2019

On the Widely Differing Effects of Free Trade Agreements: Lessons from Twenty Years of Trade Integration

Scott L. Baier

Yoto V. Yotov

Thomas Zylkin
University of Richmond, tzylkin@richmond.edu

Recommended Citation

Baier, Scott L., Yoto V. Yotov and Thomas Zylkin. "On the Widely Differing Effects of Free Trade Agreements: Lessons from Twenty Years of Trade Integration." *Journal of International Economics* 116 (2019): 206-226. https://doi.org/10.1016/j.jinteco.2018.11.002.

On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration*

Scott L. Baier
Clemson University

Yoto V. Yotov
Drexel University, ERI-BAS

Thomas Zylkin†
University of Richmond

November 4, 2018

Abstract. We develop a novel two stage methodology that allows us to study the empirical determinants of the *ex post* effects of past free trade agreements (FTAs) as well as obtain *ex ante* predictions for the effects of future FTAs. We first identify 908 unique estimates of the effects of FTAs on different trading pairs for the years 1986-2006. We then employ these estimates as our dependent variable in a "second stage" analysis characterizing the heterogeneity in these effects. Interestingly, most of this heterogeneity ($\sim 2/3$) occurs within FTAs (rather than across different FTAs), with asymmetric effects within pairs (on exports vs. imports) also playing an important role. Our second stage analysis provides several intuitive explanations behind these variations. Even within the same agreement, FTA effects are weaker for more distant pairs and for pairs with otherwise high levels of *ex ante* trade frictions. The effects of new FTAs are similarly weaker for pairs with existing agreements already in place. In addition, we are able to relate asymmetries in FTA effects to each country's ability to influence the other's terms of trade. Out-of-sample predictions incorporating these insights enable us to predict direction-specific effects of future FTAs between any pair of countries. A simulation of the general equilibrium effects of TTIP demonstrates the significance of our methods.

JEL Classification Codes: F13, F14, F15

Keywords: Free Trade Agreements, International Trade, Gravity

*We thank Julian Hinz for thoughtful discussant comments and seminar and conference participants at the Australasia Trade Workshop, the Midwest Trade Meetings, and the International Monetary Fund. Thomas Zylkin is grateful for research support from the NUS Strategic Research Grant (WBS: R-109-000-183-646) awarded to the Global Production Networks Centre (GPN@NUS) for the project titled "Global Production Networks, Global Value Chains, and East Asian Development".

†*Contact information*: Baier: John E. Walker Department of Economics, Clemson University, Clemson, SC, USA 29634. Email: sbaier@clemson.edu. Yotov: School of Economics, Drexel University, Philadelphia, PA, USA 19104. ifo Center for International Economics, ifo Institute, CESifo. Economic Research Institute, Bulgarian Academy of Sciences, Sofia, Bulgaria. E-mail: yotov@drexel.edu. Zylkin: Robins School of Business, University of Richmond, Richmond, VA, USA 23217. E-mail: tzylkin@richmond.edu.

1 Introduction

After a quarter-century of unprecedented trade integration, the world may be taking a momentary pause to re-evaluate the economic impact of free trade agreements. The past few years have seen the United Kingdom threaten to leave the E.U., the United States insist on the re-negotiation of NAFTA, and the high profile failures of TTIP and TPP - two would-be "mega deals" which together would have made 60% of the world's production more interdependent by eliminating barriers to trade. While economic integration has always been controversial, a common theme in the current backlash against FTAs is the perception that past FTAs have not provided the economic benefits promised by policymakers at the time of their signing. Since the broader trend towards ever larger and more comprehensive trade deals seems unlikely to abate, the current moment thus offers an opportunity to take stock of the heterogeneous effects of past FTAs on bilateral trade flows *ex post* to see if these effects reflect the wisdom of established theory and to determine what lessons we can draw to better predict the effects of future FTAs *ex ante*.

The proliferation of new FTAs in recent years offers a useful historical lens for studying how trade agreements achieve liberalization that goes beyond the simple elimination of tariffs. Since 1986, there have been more than 350 new trade agreements notified to the WTO, which have differed in their aim, breadth, and scope. Broadly speaking, however, a shared objective of many of these agreements has been to achieve "deep" integration, i.e., economic integration that goes beyond tariff reduction and extends into policies that are more difficult for the econometrician to observe and to quantify. Leading econometric studies of the "average partial effect" of FTAs on trade, such as Baier & Bergstrand (2007) and Anderson & Yotov (2016), generally support this view, as the estimates they obtain appear too large to be explained by tariff reductions alone.[1]

As a result, those wishing to model the effects of FTAs *ex ante* face a fundamental problem: how to assess their initial, partial equilibrium impact on bilateral trade. If FTAs affect trade

[1] For example, Baier *et al.* (2014) find that "deeper" agreements have stronger effects and Baier *et al.* (2018) find that *ex post* FTA effects can depend on *ex ante* geographic and institutional frictions, such as bilateral distance and the sharing of a common legal system. Incorporating more specific variation in agreement types (e.g., the inclusion of certain provisions) should be regarded as an important direction to pursue, but efforts thus far have only yielded mixed results (see: Kohl *et al.* 2016).

only through tariffs, this partial effect could be computed directly (assuming a constant elasticity). However, in combination, the small current levels of tariffs and the large FTA estimates from the existing literature support the conclusion that the effects of FTAs on bilateral trade go far beyond the simple elimination of tariffs. To allow for such possibilities, a growing number of researchers use econometric estimates of the *ex post* effects of existing FTAs (often just a single average estimate) as a proxy for the effects of future agreements.[2] However, aside from variation in tariffs (or, more nebulously, in "non-tariff measures", NTMs), few theoretically-grounded arguments exist for why these partial effects might differ systematically across different agreements.

Accordingly, the main goal of this paper is to develop methods and hypotheses that will identify meaningful, theoretically-motivated sources of variation for predicting the effects of trade deals *ex ante*. In particular, we pursue a "two stage" estimation procedure for quantifying and studying heterogeneity in the effects of FTAs, using data on trade and production for 70 countries over the period 1986-2006. In a first stage, we expand on the methods of Baier & Bergstrand (2007) and Bergstrand *et al.* (2015) to obtain agreement-specific effects for each FTA signed during the period, as well as "direction-of-trade"-specific estimates for each member pairing within a given agreement. This stage of the analysis delivers a total of 908 direction-specific, widely-varying FTA estimates, which we then use as dependent variables in a second stage that studies the determinants of the variance in the FTA estimates.

To help explain this heterogeneity, we seek guidance from theory. Specifically, we exploit the basic structure of a standard multi-country trade model to examine two novel sources of variation. First, to the extent that some trade frictions are induced by trade policies and domestic regulations, pairs of countries with higher levels of trade frictions *ex ante* should have more potential for larger FTA partial effects *ex post*. Second, drawing on the influential "terms of trade" arguments of Bagwell & Staiger (1999, 2005), countries with less "market power" over their own terms of trade should grant relatively smaller concessions when they sign FTAs, because they are likely

[2] Aichele *et al.* (2014), Felbermayr *et al.* (2015), Anderson *et al.* (2015), and Carrère *et al.* (2015) each use econometric estimates of the average effect of past agreements to model the effects of TTIP (as well as TPP, in the case of Carrère *et al.*, 2015).

already close to their "politically optimal" set of trade policies. To operationalize these insights, we introduce two indices: (i) a comprehensive index of "pre-FTA trade barriers", which we obtain naturally from our econometric first stage model, and (ii) a simulated measure of each member of an agreement's "terms of trade sensitivity", which, to emphasize the connection with the theories of Bagwell & Staiger, we also refer to as "revealed market power". Our analysis finds robust support for both of these hypotheses.

We also confirm two other intuitively plausible hypotheses that can account for a fraction of the observed heterogeneity. First, countries with a prior trade agreements already in place tend to have weaker partial effects from any subsequent agreements. Second, even after controlling for the level of existing trade frictions, FTA partial effects are weaker for countries which are further apart geographically. This may be because more distant countries are less sensitive to changes in trade policies (as emphasized in Baier *et al.*, 2018) or perhaps because they find it more difficult to coordinate on deeper integration because of weaker cultural affinities.

One other variable that has been highlighted in the broader literature on FTAs also draws our interest. Following the work of Kehoe & Ruhl (2013) on the "new goods" margin of trade - as well as subsequent work by Kehoe *et al.* (2015) - we test whether countries that trade a relatively small range of products *ex ante* have more potential for "explosive" trade creation after the signing of an FTA. Interestingly, we find that the number of products trade *ex ante* is, if anything, *positively* related to the amount of trade creation *ex post*, seemingly contradicting these earlier findings. Two remarks help reconcile this discrepancy. First, to the extent that a low traded goods margin manifests itself in a "gravity" framework as high trade costs, our comprehensive index of "pre-FTA trade barriers" already takes this margin into account. Second, the distinction between "across-" versus "within-" agreement heterogeneity is again important here. In particular, when we restrict our focus to asymmetric FTA effects within the same pair of countries - better approximating the case study design used in Kehoe & Ruhl (2013) and Kehoe *et al.* (2015) - we do find that a low traded goods margin helps predict asymmetries in trade creation.

In light of recent trends towards larger multilateral trade blocs, an especially appealing aspect of our two stage approach is that we can easily narrow our focus to heterogeneous effects that might occur within individual agreements.[3] Interestingly, we find that variation in FTA effects *across* different agreements (i.e., the difference between NAFTA and Mercosur) explains only about one-third of the variation in our first stage estimates of FTA effects. Of the remaining two-thirds, which are due to "within agreement" heterogeneity, almost half (i.e., almost one-third of the total) is due to asymmetric effects "within pairs". With the exception of prior agreements, our key variables of interest remain relevant when we limit attention to heterogeneity within agreements. We are also specifically able to relate asymmetries in trade creation within pairs to differences in market power, as would be predicted by terms-of-trade theory.

