs featuring around 65 essential provisions. Using the Agadir agreement as an example, the deepening of shallow PTAs to more closely match moderately deep PTAs could result in economic gains for their members.

While this work demonstrates useful new dimensions of PTA effects, there still remain many important directions for future research. For example, quadratic effects of depth are one way to estimate PTA impacts but may not be the best approach. Future work could examine alternative specifications that may offer a more accurate reflection of the effects of PTA depth. It would also be beneficial to look more closely at individual provisions. As evidenced by this paper's findings, it is likely the case that some provisions within an agreement facilitate trade while others limit it. Further, some provisions may stimulate trade but in a way that is not necessarily reflected in typical bilateral trade data. Increased foreign direct investment, for example, may actually increase international trade activity while decreasing cross-border trade flows.

Methods that allow for the direct consideration of provision-level heterogeneity could provide valuable insight into the full range of impacts that a modern deep PTA may have.

References

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review 69*(1), 106–116. https://www.jstor.org/stable/1802501.

Anderson, J. E., M. Larch, and Y. V. Yotov (2018). GEPPML: General equilibrium analysis with PPML. *The World Economy 41*(10), 2750–2782. https://doi.org/10.1111/twec.12664.

Anderson, J. E. and E. van Wincoop (2003). Gravity with gravitas: A solution to the border problem. *American Economic Review 93*, 170–192. https://doi.org/10.1257/000282803321455214.

Anderson, J. E. and Y. V. Yotov (2016, March). Terms of trade and global efficiency effects of free trade agreements, 1990–2002. *Journal of International Economics 99*, 279–298.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130. https://doi.org/10.1257/aer.102.1.94.

Baier, S., J. Bergstrand, and M. Feng (2014). Economic integration agreements and the margins of international trade. *Journal of International Economics 93*(2), 339–350. https://doi.org/10.1016/j.jinteco.2014.03.005.

Baier, S. L. and J. H. Bergstrand (2007). Do trade agreements actually increase members' international trade? *Journal of International Economics 71*(1), 72–95. https://doi.org/10.1016/j.jinteco.2006.02.005.

Baier, S. L. and N. R. Regmi (2023). Using machine learning to capture heterogeneity in trade agreements. *Open Economies Review 34*, 863–894. https://doi.org/10.1007/s11079-022-09685-3.

Baier, S. L., Y. V. Yotov, and T. Zylkin (2019). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *Journal of International Economics 116*, 206–226.

Bergstrand, J. H., M. Larch, and Y. V. Yotov (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review 78*, 307–327. https://doi.org/10.1016/j.euroecorev.2015.06.003.

Borchert, I., M. Larch, S. Shikher, and Y. V. Yotov (2021). The international trade and production database for estimation (itpd-e). *International Economics 166*, 140–166. https://doi.org/10.1016/j.inteco.2020.08.001.

Borchert, I., M. Larch, S. Shikher, and Y. V. Yotov (2022, July). The international trade and production database for estimation - release 2 (itpd-e-r02). Office of Economics Working Paper 2022–07–A, U.S. International Trade Commission. https://www.usitc.gov/publications/332/working_papers/itpd_e_r02_usitc_wp.pdf.

Brakman, S., H. Garretsen, and T. Kohl (2018). Consequences of Brexit and options for a 'global Britain'. *Papers in Regional Science 97*(1), 55–72.

Breinlich, H., V. Corradi, N. Rocha, M. Ruta, J. Santos Silva, and T. Zylkin (2021). Machine learning in international trade research — evaluating the impact of trade agreements. Policy Research working paper no. WPS 9629, World Bank Group, Washington, D.C. https://doi.org/10.1596/1813-9450-9629.

Claudia, H., A. Osnago, and M. Ruta (2017). Horizontal depth: A new database on the content of preferential trade agreements. Policy Research working paper no. WPS 7981, World Bank Group, Washington, D.C. http://hdl.handle.net/10986/26148.

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models* (2nd ed.). Chapman Hall/CRC.

Dür, A., L. Baccini, and M. Elsig (2014). The design of international trade agreements: Introducing a new database. *The Review of International Organizations 9*(3), 353–375. https://doi.org/10.1007/s11558-013-9179-8.