We subject these insights to a battery of other controls that may be plausibly linked to trade creation, including "gravity" relationships, depth of the agreements, factor endowment differences, institutional frictions to trade, and *ex ante* tariff policies. These additional regressions reveal several useful auxiliary findings. For example, we find that our preferred second stage model significantly outperforms observed tariffs in explaining the heterogeneity in our FTA estimates, especially when we consider heterogeneity "within agreements" and "within pairs". Information on factor endowments and/or institutional development similarly only seems useful for predicting the heterogeneity that occurs across agreements; they are silent as to why the same agreement could have different effects for different members. We do, however, find robust evidence that FTA effects are larger for countries with larger economic size (GDP).

By design, our two stage approach is well-suited for developing and validating a model for making *ex ante* predictions. Drawing on the machine learning literature, we use an "out of sample" prediction analysis to test if an empirical model fitted from the first stage estimates of one set of agreements can reliably predict the partial effects found in the excluded agreements.[4] As an illustration, we use the predictive model developed from this out-of-sample analysis to generate

[3] Very few papers in the literature allow the same agreement to have different effects on different members. Baier *et al.* (2018) allows for this type of heterogeneity by interacting their EIA dummies with *ex ante* pair-specific relationships, rather than conditioning on agreement-specific variables. Our two stage approach combines the perspective of Baier *et al.* (2018) with that of Zylkin (2015), who allows for direction-of-trade-specific effects within NAFTA.

[4] In practice, we drop one agreement at a time, then try to predict its effects out of sample.

unique, direction-of-trade-specific predicted partial effects of a revived TTIP on trade between the U.S. and E.U. Compared with an alternate scenario in which TTIP has the same partial effect on all TTIP trade flows (as is typically assumed in other analyses), we find that allowing for heterogeneous partial effects has important consequences for how TTIP affects welfare, which follow directly from the empirical insights we document in the second stage analysis. In this way, our three-part approach offers a cohesive methodology for studying and predicting heterogeneous effects of FTAs: the first stage creates preliminary objects of interest, the second stage provides a thorough empirical deconstruction, and the third part demonstrates a novel method for making *ex ante* predictions that highlights and validates the practical usefulness of the first two parts.

Naturally, our approach is complementary to the prevailing, non-econometric methods that have been used to model the effect of trade policies *ex ante*. Because FTAs have shifted their focus away from tariffs, applied work in the CGE literature increasingly aims to quantify the impact of non-tariff provisions of FTAs on so-called "non-tariff barriers". However, as discussed in Fugazza & Maur (2008), because of the complexity of these issues, even the best-possible estimates of non-tariff barriers must be interpreted with caution and model results based on these estimates may be highly fragile to minor variations in methodology. Our methods admittedly lack the specificity of a fully-specified CGE framework. Nonetheless, we are able to identify several broad sources of variation in FTA effects that have been previously overlooked, have strong theoretical and intuitive appeal, and appear to have robust support in the data.

Lastly, because we allow FTA effects to differ by agreement, we also contribute to a longstanding literature that has examined the effects of individual agreements. This literature begins with the seminal work of Tinbergen (1962), who found only small effects for the Benelux and British Commonwealth preference arrangements, and also includes other influential studies by Frankel & Wei (1997) and Carrère (2006), who allow for differences across several major modern regional trading blocs. Methodologically, the most related work in this area is Kohl (2014), who observes that FTA effects may differ based on WTO membership and on the institutional quality of an agreement. By and large, these studies have not found that most FTAs have increased trade. For example, Kohl

(2014) finds only 27% of FTAs have had positive and significant effects on trade. We, however, find positive effects for a majority (57%) of the agreements in our study. This could be for three reasons. First, we follow the econometric recommendations of Santos Silva & Tenreyro (2006) in using Poisson PML, as opposed to OLS, in order to account for heteroskedasticity of trade data and to be able to use the information contained in zero trade flows. Second, we include consistently-measured internal trade flows, which enable us to capture the possibility that increases in trade between liberalizing countries may actually be at the expense of internal trade.[5] Third, bilateral trade flows may not instantaneously respond to the implementation of an FTA. To account for the phasing-in of different provisions and other adjustments that may accompany FTAs, we allow agreements to have lagged effects that accumulate over time.

The following section describes the first stage estimation procedure. Section 3 adds details on how we construct our data and key variables. Section 4 summarizes our first stage FTA estimates. These estimates are then used as the dependent variables in the second stage analysis, which is contained in Section 5. Section 6 uses the insights from the second stage to forecast the effects of TTIP. Finally, Section 7 adds concluding remarks.

2 Decomposition of FTA Effects

This section describes how we recover heterogeneous estimates of the effects of FTAs. We start with a brief review of the structural gravity model, which delivers an estimating equation for the "average partial effect" of an FTA. We then discuss how to decompose this average effect into successively nested layers of heterogeneity, starting with the level of the individual agreement and then allowing for an increasingly more detailed heterogeneity within agreements as well.

2.1 Structural Gravity

Our starting point is a simple, generalized version of the "structural gravity" equation, as originally derived by Anderson (1979) and as popularized by Eaton & Kortum (2002) and Anderson & van

[5] This idea has been explored in the context of FTAs by Dai *et al.* (2014) and by Bergstrand *et al.* (2015).

Wincoop (2003) and subsequently extended to the panel dimension by Baier & Bergstrand (2007).[6] Let X_{ij} denote the value of exports from an origin country i to a destination country j. The gravity equation for these trade flows is

$$X_{ij} = \frac{A_i w_i^{-\theta} \tau_{ij}^{-\theta}}{\sum_l A_l w_l^{-\theta} \tau_{lj}^{-\theta}} E_j. \tag{1}$$

In (1), E_j is the total expenditure by purchasers in j on goods across all different origins (including goods produced domestically in j). The share of j's expenditure allocated specifically to products from any one origin i is then directly dependent on the following three factors: A_i, the overall quality of the production technologies available in i; w_i, the wage in i; and τ_{ij}, the "iceberg" trade cost required to send goods from i to j. Goods from different origins are assumed to be imperfectly substitutable; therefore, the effects of production costs and trade costs on trade are subject to a constant trade elasticity $\theta > 1$. Importantly, all cost factors only weigh on trade relative to the overall degree of competition in j's import market, which is accounted for via the summation term in the denominator of (1). Noting that this summation term is specific to the importing country (because it sums across all origins), a more compact way of writing (1) is

$$X_{ij} = \frac{A_i w_i^{-\theta} \tau_{ij}^{-\theta}}{P_j^{-\theta}} E_j, \tag{2}$$

where $P_j^{-\theta} = \sum_i A_i w_i^{-\theta} \tau_{ij}^{-\theta}$. As noted by Anderson & van Wincoop (2003), $P_j^{-\theta}$ serves as a useful aggregate of all bilateral trade costs faced by consumers in importer j.[7] Writing the structural gravity equation as in (2) - with distinct i, j, and i-by-j components - lends itself naturally to deriving a "fixed effects" estimation equation for trade flows, as we demonstrate below.

[6] More recently, Arkolakis *et al.* (2012), Costinot & Rodríguez-Clare (2014), and Head & Mayer (2014) have each shown that the basic empirical structure implied by these earlier papers - gravity estimation with exporter and importer fixed effects - is consistent with a wide range of other trade models, including Armington (1969), Krugman (1980), Melitz (2003), and Melitz & Ottaviano (2008).

[7] Anderson & van Wincoop (2003) famously refer to $P_j^{-\theta}$ as j's "inward multilateral resistance". They also show how to derive a seller-side analogue to $P_j^{-\theta}$ (the *"outward* multilateral resistance") that similarly aggregates bilateral trade costs for producers in exporter i. By performing this exercise, Anderson & van Wincoop (2003) show how (1) can be re-written in the form of a more traditional gravity equation, in which both countries' "economic mass" (i.e., GDP) enters directly.

2.2 Panel Econometric Implementation

From an empirical perspective, our primary object of interest is the combined trade frictions parameter $\tau_{ij}^{-\theta}$. We are not, however, interested in the general determinants of trade frictions - e.g., geographical distance, historical affinities, etc. - which have been widely explored in the broader gravity literature.[8] Rather, we wish to focus on how FTAs have shaped changes in $\tau_{ij}^{-\theta}$ over time. More specifically, we aim to shed light on how different agreements have had very different effects on $\tau_{ij}^{-\theta}$ and, furthermore, how changes in $\tau_{ij}^{-\theta}$ may vary widely even within the same agreement.

Our first step in this direction is to follow Baier & Bergstrand (2007) in deriving a panel implementation of (2) that permits identification of an *average* FTA effect across all the agreements in our sample. To ease this derivation, we first add a time subscript, t, as well as an error term, $\epsilon_{ij,t}$, and re-write (2) in exponential form:

$$X_{ij,t} = \exp\left(\ln A_{i,t} w_{i,t}^{-\theta} + \ln \frac{E_{j,t}}{P_{j,t}^{-\theta}} + \ln \tau_{ij,t}^{-\theta}\right) + \epsilon_{ij,t}. \tag{3}$$

Next, use the following generic functional form for the trade costs term $\ln \tau_{ij,t}^{-\theta}$

$$\ln \tau_{ij,t}^{-\theta} = Z_{ij}\delta + \beta_1 FTA_{ij,t} + \beta_2 FTA_{ij,t-5} + u_{ij,t},$$

where Z_{ij} can be thought of a set of time-invariant controls for the general level of trade costs between i and j with coefficient vector δ. For more traditional gravity applications, the contents of Z_{ij} would normally be specified to include geographical and/or historical ties between countries, as discussed above. For our purposes, however, the main variables of interest are $FTA_{ij,t}$, a 0/1 indicator for if i and j belong to a common free trade agreement at time t, and $FTA_{ij,t-5}$, a 5 year lag of $FTA_{ij,t}$. The inclusion of this lagged term accounts for the possibility that FTA effects may "phase-in" over time, as has been previously shown by such papers as Baier & Bergstrand (2007) and Anderson & Yotov (2016).[9]

[8] For a thorough reference, see Anderson & van Wincoop (2004).

[9] Our specific choice of a single 5 year lag reflects a compromise between allowing for phasing-in effects and trying to identify lags for as many agreements as possible. However, we have also thoroughly examined robustness to other possible ways of specifying the timing of effects. For the aggregate analysis, we have experimented with 2-12 year lags as well as with 1-6 year leads. The results, shown in our Online Appendix, reveal the average phasing-in period

A concern that arises is that there are components of Z_{ij} that are unobserved and correlated with FTAs, and not accounting for these factors will result in biased coefficient estimates. A key insight from Baier & Bergstrand (2007) is that specific knowledge of δ is neither necessary nor sufficient in order to obtain consistent estimates of the effects of FTAs. Instead, drawing on standard panel estimation techniques described in Wooldridge (2002), Baier & Bergstrand (2007) recommend using pair-specific fixed effects in place of $Z_{ij}\delta$, such that the time dimension of the panel identifies the (average) causal effect of FTAs on trade. With this same strategy in mind, our baseline specification for estimating the average effect of FTAs on trade barriers becomes

$$X_{ij,t} = \exp\left(\eta_{i,t} + \psi_{j,t} + \gamma_{\overline{ij}} + \beta_1 FTA_{ij,t} + \beta_2 FTA_{ij,t-5}\right) + \varepsilon_{ij,t}. \tag{4}$$

Here, $\eta_{i,t}$ and $\psi_{j,t}$ are, respectively, time-varying exporter- and importer- fixed effects meant to absorb the $\ln A_{i,t} w_{i,t}^{-\theta}$ and $\ln E_{j,t}/P_{j,t}^{-\theta}$ terms in (3), which are endogenous and cannot be observed directly. In addition, $\eta_{i,t}$ and $\psi_{j,t}$ also effectively control for all country-level factors on the exporter and on the importer side, respectively. $\gamma_{\overline{ij}}$ is a (symmetric) pair-wise fixed effect that strips out all time-invariant determinants of trade barriers between i and j.[10] Lastly, we treat the additive residual term $\varepsilon_{ij,t}$ as both reflecting measurement error in trade values as well as now also absorbing the error term in $\ln \tau_{ij,t}^{-\theta}$ above. Following the recommendations of Santos Silva & Tenreyro (2006, 2011), we estimate (4) - as well as all subsequent specifications described in this section - using Poisson Pseudo-maximum Likelihood ("PPML").[11]

We obtain our final baseline specification for estimating the effects of FTAs by incorporating

of the FTAs in our sample is well-approximated by a 5 year lag. We also find that none of the leads are statistically or economically significant.

[10] Baier & Bergstrand (2007) estimate two versions of (4), one with bilateral fixed effects and one in log-differences, such that the pair fixed effects drop out of the estimation. Because we estimate in levels using symmetric fixed effects, our specific implementation draws more directly on Anderson & Yotov (2016). Later, when we allow for asymmetric FTA effects, it is necessary to also use asymmetric pair effects in order to obtain unbiased estimates. In general, however, imposing symmetry on the pair effects leads to more efficient estimates, because it reduces the number of parameters to be estimated. Thus, we stick with symmetric pair effects for our initial specifications.

[11] By using PPML, we implicitly assume that the the variance of $\varepsilon_{ij,t}$ is proportional to the conditional mean of $X_{ij,t}$. Santos Silva & Tenreyro (2006, 2011) show that PPML generates relatively robust results even when this assumption is not satisfied and/or the data features many zero trade values. Arvis & Shepherd (2013) and Fally (2015) offer further reasons for PPML's suitability in the gravity context. Egger & Staub (2015) and Head & Mayer (2014) provide a broader comparison between PPML and other, non-GLM estimators not considered in Santos Silva & Tenreyro (2006). Piermartini & Yotov (2016) offer broader discussion of the econometric challenges associated with gravity estimation as well as recommended best practices.

the methods of Bergstrand *et al.* (2015), who argue that FTA estimates based on specification (4) may be biased upward because they may be capturing the effects of globalization. Adapting a related idea from Yotov (2012), the simple adjustment proposed by Bergstrand *et al.* (2015) is to explicitly control for the effects of globalization in the gravity model by introducing a set of globalization dummies. Applied to our setting, this adjustment results in the following econometric model:

$$X_{ij,t} = \exp\left(\eta_{i,t} + \psi_{j,t} + \gamma_{\overline{ij}} + \beta_1 FTA_{ij,t} + \beta_2 FTA_{ij,t-5} + \sum_t b_t\right) + \varepsilon_{ij,t}, \tag{5}$$

where the added term $\sum_t b_t$ is a set of dummies that equal 1 for international trade observations (as opposed to *internal* trade, X_{ii}) at each time t. The coefficients on these time-varying border dummies, the b_t's, capture the process of globalization over time, as all countries trade more with each other and less with their own internal markets.

The average "total" (or "cumulative") effect of FTAs on trade after accounting for phasing-in can be constructed as $\beta \equiv \beta_1 + \beta_2$. The specific interpretation of β can be described in one of two ways. From a strictly econometric point of view, β is the total average *partial* effect of an FTA on bilateral trade flows, noting that FTAs also influence trade through the country-specific terms $\eta_{i,t}$ and $\psi_{j,t}$. A second, more structural interpretation is that β gives the average treatment effect of an FTA specifically on "trade costs" - i.e., its effect on the $\ln \tau_{ij,t}^{-\theta}$ term in (3). To ease this latter interpretation, note that the combined term $\exp(\gamma_{\overline{ij}} + b_t + \beta_1 FTA_{ij,t} + \beta_2 FTA_{ij,t-5})$ describes the predicted level of $\tau_{ij,t}^{-\theta}$ in place between countries i and j at time t. To comment more thoroughly on the identification of β, note that increases in trade between i and j do not translate directly to implied reductions in the trade cost term $\ln \tau_{ij,t}^{-\theta}$. Instead, due to the presence of the time-varying exporter and importer fixed effects, the impact of FTAs is only identifiable when trade increases between i and j relative to each country's trade with all other partners. Importantly, and consistent with theory, the set of outside partners for each country includes the value for X_{ii}, the value of sales to one's own market, or "internal trade". We regard accounting for internal trade as a key feature of our empirical approach. As documented empirically in Dai *et al.* (2014), including internal trade in the estimation should lead to larger, more precise estimates of FTA effects.[12]

2.3 Allowing for FTA heterogeneity

With the wide adoption of Baier & Bergstrand's methods, most estimates of the effects of FTAs in the prior literature generally find positive and significant "average" results.[13] However, for the purposes of policy analysis, an obvious weakness of estimating an "average" FTA effect is that the effects of a given agreement may be substantially different from the average; thus, it may not be appropriate to apply an average estimate of the effects of all existing FTAs for making *ex ante* predictions about the effects of specific FTAs. To capture and analyze this potential heterogeneity in FTA effects, we expand on the initial specification shown in (5) in three successive steps:

First, we consider a specification where FTA effects are allowed to vary at the level of the underlying agreement, similar to the approach taken in Kohl (2014):[14]

$$X_{ij,t} = \exp\left(\eta_{i,t} + \psi_{j,t} + \gamma_{\overline{ij}} + \sum_{A} \beta_{1,A} FTA_{ij,t} + \sum_{A} \beta_{2,A} FTA_{ij,t-5} + \sum_{t} b_t\right) + \varepsilon_{ij,t}, \tag{6}$$

where we allow for a distinct average partial effect - $\beta_A \equiv \beta_{1,A} + \beta_{2,A}$ - for each individual agreement, using superscript A to index by agreement and also allowing for agreement-specific lags. This initial refinement allows us to make useful statements about which FTAs in our sample have been more successful than others about promoting trade. However, it is silent about the possibility that the same agreement may not affect all countries involved in exactly the same way.

Second, we allow for further heterogeneity at the level of each trading pair within an agreement. For example, we allow Sweden's accession to the E.U. in 1995 to have different effects on its trade barriers with Germany vs. its trade barriers with the U.K. The resulting specification is

$$X_{ij,t} = \exp\left(\eta_{i,t} + \psi_{j,t} + \gamma_{\overline{ij}} + \sum_{A} \sum_{p \in A} \beta_{1,A:p} FTA_{ij,t} + \sum_{A} \sum_{p \in A} \beta_{2,A:p} FTA_{ij,t-5} + \sum_{t} b_t\right) + \varepsilon_{ij,t}, \tag{7}$$

where each $p \in A$ is a pair of countries (i, j) belonging to agreement A, counting (i, j) and (j, i) as

[12] Indeed, as we will later show, our accounting for internal trade allows us to obtain substantially more "optimistic" estimates of the effects of individual agreements than the prior literature.

[13] We note some exceptions to this statement: Frankel *et al.* (1997); Ghosh & Yamarik (2004); Head & Mayer (2014). These papers differ methodologically from the approach used in Baier & Bergstrand (2007), however.

[14] For earlier work, see also: Frankel & Wei (1997), Soloaga & Winters (2001), and Carrère (2006).

the same pair. $\beta_{A:p} \equiv \beta_{1,A:p} + \beta_{2,A:p}$ then gives us a corresponding set of agreement-pair-specific FTA estimates.