Feenstra, R. C. (2002). Border effects and the gravity equation: Consistent methods for estimation. *Scottish Journal of Political Economy 49*(5), 491–506. https://doi.org/10.1111/1467-9485.00244.

Fontagné, L., N. Rocha, M. Ruta, and G. Santoni (2023). The economic impact of deepening trade agreements. *The World Bank Economic Review 37*(3), 366–388. https://doi.org10.1093/wber/lhad005.

Gurevich, T. and P. R. Herman (2018, February). The dynamic gravity dataset: 1948-2016. Office of Economics Working Paper 2018-02-A, U.S. International Trade Commission. https://www.usitc.gov/publications/332/working_papers/gurevich_herman_2018_dynamic_gravity_dataset_201802a.pdf.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4, pp. 131–195. Elsevier. https://doi.org/10.1016/B978-0-444-54314-1.00003-3.

Herman, P. (2021). Modeling complex network patterns in international trade. *Review of World Economics 158*, 127–179. https://doi.org/10.1007/s10290-021-00429-y.

Herman, P. R. and S. Oliver (2023). Trade, policy, and economic development in the digital economy. *Journal of Development Economics 164*, 103135. https://doi.org/10.1016/j.jdeveco.2023.103135.

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. https://doi.org/10.1111/j.1467-9701.2010.01273.x.

Kohl, T., S. Brakman, and H. Garretsen (2016). Do trade agreements stimulate international trade differently? evidence from 296 trade agreements. *The World Economy 39*(1), 97–131. https://doi.org/10.1111/twec.12272.

Kourtelis, C. (2021). The agadir agreement: The capability traps of isomorphic mimicry. *World Trade Review 20*, 306–320. https://doi.org/10.1017/S1474745620000488.

Laget, E., A. Osnago, N. Rocha, and M. Ruta (2020). Deep trade agreements and global value chains. *Review of Industrial Organization 57*, 379–410. https://doi.org/10.1007/s11151-020-09780-0.

Larch, M. and Y. V. Yotov (2024). Estimating the effects of trade agreements: Lessons from 60 years of methods and data. *The World Economy 47*(5), 1771–1799. https://doi.org/10.1111/twec.13569.

Lee, W., A. Mulabdic, and M. Ruta (2023). Third-country effects of regional trade agreements: A firm-level analysis. *Journal of International Economics 140*, 103688. https://doi.org/10.1016/j.jinteco.2022.103688.

Lefebvre, K., N. Rocha, and M. Ruta (2023). Containing Chinese state-owned enterprises? the role of deep trade agreements. *Review of World Economics 159*, 887–920. https://doi.org/10.1007/s10290-022-00484-z.

Lenzen, M., K. Kanemoto, D. Moran, and A. Geschke (2012). Mapping the structure of the world economy. *Environmental Science & Technology 46*(15), 8374–8381. https://doi.org/10.1021/es300171x.

Lenzen, M., D. Moran, K. Kanemoto, and A. Geschke (2013). Building Eora: A global multi-region input–output database at high country and sector resolution. *Economic Systems Research 25*(1), 20–49. https://doi.org/10.1080/09535314.2013.769938.

Magee, C. S. (2008). New measures of trade creation and trade diversion. *Journal of International Economics 75*(2), 349–362. https://doi.org/10.1016/j.jinteco.2008.03.006.

Mattoo, A., A. Mulabdic, and M. Ruta (2022). Trade creation and diversion in deep trade agreements. *Canadian Journal of Economics 55*(3), 1598–1637. https://doi.org/10.1111/caje.12611.

Mattoo, A., N. Rocha, and M. Ruta (2020). *Handbook of Deep Trade Agreements*. Washington, D.C.: World Bank. http://hdl.handle.net/10986/34055.

Monteiro, J.-A. (2020). Structural gravity manufacturing sector dataset: 1980–2016. World trade organization, Geneva. https://www.wto.org/english/res_e/reser_e/structural_gravity_e.htm.

Orefice, G. and N. Rocha (2014). Deep integration and production networks: An empirical analysis. *The World Economy 37*(1), 106–136. https://doi.org/10.1111/twec.12076.