Third, we consider the possibility that, even within a given pair, an FTA may not affect trade in both directions symmetrically. For this last refinement, let $d \in A$ denote a unique "directional pair" of countries $(\overrightarrow{i,j})$ belonging to agreement A, where the notation $(\overrightarrow{i,j})$ refers specifically to the effect on trade flows where i is the exporter and j is the importer. We thus estimate two effects for each agreement-pair, one for each direction of trade. In addition, since FTAs no longer affect each partner symmetrically within a given pair, we now also introduce an asymmetric pair fixed effect "$\gamma_{\overrightarrow{ij}}$", which varies by direction as well. We then have the following econometric model:

$$X_{ij,t} = \exp\left(\eta_{i,t} + \psi_{j,t} + \gamma_{\overrightarrow{ij}} + \sum_{A}\sum_{d\in A}\beta_{1,A:d}FTA_{ij,t} + \sum_{A}\sum_{d\in A}\beta_{2,A:d}FTA_{ij,t-5} + \sum_{t} b_t\right) + \varepsilon_{ij,t}, \qquad (8)$$

$\beta_{A:d} \equiv \beta_{1,A:d} + \beta_{2,A:d}$ then gives us a unique set of direction-specific estimates which we will soon use for our "second stage" analysis of the empirical determinants of trade integration.

Before proceeding further, we pause to clarify two important details. First, the heterogeneous FTA estimates that will be obtained from specifications (6), (7), and (8) should be interpreted with care. While Baier & Bergstrand (2007) has emerged as the standard method for consistently estimating the *average* treatment effect of FTAs, the same cannot be said when we pull apart our average "β" to obtain increasingly more finely-grained coefficients, which we should regard as being estimated with at least some unobserved error. Instead, our preferred method draws on the suggested approach of Lewis & Linzer (2005): even if the individual $\beta_{A:d}$'s we estimate from contain some unobserved noise, we can still investigate heterogeneity in the causal effects of FTAs by using a "second stage" regression analysis to extract some useful signals from that noise.[15]

Second, our desire to allow for individual lagged effects creates an additional complication because it is not possible to estimate 5 year lagged effects for the agreements that form within

[15] Formally speaking, what we have in our $\beta_{A:d}$'s (for example) is a measure of how a particular $\tau_{ij,t}$ changed for a given FTA and direction of trade. These direction-specific changes in trade barriers could be due to the FTAs themselves or could be because of "noisier" reasons (changes in wind patterns, for instance). What we require for causal interpretation of our *second stage* estimates is that the "noise" in our first stage is uncorrelated with our second stage covariates. We return to this consideration in more detail in Section 5.

the last 5 years. While we could drop these later agreements from the sample, doing so effectively discards the useful information in the contemporaneous FTA effects $\beta_{1,A}$, $\beta_{1,A:p}$, and $\beta_{1,A:d}$. Instead, our preferred way of addressing this issue is to use the direction-specific estimates (the $\beta_{A:d}$'s) from the first 15 years of data to estimate the following auxiliary regression

$$\beta_{2,A:d} = \pi_0 + \pi_1 \beta_{1,A:d} + \upsilon_{A:d}, \tag{9}$$

such that we may then infer the needed remaining lags using $\beta_{2,A:d} = \pi_0 + \pi_1 \beta_{1,A:d}$. We use the same coefficients to infer missing values for $\beta_{2,A}$'s and $\beta_{2,A:p}$'s where needed.[16]

3 Data Construction

This section describes the sources and the construction of the data with emphasis on several specially constructed indices that are theoretically motivated and used as key regressors in our second stage analysis in Section 5. These indices include our novel measures of "pre-FTA trade barriers" and "revealed market power", as well as the "new goods" margin of Kehoe & Ruhl (2013).

3.1 Trade and FTA Data for First Stage Analysis

Trade. We construct a data set with information on manufacturing production and trade for a sample of 70 countries over the twenty year period 1986 - 2006.[17] Table A.1 of the Data Appendix lists the countries included. For computational reasons, we combine 17 countries which do not form any FTAs during the period into a single "Rest of the World" aggregate region. Thus, in the end, we arrive at a balanced panel of 53 trading regions observed over the 21 year period 1986-2006.[18] Our primary source for bilateral trade flows is UN COMTRADE. Since partner countries tend to report different values for same trade flow, we generally use the mean of reported values

[16] We also find the results of this regression to be of interest for their own sake and discuss them further in Section 4. In addition, we perform an extensive robustness analysis in the Online Appendix to show our results are robust to alternative choices regarding missing lags.

[17] We use manufacturing data both because of its wide coverage and completeness and also to maintain comparability with similar studies.

[18] All first stage regressions make use of every single year of the data. Cheng & Wall (2005) have argued against using consecutive years in gravity regressions on the basis that wider intervals allow more time for trade to adjust to changes in trade costs. We have found similar results for the case of four year intervals. We favor using every year because it gives us maximal degrees of freedom for identifying direction-specific FTA effects.

when possible. If either country fails to report a value, we use the non-missing value.

An important feature of our dataset is that it includes values for "internal trade" flows (a.k.a. "domestic sales"). To construct internal trade observations, we combine data on industry-level gross output from two main sources: the CEPII TradeProd database and UNIDO IndStat. We have selected both the sample of countries and the period of study in order to achieve the widest possible use of the available production data from these sources. Since production values in TradeProd are largely taken from earlier versions of UNIDO IndStat - and further augmented using the World Bank "Trade, Production, and Protection" database by Nicita & Olarreaga (2007) - we generally use TradeProd to provide production data for earlier years and data from UNIDO to fill in later years where needed.[19] We also cross-check against the World Bank data to fill in some additional missing values from the beginning of the period. We then construct internal trade values as the difference between the value of a country's gross output and the value of its total exports to other markets. In some isolated cases, however, it is not possible to calculate internal trade values because the production data is either missing or implies a negative value for internal trade. We address these issues in a series of steps. First, we apply linear interpolation between non-missing values whenever possible. Second, if values are negative or missing only for a particular industry, we apply the average share of expenditure spent on domestic output by that country on other (non-missing) industries. Finally, we extrapolate any remaining missing production values at the beginning or end of the sample using the evolution of that country's industry-level exports.[20]

Free Trade Agreements. Our starting point for FTA data is the set of FTAs used in Baier & Bergstrand (2007). We update and cross-check this data against information available via the WTO's website as well as the NSF-Kellogg Database on Economic Integration Agreements. Table A.2 of the Data Appendix provides a complete summary of the agreements included in our study.

[19] Specifically, the TradeProd data is reported in the ISIC Rev. 2 industry classification, covering 1980 to 2006, and the IndStat data reports values using the ISIC Rev. 3 classification, covering 1995 to 2009. Because ISIC Rev. 2 industry codes do not map one-to-one to the Rev. 3 industry codes, we construct country-specific concordances based on matched years, using the 4 digit level of industry detail whenever possible. For these matched years, the correlation between the original ISIC Rev. 2 production values and the (post-concordance) IndStat values is .990.

[20] We also experimented with using the U.S. GDP deflator as an alternate basis for extrapolating missing output, following the procedure used in Anderson & Yotov (2016). This method makes virtually no difference for our results.

Overall, our FTA data cover65 different agreements, including 8 multilateral trading blocs, 32 bilateral FTAs, and 25 agreements between multilateral blocs and outside partners.[21] Within these 65 agreements, there are 455 different agreement-pairs, counting as separate any instance where two countries that are already joined via a prior agreement become part of a second agreement (e.g., Canada and the U.S. in the case of NAFTA). Since we estimate two effects per agreement-pair, we would ordinarily be able to estimate $2 \cdot 455 = 910$ distinct FTA effects in total. However, because Iceland never exports to Romania before the signing of the EFTA-Romania FTA in 1993, we are not able to obtain a directional estimate for EFTA-Romania on Iceland-Romania exports. We therefore drop the Iceland-Romania pair from the latter two sets of estimates, leaving us with 454 pair-specific estimates and 908 direction-specific estimates to be used in the subsequent analysis.

3.2 Key Covariates for Second Stage Analysis

FTA Estimates. To construct the dependent variable for the second-stage analysis of the determinants of the impact of FTAs, we combine the estimates of the current and lagged effects of FTAs. The Online Appendix offers robustness experiments with alternative treatments of the lags.

Pre-FTA Trade Barriers. As originally observed by Baier & Bergstrand (2007), the main advantage of using a panel specification with pair fixed effects to identify the effects of trade policies is that the pair fixed effects effectively absorb all bilateral trade frictions in the cross-section. Importantly, this includes any "unobservable" component of trade costs, which otherwise would enter the error term and potentially lead to inconsistent estimates. Therefore, the pair fixed effect $\gamma_{\overrightarrow{ij}}$ in (8) contains potentially very useful information about the full level of *ex ante* trade barriers between any potential FTA pair, including any unobservable trade costs. However, because $\gamma_{\overrightarrow{ij}}$ is *direction*-specific, and because of collinearity between $\gamma_{\overrightarrow{ij}}$ and the exporter and importer fixed effects $\eta_{i,t}$ and $\psi_{j,t}$, we cannot directly interpret the values for $\gamma_{\overrightarrow{ij}}$ we recover from the estimation as reflecting "trade barriers" (i.e., the $\tau_{ij}^{-\theta}$'s). In principle, however, we can identify the average trade level of *ex ante* trade barriers for trade between a given pair i, j by imposing symmetry on

[21] We do not estimate a separate effect for EFTA, since it precedes our study.

the $\gamma_{\overrightarrow{ij}}$'s. In practice, we can perform the following regression:

$$X_{ij,t} = \exp\left(\eta_{i,t} + \psi_{j,t} + \gamma_{\overline{ij}} + \sum_t b_t + \sum_A \sum_{d\in A} \widetilde{\beta}_{1,A:d} FTA_{ij,t} + \sum_A \sum_{d\in A} \widetilde{\beta}_{2,A:d} FTA_{ij,t-5}\right) + \varepsilon_{ij,t}, \quad (10)$$

where the tilde superscript on the set of the partial FTA effects $\widetilde{\beta}_{1,A:d}$ and $\widetilde{\beta}_{2,A:d}$ reflects that fact that we are *constraining* these to be the same as we estimated previously from (7) using directional fixed effects. The $\gamma_{\overline{ij}}$'s are similarly constrained to be symmetric within pairs.

The combined term "$\gamma_{\overline{ij}} + b_{t-1}$" then provides a measure of the average level of trade barriers between i and j in the year before the signing of an FTA at time t. Since $\gamma_{\overline{ij}} + b_{t-1}$ is an *inverse* measure of pre-FTA trade frictions between i and j, we expect that it should enter with a negative sign in the second stage analysis. Because it is an "inclusive" measure of these frictions, it can plausibly control for a variety of different obstacles to trade that could potentially be eliminated by an FTA and would be difficult to capture otherwise, including idiosyncratic differences in domestic regulations that inhibit trade (as emphasized in Baier & Bergstrand, 2007), the trade-muting effects of uncertainty over future trade policies (c.f., Limão, 2016, Sec. 4.3), as well as each country's unilateral incentives to impose restrictive trade policies *ex ante*. Furthermore, since this measure holistically controls for the full magnitude of *ex ante* trade barriers between a given pair of countries, we can then also include standard proxies for trade costs (ln DIST, COLONY, etc.) and these will now in turn give us more specific inferences about the roles these variables play in explaining the first stage partial effects (rather than also reflecting the role these variables play in determining the magnitude of initial trade barriers). In our empirical analysis, we will simply refer (with some abuse in terminology) to the combined term "$\gamma_{\overline{ij}} + b_{t-1}$" as our "first stage pair fixed effect". Note also that, in the year before an FTA, we have that $\tau_{ij,t-1}^{-\theta} = e^{\gamma_{\overline{ij}} + b_{t-1}}$.

Revealed Market Power. Despite its inclusiveness, a key weakness of the pre-FTA trade barriers index just-described is that it is strictly symmetric and does not allow us to speak to why we should observe asymmetries in trade creation within agreements. One plausible theoretical reason for asymmetries are through differences in each country's terms of trade-related incentives for restricting trade *ex ante*. We therefore also introduce a measure of each FTA country's "terms of trade sensitivity" (or "revealed market power"), which we will use to examine whether asymmetries in our estimated FTA effects can be rationalized based on terms-of-trade theory.

We derive this latter index as follows. First, we take a given set of FTA-signing countries in the year before they entered the agreement (e.g., the U.S., Canada, and Mexico in 1993). We then use a standard multi-country general equilibrium model (described further in Section 6) to simulate the change in each country's "terms of trade" as a result of a symmetric reduction in trade barriers (i.e., a common "partial effect" $\bar{\beta}$) applied to all trade flows within this set of countries. The index we use for the change in a country's terms of trade is

$$\widehat{ToT}_{A:j} = \frac{\widehat{w}_j}{\widehat{P}_j}, \tag{11}$$

which, following Anderson & Yotov (2016), uses the ratio of the change in a country's producer price $\widehat{w}_j$ to the change in its purchasing price $\widehat{P}_j$ (which also gives the change in its real wage).[22] We repeat this process for every FTA in the sample, using the same common partial effect each time (we use the overall average partial effect estimated from (5), $\bar{\beta} = 0.293$) and a (typical) trade elasticity of $\theta = 4$. This procedure will deliver $\widehat{ToT}_{A:j}$'s corresponding to each country j in every agreement A in our sample, which are directly comparable with one another and should not be systematically related to the partial effects estimated from the first stage, other than for the reasons we are investigating.[23] Furthermore, similar to our (symmetric) "Pre-FTA Trade Barriers" index, $\widehat{ToT}_{A:j}$ is also a relatively "inclusive" measure of terms of trade incentives in that it plausibly reflects incentives to use other trade policies aside from tariffs.

The connection to theory follows from what the resulting $\widehat{ToT}_{A:j}$ indices reveal about the sensitivity of a country's terms of trade with respect to the trade barriers of its prospective FTA partners. As elementary trade theory suggests, in the absence of a trade agreement, countries with more abil-

[22]Another, perhaps more-standard approach to computing terms of trade would be to focus on the ratio of *export* prices to *import* prices. The main issue we encounter in trying to implement an analogous measure based on export prices and import prices is that the latter approach requires that we first "artificially balance" global trade, so that the results are not sensitive to the choice of numeraire. Aside from ease of computation, our preferred formulation of terms of trade also has the advantage of retaining the intuition of the standard definition (see Anderson & Yotov, 2016, p. 282.) The Online Appendix adds further discussion and explores various alternative ways of assessing market power.

[23]That is to say, we can be assured the first stage estimates should not be mechanically picking up these general equilibrium price changes because the exporter-time and importer-time fixed effects explicitly control for them.

ity to depress the producer prices of other countries with their trade policy decisions - i.e., more "market power" - have stronger incentives to set high policy barriers to trade *ex ante*. Bagwell & Staiger (1999) then show that, to achieve an "efficient" trade agreement, both countries must fully internalize the externalities their trade policies impose on the other's terms of trade, such that countries with more market power will always make larger trade concessions in any efficient agreement.[24] Our $\widehat{ToT}_{A:j}$ measure captures this idea by revealing the differential responses of each member's terms of trade to a common reduction in trade barriers. For a country with relatively low market power, the computed values of $\widehat{ToT}_{A:j}$ will be relatively high, since the increase in their producer price $\widehat{w}_j$ from a symmetric trade agreement will be large relative to changes in producer prices in their prospective partners, which enter the $\widehat{ToT}$ index through the change in the buyer price index $\widehat{P}_j$.[25] For a high market power country, it is the opposite, since - for a reciprocal agreement - its own price levels are relatively less affected. Thus, guided by theory, we expect the FTA partial effects we estimate - the $\beta_{A:d}$'s from the first stage - to be larger when the importing country's $\widehat{ToT}_{A:j}$ is low.[26]

Of course, the original theory of Bagwell & Staiger (1999) is a theory of multilateral tariff negotiations via the WTO rather than a theory of FTAs *per se*. Since Bagwell & Staiger (1999)'s seminal contribution, several subsequent frameworks help to clarify the applicability of terms-of-trade theory to our setting with FTAs. First, even if we expect WTO members to have already achieved their globally efficient trade policy levels, Bagwell & Staiger (2005) demonstrate that countries which subsequently negotiate preferential FTAs still have a "bilateral opportunism" incentive to lower their trade policies further in order to distort world prices to their mutual advantage. Second, Staiger & Sykes (2011) and Staiger (2012) have shown how terms-of-trade theory

[24]Because low market power countries do not make large trade concessions, some form of "side payment" may be needed to make the efficient agreement feasible. Limão (2007) models these side considerations more formally as cooperation on "non-trade objectives". Hofmann *et al.* (2017) document that provisions reflecting non-trade objectives have become increasingly commonplace in modern FTAs.

[25]Eicher & Henn (2011) explore a similar idea in the case of WTO partial effects. Bagwell & Staiger (2011) and Ludema & Mayda (2013) also explore a similar idea in the case of WTO negotiated tariff reductions.

[26]We note that "market power" is not the same thing as "size". In a many-country world with costly trade frictions, a country's influence on the terms of trade of others depends not only on size differences, but also on how the $N \times N$ system of general equilibrium relationships between all pairs of countries responds to a given change in trade policies, which is captured by our theory-consistent $\widehat{ToT}$ index.

may be applied to non-tariff barriers when tariff policies are constrained by the WTO.[27] Third, Limão (2007) illustrates how FTA provisions on "non-trade objectives" (such as intellectual property rights, security cooperation, and labor and environmental standards) create additional avenues for low-market power countries to compensate high-market power countries for increased market access that are not available though the WTO. Under any of these frameworks, the balance of trade concessions within FTAs should flow from countries with more ability to manipulate the terms of trade to countries with less, consistent with the hypothesis stated above.

New Goods Margin. For some specifications, we also follow Kehoe & Ruhl (2013) in accounting for the possibility that country pairs that trade a smaller range of product varieties *ex ante* have more potential for "explosive" trade growth *ex post*. The measure of the number of traded products we use is the Hummels-Klenow decomposition of the "extensive margin" of trade (Hummels & Klenow, 2005), using Kehoe & Ruhl (2013)'s (pair specific) "least traded goods" cutoff to determine whether to count a variety as traded or not. Specifically, the extensive margin of trade from i to j at time $t-1$ (i.e., the year before an agreement) is constructed as

$$\text{Ext. Margin}_{ij,t-1} = \frac{\sum_{p \in \Omega_{ij}} X_{Wj,t-1}}{\sum_{p \in \Omega_{Wj}} X_{Wj,t-1}}. \tag{12}$$

$X_{Wj,t-1}$ is the volume of trade that each importer j receives from the world at time $t-1$. Each variety p in (12) is an 5 digit SITC product variety, obtained from COMTRADE and assembled using the same procedures described above for aggregate trade. As in Kehoe & Ruhl (2013), p is only assigned to the "traded goods set" Ω_{ij} if, when varieties are sorted by trade volume, bilateral trade volume in p lies above the 10th percentile. (12) thus gives us a flexible measure of the share of products exported from i to j, weighted by each product's contribution to j's total imports.

[27] The theoretical ideas highlighted in Staiger & Sykes (2011) and Staiger (2012) are supported by Kee *et al.* (2009)'s empirical finding that tariffs and non-tariff barriers tend to be policy substitutes for one another. Our own results suggest asymmetries in market power are independently related to asymmetries in post-FTA trade creation even when *ex ante* tariff levels are taken into account.