Osnago, A., N. Rocha, and M. Ruta (2017). Do deep trade agreements boost vertical FDI? *The World Bank Economic Review 30*(Supplement), s119–s125. https://doi.org/10.1093/wber/lhw020.

Roy, J. (2010). Do customs union members engage in more bilateral trade than free-trade agreement members? *Review of International Economics 18*(4), 663–681. https://doi.org/10.1111/j.1467-9396.2010.00904.x.

Santos Silva, J. M. C. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and Statistics 88*(4), 641–658. https://doi.org/10.1162/rest.88.4.641.

Treffler, D. (2004). The long and short of the canada-u. s. free trade agreement. *American Economic Review 94*(4), 870–895. https://www.jstor.org/stable/3592797.

U.S. International Trade Commission (USITC) (2019). U.S.-Mexico-Canada trade agreement: Likely impact on the u.s. economy and on specific industry sectors. Publication no. 4889. https://www.usitc.gov/publications/332/pub4889.pdf.

Weidner, M. and T. Zylkin (2021). Bias and consistency in three-way gravity models. *Journal of International Economics 132*, 103513. https://doi.org/10.1016/j.jinteco.2021.103513.

Xiong, B. and J. Beghin (2014). Disentangling demand-enhancing and trade-cost effects of maximum residue regulations. *Economic Inquiry 52*(3), 1190–1203. https://doi.org/10.1111/ecin.12082.

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, and M. Larch (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model (Online Revised Version)*. World Trade Organization and the United Nations Conference on Trade and Development. https://doi.org/10.30875/abc0167e-en.