4 Summarizing FTA estimates

Using the first stage econometric specifications presented in equations (5)-(8), we generate 4 distinct sets of FTA estimates. Each set of estimates consists of current and lagged FTA effects. For expositional simplicity, the presentation in this section focuses on "total" FTA effects (i.e., the sum of the corresponding current and lagged FTA estimates). Following Larch *et al.* (2017), all first stage standard errors are "three-way" clustered by exporter, importer, and year.[28]

Average FTA estimate. We start by briefly discussing the average (across all agreements and pairs) total FTA effect that we obtain based on (5). This specification corresponds to the standard approach in the literature and is the easiest to describe: we estimate an average total FTA effect of $\beta = 0.293$, with a standard error of 0.105 ($p = 0.005$). This estimate yields the interpretation that, on average, FTAs have a partial effect of $(e^{0.293} - 1) * 100 = 34.0\%$ on trade flows. Alternatively, using a typical value for the trade elasticity of $\theta = 4$ in combination with gravity theory, our average estimate implies a $(1 - e^{-\frac{0.293}{4}}) * 100 = 7.1\%$ average decline in bilateral trade frictions.[29]

Inferring lags for later agreements. As noted, we would like to extract as much information as possible from our first stage estimates of later agreements while still acknowledging that "phasing-in" is an important component of FTA-related trade growth. After first obtaining estimates of $\beta_{1,A:d}$ and $\beta_{1,A:d}$ from (8), we subsequently estimate (using (9))

$$\beta_{2,A:d} = 0.141 + 0.201 \cdot \beta_{1,A:d} + \upsilon_{A:d}. \tag{13}$$

With "robust" standard errors, both of the estimates in (13) are highly statistically significant ($p < 0.001$ in both cases.) Importantly, in addition to allowing us to include later agreements in our analysis, this method also generates results that contribute to the literature on the timing of

[28]Egger & Tarlea (2015) and Larch *et al.* (2017) discuss how three-way clustering leads to significantly more conservative inferences of gravity estimates versus other standard approaches. For our purposes, an added advantage of three-way clustering is that estimates become overall less precise when we examine increasingly disaggregated FTA effects, as is seen in the accompanying figures in this section. We have found the same is not true when clustering by country-pair, for example.

[29]The lagged term $\beta_2 = .199$ ($s.e. = .054$) turns out to be both larger and more significant than the current term $\beta_1 = 0.093$ ($s.e. = .077$), confirming the importance of allowing for phasing-in effects of FTAs. More detailed estimates, shown in the Online Appendix, suggest that trade growth on average peaks 3-6 years after the signing of an agreement, then falls off shortly thereafter.

FTA effects (c.f., Magee, 2008; Baier *et al.*, 2014). In particular, our finding that $\pi_0 > 0$ implies that lagged effects of FTAs are more likely to be positive than their corresponding initial effects. Likewise, $\pi_1 > 0$ suggests that agreements with larger initial effects should also be expected to have larger lagged effects.

Agreement-specific estimates. Our estimates of agreement-specific effects obtained by estimating (6) are shown in Table 1, grouped by sign and significance and listed in descending order. As an alternative means of conveying the heterogeneity in these effects, we also offer a graphical depiction in Fig. 1, which presents the distribution of our estimates with their associated 95% confidence bounds.[30] Several features of these results stand out. First, not all the agreement-specific effects that we estimate are positive and statistically significant. We find that 38.5% of our estimates (25/65) are statistically insignificant at the $p = 0.05$ significance level, and 7.6% of our point estimates (5/65) are negative and significant. Nonetheless, the fact that we find positive, statistically significant partial effects for the majority (53.9%) of the agreements in our sample is re-assuring given the more mixed results found in the prior literature. Kohl (2014), for example, only obtains positive and significant effects for 27% of the agreements in his study. We attribute the more "optimistic" findings from our analysis to the inclusion of internal trade values in our estimation. As shown in Dai *et al.* (2014), internal trade is an important component of the overall reference group for estimating theoretically consistent effects of trade policies; thus, estimations that include internal trade generally obtain larger, more precisely estimated FTA effects.[31]

Second, we notice that some countries in our sample generally seem to have had consistently larger (partial equilibrium) impacts from trade agreements. Central and Eastern European countries in particular (e.g., Bulgaria, Hungary, Romania, Poland) are well-represented in the first column

[30] For later agreements, we impute the displayed standard errors by multiplying standard error we obtain for the relevant initial effect (i.e., $\beta_{1,A}$, $\beta_{1,A:p}$, $\beta_{1,A:d}$) by $(1+\pi_1)$.

[31] Indeed, comparing our results with those of Kohl (2014) reveals that our estimates are larger and more significant across the board, while otherwise mostly similar in relative magnitudes. In our own experiments, we have found that the percentage of FTAs with positive and significant effects falls substantially when we drop internal trade observations or if we estimate the first stage using OLS (with logged trade flows as the dependent variable) instead of PPML. The ranking of the different effects from largest to smallest still closely resembles the one shown in Table 1 in each case. We have also experimented with estimating these effects with vs. without lags and with different lag intervals. Again, the ranking of effects remains very similar to Table 1 in all cases.

of Table 1. We also note generally strong effects for agreements signed by Israel, Turkey, Mexico, Mercosur, and the Andean Community. On the other hand, with the exception of Israel and Turkey, other Mediterranean nations - such as Egypt, Tunisia, and Morocco - have generally experienced more modest effects. We will return to this issue of whether some countries consistently experience larger FTA effects than others in our second stage analysis in Section 5.

Overall, the estimates from Table 1 and Fig. 1 confirm that FTAs have had very heterogeneous effects on trade. The degree of heterogeneity we document echoes some earlier findings in the literature, although we generally observe a more optimistic picture of the efficacy of FTAs on an agreement-by-agreement basis. While these factual considerations are useful to note, there remain three important avenues along which we would like to deepen the analysis. First, many agreements involve three or more countries; thus, we wish to examine how the same agreement will affect different pairs of member countries. Second, policymakers generally want to know how trade policies will affect their countries specifically; thus, it is important to allow for direction-specific FTA effects. Third, as noted in Baier *et al.* (2018), even the most finely-tuned appraisals of *past* FTAs do not by themselves tell us anything about what the effects of *future* FTAs will look like. These considerations motivate our more detailed pair- and direction-specific FTA estimates.

Pair-specific and direction-specific estimates. Given the number of estimates we can potentially obtain from specifications (7) and (8) - 455 and 910, respectively - it is not practical to present a full listing of the many pair- and direction-specific effects we estimate. Instead, we summarize the heterogeneity we observe in our pair- and direction-specific estimates in two complementary ways. First, in Table 2, we offer a snapshot of the substantial variation in partial effects that can be observed within a single agreement, using the E.U. as our example. Second, in Fig. 2 and Fig. 3, we add graphical depictions of how the distribution of partial effects changes when we allow for more specificity, i.e. for variation at the agreement-pair level and for directional effects at the agreement-pair level, respectively. The distributions shown in Fig. 2 can be compared directly with the distribution of agreement-specific effects shown in Fig. 1. The two plots shown in Fig. 3 then more specifically highlight, respectively, the degree of pair-wise heterogeneity within each agreement and the degree of asymmetry within each pair.

We focus on the E.U. in Table 2 because it is by far the largest agreement in our sample. There are 98 distinct pair-specific effects -and 196 direction-specific effects - that can be estimated within the E.U. alone. Rather than show all these estimates, Table 2 presents (roughly) the upper and lower quartiles from the pair-specific effects, as well as some representative examples of asymmetries within pairs. Both panels of Figure 3 also place these EU-specific estimates in the context of the heterogeneity we observe within other agreements, such as NAFTA and the E.U.'s agreements with Poland and Chile.

From the top panel of Table 2, we can clearly see a very wide variance in the effects of recent E.U. accessions across different E.U. pairs. Consistent with our expectation that more finely-grained estimates should exhibit relatively more noise, the ranking of the various estimates seems somewhat more random than the ordering seen in Table 1. Smaller countries, such as Malta and Cyprus, appear regularly at the extremes of both lists. Other, more economically large countries - such as the U.K., Hungary, and Portugal - also appear multiple times on both lists.

The lower panel of Table 2 shows some representative examples of E.U. pairs with strongly asymmetric E.U. pairs. For brevity, we only show these effects for three acceding countries (Poland, Austria, and Sweden) and three existing E.U. members (Netherlands, Spain, and the U.K.), though these examples are generally representative. Here, we do see strong, country-specific patterns, confirming again the general pattern suggested by Table 1. Poland's accession has led more so to increases in exports from Poland to existing E.U. members than to increases in trade in the other direction. This dynamic does not appear to be a regular function of E.U. accession, however: asymmetries in the Spain-Austria and Netherlands-Austria pairs both favor trade flows *to* the acceding country (in these cases, Austria) rather than the other way around. We also see a similar pattern for Sweden, another acceding country, in its pairings with these countries. Overall, these examples suggest that, even within agreements, there can be large, *country-specific* patterns of effects that are worth investigating further.

The graphical depictions of the distributions of our various estimates from Fig. 2 and 3 echo these same messages and provide a broader overview.[32] The wider confidence intervals seen in each panel of Figure 2 confirm that, as we move to increasingly more detailed estimates, it becomes increasingly more difficult to obtain statistical precision. Indeed, only 40.4% of our agreement-pair-specific estimates (shown in the left panel of Figure 2) are positive and statistically significant, with this same percentage narrowing to 32.0% for our direction-specific estimates (shown in the right panel of Figure 2). Figure 3 then uses different colors to separate out these effects by the different agreements they belong to. The left panel of Figure 3, which plots agreement-specific estimates against pair-specific estimates, confirms there is generally substantial heterogeneity within the same agreement whenever multiple pairs are involved. Finally, the right panel of Figure 3, which plots directional effects within the same pair against one another, then shows that asymmetries within pairs are also very prevalent, but - at the same time - directional FTA effects within the same pair tend to be highly correlated with one another.

Admittedly, as we focus on increasingly specific estimates of changes in trade costs associated with FTAs, we also increase the likelihood that our estimates reflect omitted factors that may enter specifications (6)-(8) via the error term. Causal interpretation of these more specific FTA effects would require that these effects are directly reflected in changes in trade that occur around the time of the agreement. While this assumption may not strictly hold in all (or even most) cases, presenting these results still allows for a broad description of the vast heterogeneity we observe in the *ex post* partial effects of FTAs in our sample, especially within agreements. In addition, while we acknowledge that individual FTA effects may be measured with unobservable error, taken collectively, these effects can still be analyzed to determine what factors may be expected to be associated with stronger or weaker effects *ex ante*. This is the focus of our second stage analysis.

[32] Baier *et al.* (2018) show similar figures using a random effects model, under the assumption that FTA effects follow a two-tailed normal distribution. We are essentially showing, by estimating explicit dummies for FTA effects, that they do indeed follow a two-tailed distribution.

5 On the Determinants of the Effects of FTAs

In this section, we capitalize on the rich database of FTA estimates that we have constructed in order to study their determinants. We start with a description of our econometric approach and a general characterization of the heterogeneity we observe in our first stage FTA estimates. We then launch into a "second stage" analysis, which takes the direction-agreement-specific partial effects from our estimation of (8) as our dependent variable and regresses them on various covariates of interest in order to gauge and decompose the determinants of these effects. As part of this analysis, we introduce two new variables that prove to be important determinants of the effects of FTAs: (i) our comprehensive index of "pre-FTA trade barriers", which we hope to use to explain the overall magnitude of FTA partial effects, and (ii) our index of "revealed market power", which we will use to investigate potential asymmetries in FTA effects as well as to test some prominent theoretical predictions drawn from Bagwell & Staiger (1999). Importantly, our two stage design makes it easy for us to test how these and other variables contribute to heterogeneous effects within the same agreement and/or within the same pair. The insights that we obtain in this section will feed into the out-of-sample and *ex ante* prediction analysis that will follow in Section 6.

5.1 Econometric Approach

Before turning to the full-blown second stage analysis, we pause to consider two preliminary matters of interest. First, because our first stage estimates of FTA partial effects have been potentially estimated with unobservable error, it is important to discuss how this this error enters the second stage and how our analysis may be structured to account for it. An exceptionally useful reference in this regard is Lewis & Linzer (2005), who carefully examine the consequences of various different ways of weighting a "second stage" dependent variable to account for residual error variance from a prior stage. As Lewis & Linzer (2005) demonstrate, so long as a White (1980) correction for heteroskedasticity-robust standard errors is used, simply using (unweighted) OLS to estimate the second stage will enable us to obtain conservative, if inefficient, inferences of our second stage parameters. That is, while other weighting methods might lead to more efficient estimates, the White

(1980) correction should still generate reliable consistent estimates of our second stage standard errors and confidence intervals, such that we should not be worried about "over-confidence" in our inferences. Lewis & Linzer (2005) also demonstrate that other popular weighting methods for dealing with first stage error can actually perform far worse in terms of efficiency and generally conclude that OLS with heteroskedasticity-robust standard errors is "probably the best approach" in most cases. With these recommendations in mind, we adhere strictly to OLS with robust standard errors for all of our main second stage regressions.[33]

Second, as a direct lead-in for the second stage, we perform a standard "analysis of variance" exercise, sequentially adding *agreement-specific* followed by (symmetric) *pair-agreement-specific* dummy variables as regressors in order to absorb all variance that is associated with heterogeneity "across agreements" and (respectively) "across pairs within agreements". The remaining variance, that is not specific to either an agreement or a pair within agreement, then strictly reflects residual asymmetries in estimated FTA effects within pairs of countries. The main results from this initial decomposition exercise are as follows. First, the agreement-level dummies we use to absorb differences *across* agreements collectively only explain 35.5% of the overall variation in our estimated directional FTA effects (measured by R^2). When we parse the estimates further by next adding pair-agreement-level dummies, the share of explained variation increases to 70.4%, suggesting that 70.4% − 35.5% = 34.9% can be explained by heterogeneity across different pairs within the same agreement. The remaining unexplained variation, 29.6%, specifically reflects asymmetries within the same pair. Notably, this last term is of comparable magnitude to the other two components.

Usefully, this exercise also serves to preview some of the techniques we will use in the analysis that follows to isolate how different variables may explain different aspects of the overall heterogeneity in our estimates. For example, regressing our FTA estimates on a set of observables in the presence of agreement-level fixed effects will allow us to narrowly focus on potential sources of "within agreement" heterogeneity. Similarly, the introduction of agreement-by-pair fixed effects

[33] In the Online Appendix, we also offer a more detailed treatment of the other weighting approaches considered in Lewis & Linzer (2005). Auxiliary results using Lewis & Linzer (2005)'s other preferred alternative - an efficient weighting method proposed by Hanushek (1974) - are very similar to our baseline results using OLS.

will, much as in our initial decomposition exercise, allow us to narrow the focus further to sources of "asymmetries within pairs". Generally speaking, heterogeneity in FTA partial effects within agreements is a topic which has not received much attention in the literature, despite its intuitive appeal. The opportunity to selectively vary our focus in this way - made significantly easier by approaching the problem in two stages - will thus be a key aspect of our overall methodology.

5.2 Decomposing the Sources of FTA Heterogeneity

The preceding discussion highlighted broad patterns of heterogeneity in our first stage FTA estimates. In earlier sections, we have also described at length some attractive, theoretically-guided indices for potentially accounting for some of this variance. The stage is set for a more detailed investigation: What can we say about the empirical determinants of FTA effects? In short order, we will introduce a wide variety of covariates that may hold sway in this context, including "economic geography" variables, institutional factors, as well as neoclassical determinants of trade. For now, however, we start with only a basic specification, drawing solely on a few key hypotheses:

$$\beta_{A:d} = \alpha_0 + \alpha_1 \ln \text{First stage pair FE}_{ij} + \alpha_2 \ln \widehat{ToT}_{A:j} + \alpha_3 \text{Ext. Margin}_{ij,t-1} + \nu_{ij}. \tag{14}$$

As we have established in Section 3.2, each of the key variables in (14) has an intuitive expected sign. The first stage pair fixed effect term, for example, provides a novel and comprehensive (inverse) measure of the level of trade frictions between i and j just before the signing of their agreement. Intuitively, country-pairs with a lower first stage pair fixed effect suffer from higher *ex ante* bilateral trade frictions and, therefore, have more potential for larger trade creation effects from FTAs *ex post*.[34] We therefore expect the sign of α_1 to be negative. $\ln \widehat{ToT}_{A:j}$, meanwhile, measures the sensitivity of the importer's terms of trade based on the simulation procedure described in Section 3.2. A smaller $\ln \widehat{ToT}_{A:j}$ indicates that the importing country has relatively more influence over the terms of trade over its partners. Thus, we expect that $\alpha_2 < 0$, in accordance with the influential arguments of Bagwell & Staiger (1999). We also expect that α_3, the coefficient

[34] As discussed in the Data Section, the pair fixed effect term used in our second stage also incorporates our control for the cumulative amount of non-FTA-related "globalization" that has taken place in the years leading up to the signing of each agreement.

on the *ex ante* extensive margin, to be < 0, incorporating the arguments of Kehoe & Ruhl (2013).

Column 1 of Table 3 shows the results from this simple specification testing each of the above hypotheses. The first stage pair fixed effect and $\ln \widehat{ToT}_{A:j}$, enter with the expected sign and are highly statistically significant (with p-values less than 0.01). Interestingly, the coefficient for our extensive margin measure is significant and *positive*: conditional on our other covariates, countries which trade a wider range of products with one another *ex ante* experience stronger trade creation effects from FTAs *ex post*, contrary to the earlier findings by Kehoe & Ruhl (2013) and Kehoe *et al.* (2015). We offer an explanation for this finding below, where we more specifically study the heterogeneity in our FTA estimates that is due to asymmetries within pairs.

Column 2 adds several other, more standard variables that specifically draw on the "gravity" literature. These include bilateral relationships - such as log distance, contiguity, the sharing of a common language, or the presence of a prior free trade agreement - and also the GDPs and GDP per capitas of each partner.[35] For the most part, we do not impose strong priors on these added variables. We do, however, generally expect log distance ("ln DIST") to be negative, either because more distant countries are more sensitive to changes in trade policies (as emphasized in Baier *et al.*, 2018) or perhaps because they simply sign weaker agreements due to weaker cultural affinities.[36] Likewise, countries that already have a prior trade agreement ("Prior Agreement") should experience weaker trade creation, because the earlier agreements have likely already picked much of the "low hanging fruit" in terms of easily addressed barriers to trade.

The signs on ln DIST and Prior Agreement are indeed negative as expected. We also find larger FTA effects for contiguous countries ("CONTIG") and for countries with common legal system ("LEGAL"). The positive and significant result we observe for CONTIG may be due to the same

[35] Aside from prior agreement, the added bilateral variables are from CEPII. GDP and GDP per capita are from the Penn World Table.