A Additional results

Table 6: Expanded collection of GE impacts of a 63 provisions Agadir agreement

Country	ISO	Producer prices	IMR	OMR	Real output	Exports	Imports	Domestic trade
Egypt	EGY	0.0481	0.0257	-0.0480	0.0223	0.5491	0.3067	-0.0291
Jordan	JOR	-0.0275	-0.0616	0.0275	0.0341	0.5492	0.2136	-0.2506
Morocco	MAR	-0.0041	-0.0243	0.0041	0.0202	0.2096	0.1426	-0.1203
Tunisia	TUN	0.0044	-0.0255	-0.0044	0.0299	0.2224	0.1670	-0.1616
Argentina	ARG	-0.0023	-0.0018	0.0023	-0.0005	-0.0009	-0.0022	0.0074
Armenia	ARE	-0.0173	0.0027	0.0175	-0.0199	-0.0657	-0.0162	0.0966
Australia	AUS	0.0006	0.0006	-0.0006	0.0000	0.0103	0.0112	0.0099
Austria	AUT	-0.0011	-0.0004	0.0010	-0.0006	0.0074	0.0076	0.0103
Belgium-Luxembourg	BLX	-0.0004	0.0004	0.0004	-0.0008	0.0078	0.0117	0.0128
Bengladesh	BGD	0.0003	0.0009	-0.0003	-0.0007	-0.0020	0.0102	0.0133
Brazil	BRA	-0.0015	-0.0013	0.0015	-0.0002	0.0033	0.0001	0.0072
Canada	CAN	0.0004	0.0005	-0.0004	-0.0001	0.0092	0.0101	0.0102
Chile	CHL	0.0003	0.0001	-0.0003	0.0001	0.0106	0.0101	0.0085
China	CHN	0.0012	0.0016	-0.0012	-0.0005	-0.0066	0.0126	0.0137
Colombia	COL	0.0004	0.0004	-0.0004	0.0000	0.0113	0.0106	0.0094
Czechia	CZE	-0.0004	-0.0004	0.0004	0.0000	0.0082	0.0083	0.0077
Denmark	DNK	-0.0007	-0.0004	0.0007	-0.0003	0.0078	0.0081	0.0094
Finland	FIN	-0.0008	-0.0007	0.0008	-0.0001	0.0079	0.0068	0.0078
France	FRA	-0.0010	-0.0005	0.0010	-0.0006	0.0060	0.0064	0.0099
Germany	DEU	-0.0007	0.0000	0.0007	-0.0007	0.0055	0.0088	0.0115
Hong Kong	HKG	0.0010	0.0006	-0.0011	0.0004	0.0111	0.0110	0.0083
Hungary	HUN	-0.0004	-0.0006	0.0004	0.0001	0.0081	0.0081	0.0071
India	IND	0.0022	0.0034	-0.0022	-0.0012	-0.0195	0.0271	0.0205
Indonesia	IDN	0.0007	0.0012	-0.0007	-0.0005	0.0067	0.0129	0.0130
Iran	IRN	-1.9633	0.0000	0.0000	0.0000	-2.1462	-2.0126	-1.9969
Ireland	IRL	-0.0002	-0.0002	0.0002	0.0000	0.0089	0.0088	0.0081
Israel	ISR	0.0002	0.0000	-0.0002	0.0002	0.0105	0.0100	0.0079
Italy	ITA	-0.0009	0.0001	0.0009	-0.0010	0.0021	0.0074	0.0130
Japan	JPN	0.0004	0.0005	-0.0004	-0.0001	0.0085	0.0100	0.0101
Korea	KOR	-0.0005	0.0001	0.0005	-0.0006	0.0049	0.0048	0.0114
Malaysia	MYS	0.0006	0.0008	-0.0006	-0.0002	0.0105	0.0123	0.0113
Mexico	MEX	0.0004	0.0004	-0.0004	0.0000	0.0097	0.0098	0.0094
Netherlands	NLD	-0.0006	0.0001	0.0006	-0.0007	0.0081	0.0099	0.0117
Norway	NOR	-0.0002	-0.0003	0.0002	0.0000	0.0090	0.0095	0.0081
Pakistan	PAK	0.0033	0.0060	-0.0033	-0.0027	-0.0240	0.0378	0.0319
Philippines	PHL	0.0009	0.0010	-0.0009	-0.0001	0.0093	0.0133	0.0113
Poland	POL	-0.0004	-0.0004	0.0004	0.0000	0.0077	0.0087	0.0080
Portugal	PRT	-0.0006	-0.0004	0.0006	-0.0002	0.0071	0.0085	0.0086
Romania	ROU	-0.0011	-0.0005	0.0011	-0.0006	0.0061	0.0074	0.0103
Russia	RUS	-0.0027	-0.0020	0.0027	-0.0007	0.0008	-0.0041	0.0078
Saudi Arabia	SAU	0.0008	0.0017	-0.0008	-0.0009	-0.0022	0.0186	0.0160
Singapore	SGP	0.0005	0.0036	-0.0005	-0.0031	0.0100	0.0253	0.0285
South Africa	ZAF	0.0002	0.0004	-0.0003	-0.0002	0.0061	0.0100	0.0105
Spain	ESP	-0.0010	-0.0005	0.0010	-0.0005	0.0046	0.0058	0.0096
Sweden	SWE	-0.0010	-0.0005	0.0010	-0.0005	0.0079	0.0068	0.0096
Switzerland	CHE	-0.0008	-0.0006	0.0007	-0.0002	0.0092	0.0089	0.0083
Thailand	THA	0.0004	0.0010	-0.0004	-0.0005	0.0075	0.0119	0.0127
Turkey	TUR	-0.0031	0.0011	0.0031	-0.0042	-0.0403	0.0026	0.0281
Ukraine	UKR	-0.0039	-0.0018	0.0039	-0.0020	-0.0017	-0.0021	0.0133
United Kingdom	GBR	-0.0005	-0.0003	0.0005	-0.0002	0.0078	0.0079	0.0090
United States	USA	0.0004	0.0004	-0.0005	0.0000	0.0096	0.0105	0.0095
Venezuela	VEN	0.0012	0.0013	-0.0012	-0.0001	0.0077	0.0159	0.0119
Vietnam	VNM	0.0009	0.0015	-0.0009	-0.0006	0.0067	0.0149	0.0142

Note: This table presents a full set of results for the 53 countries included in the counterfactual simulation described by Table 5 in the main text. All values represent percent changes from the baseline. "IMR" and "OMR" denote inward and outward multilateral resistances, respectively.