[36] Baier *et al.* (2018) show that if geography, culture, and institutions influence the fixed cost of exporting, then the trade cost elasticity - here, "θ" - will be heterogeneous in a Melitz-style model. For ease of exposition, the framework we use to motivate our estimation does not explicitly accommodate heterogeneous trade elasticities. Nonetheless, even if θ is not common to all countries, the combined term "$\tau_{ij,t-1}^{-\theta}$" we extract from our first stage should at the very least give us a reasonable approximation of the aggregate incidence of trade frictions on trade between i and j, since the direct effect on trade of any given level of trade frictions τ is always channeled through the trade elasticity. Our Online Appendix includes some additional experiments where we add to the second stage external estimates of country-level trade elasticities from Broda *et al.* (2006).

reasons we have already cited for the negative sign we observe on log distance. The positive and significant sign for LEGAL, however, merits further discussion, as we would ordinarily expect countries with dissimilar legal systems to have higher *ex ante* barriers to trade. The key remark to make here is that, because we already have in place an "inclusive" measure of trade costs - in the form of our first stage pair fixed effect - we may interpret these latter variables as affecting the efficacy of the FTA *independently* of how they affect initial trade costs. Conditional on the level of *ex ante* trade frictions, sharing a common legal system (for example) could conceivably make it easier for countries to coordinate on stronger trade creation measures.

Three of the added country-level gravity variables—the log GDPs of both countries and the log per capita GDP of the importer—also enter the analysis significantly. Interestingly, we find that FTA effects are generally stronger for larger countries (measured in real GDP), both on the export-side as well as on the import-side, with the exporter's GDP having more overall influence. While agreements between larger countries are associated with more trade, the combined impact is less than proportionate; that is, a one percent increase in export and importer GDP increases trade by less than one percent. The positive sign on the importer's per capita GDP implies an asymmetry in how FTAs affect trade between countries with differing levels of development, a result we will query in more detail below.

Another important aspect of including a direct measure of country "size" in our analysis is that, all else equal, we would expect size/GDP to be a reasonable proxy for a country's market power over world prices. However, while the magnitude of the coefficient for $\ln \widehat{ToT}_{A:j}$ does weaken when we add these controls, it remains robustly significant. Furthermore, the relative magnitudes of the coefficients for GDPs are not consistent with a story based on market power: if anything, they suggest FTAs promote trade from larger countries to smaller ones. We also note that, while our estimate of the coefficient on the extensive margin variable is still positive and significant, its magnitude falls by roughly half when these other controls are added.

Within-agreement heterogeneity. As noted above, a unique aspect of our two stage design is that we can easily switch our focus from analyzing broad patterns of heterogeneity (as in Table 3) to the heterogeneity that occurs more narrowly within individual agreements and (later) within individual trading pairs. In addition, this approach will enable us to test the robustness of our results in the presence of a rich set of fixed effects. Column 3 of Table 3 repeats the same specification used in column 2, only now with added *agreement-level* fixed effects, such that our estimates now reflect only the residual variation that takes place within agreements. Most of our results carry over from before, with a few notable exceptions. For example, *Prior Agreement* now becomes very close to zero and statistically insignificant. Apparently, these limits are only relevant for determining the overall impact of an agreement on all partners; conditional on a multilateral agreement being signed (where at least one pair already has an agreement and one does not), we cannot identify an effect. CONTIG likewise loses its significance. Interestingly, our earlier, surprising result for the extensive margin also disappears when add agreement fixed effects.

Within-pair heterogeneity. To narrow the analysis even further, Column 4 of Table 3 moves from including agreement-level fixed effects (as in column 3) to now including (symmetric) *agreement-by-pair* fixed effects. Because these fixed effects absorb all symmetric bilateral variables (e.g., ln DIST, etc.), there are only a few key variables that can be identified, and their estimates should be interpreted carefully as strictly reflecting determinants of asymmetric FTA effects within pairs. For example, once again, we obtain a negative and highly significant estimate on $\ln \widehat{ToT}_{A:j}$. Because our agreement-pair fixed effects absorb the average level of trade liberalization within every pair of FTA-signing countries, this result specifically indicates that asymmetries in trade barrier reductions strongly favor the exports of the country with relatively less market power. This to us is the most literal test of the "terms of trade" argument of Bagwell & Staiger (1999).

Turning to the other results in column 4, we once again confirm the exporter's economic size (i.e., its log real GDP) is relatively more important than that of the importer. However, we are not able to confirm our finding from the previous specifications that asymmetries in FTA effects favor the exports of the less developed country (in per capita GDP terms). In addition, we note that the estimate on the extensive margin is now entering negatively, albeit with marginal statistical significance ($p < 0.10$). While this is still not strong evidence in favor of the "least traded goods" hypothesis, the comparison across specifications helps clarify why our earlier results differ starkly from Kehoe & Ruhl (2013) and Kehoe *et al.* (2015), since these studies each examine changes in trade within individual FTA pairings on a case-by-case basis, which is most similar to our focus on within-pair heterogeneity in column 4.

5.3 Robustness

Having established our main specifications in columns 1-4 of Table 3, we now move on to several key sensitivity analyses encompassing some natural lines of inquiry as well as some deeper concerns about causal interpretation. Our first task is to introduce more detailed information on trade policies in place and/or the "depth" of the agreement being signed. To this end, columns 5 to 8 of Table 3 repeat the same specifications as before, only now adding the pre-FTA tariff level ("ln 1+Applied Tariff", which uses weighted-average applied tariffs taken from TRAINS), a 0/1 dummy variable for if the agreement entails the formation of a customs union and/or common market, and two other, more specific measures of FTA depth (the "Count of Enforceable Provisions" and the "Institutional Quality" of the agreement, which we take from Kohl *et al.*, 2016).[37]

The inclusion of these trade policy variables at first appears meaningful. In particular, the estimate on the *ex ante* tariff level in column 5, where we are repeating the same specification as column 1, is positive and highly significant as expected ($p < .01$). In addition, while the coefficient for Customs Union is basically zero, the other depth measures also enter positively, albeit with marginal ($p < .10$) significance.[38] Interestingly, the statistical significance of ln Applied tariff disappears in the presence of gravity controls (column 6), agreement fixed effects (column 7), or agreement-pair fixed effects (column 8). Our original main variables and controls, meanwhile, remain largely unaffected. We conclude, based on these results, that our approach allows us to

[37]Note that, aside from the customs union dummy, each of these variables causes us to drop observations from our main sample. The problems with availability of tariff data are of course well-known. Even though we use MFN tariffs to infer missing applied tariff data whenever possible, there are still 54 observations in our sample for which we cannot reasonably infer what the applied tariff would have been before the agreement. For the depth measures taken from Kohl *et al.*, 2016, we note that they also provide a third summary measure along these lines reflecting the total number of provisions included in an agreement. For most agreements in our sample, the number of enforceable provisions is the same as the number of provisions. We thus decided not to use this latter index.

[38]We are not the first to find that such "depth" indices are not robustly correlated with the amount of trade created by FTAs. Indeed, these mixed results are in line with the findings from Kohl *et al.* (2016).

capture important sources of variation in the *ex post* effects of FTAs that go beyond what we can infer based on tariffs alone.[39]

Next, we consider two further sets of variables that may plausibly be linked to larger FTA effects: (i) "institutional" factors (such as each country's rule of law, the degree of democracy, etc.) and (ii) the role of factor endowment differences and other "Neoclassically" motivated drivers of trade creation. Columns 1-4 of Table 4 investigate the role played by institutions. The institutional indicators we consider are: each country's rule of law and bureaucracy quality (each from the Institutional Country Risk Guide), their degree of democracy (from Polity IV), and the degree of "checks and balances" in their respective political systems (from the World Bank's Database of Political Institutions). For some motivation, we are interested to know whether asymmetries in FTA effects tend to favor the exports of countries with low levels of institutional development levels, based on the supposition that FTAs may help developing countries circumvent constraints on their export capacity imposed by their weak institutional environment (c.f., Anderson & Marcouiller, 2002; Manova, 2013). Additionally, the political science literature (c.f., Mansfield *et al.*, 2007) has argued that stronger checks and balances and/or a stronger degree of democracy should weaken the ability of an executive to enact more sweeping trade policy changes.

The results of these experiments offer mixed support for the notion that FTAs specifically promote the exports of countries with weak institutions. When we look at broad results without any agreement-level or agreement-pair fixed effects (columns 1 and 2 of Table 4), we generally find that FTAs have stronger effects for countries with strong legal institutions (i.e., high "Rule of Law") and/or weak bureaucratic institutions (i.e., low "Bureaucracy Quality"). While there is indeed an implied asymmetry potentially favoring exporters with weaker institutions in these initial specifications, all of these variables lose their significance when we introduce agreement fixed effects (in column 3) as well as when we add agreement-pair fixed effects (in column 4). For democracy and checks and balances, we see a similar story: in columns 1 and 2, we overwhelmingly find that

[39] A possible reason why applied tariffs are no longer significant in column 5 is because they are highly correlated with several of the controls we add in this column, including distance, exporter GDP, exporter and importer GDP per capita, and (naturally) whether or not the two countries have a prior agreement. When we regress ln Applied tariff (as the dependent variable) on these added controls, we obtain an R^2 of 0.349.

these variables are positively associated with signing stronger agreements. However, they play no role in explaining heterogeneity within agreements. Other key results remain unchanged.

Motivated by the Neoclassical "Heckscher-Ohlin" trade theory, the remaining columns of Table 4 control for factor endowment differences across countries (e.g., physical and human capital-to-labor ratios, skill-to-labor ratios). We also include the absolute difference in these relative endowments, as well as the absolute difference in (log) per capita GDPs, in order to get at various potential dimensions of comparative advantage. Data for each of these endowment measures are constructed using information from the Penn World Table. Echoing what we have seen in previous results, while we do find some significant results of note - in particular for physical capital-ratios and the difference therein - these results only appear to be meaningful for describing variations in FTA effects across different agreements. When we examine these variables in the presence of agreement fixed effects (column 7) or agreement-pair fixed effects (column 8), none are robustly significant. Our main results still retain the same sign and significance.

Finally, as we have discussed, the identification of our second stage covariates requires that the unobservable "noise" present in our first stage estimates can be treated as part of the second stage error term. To devise a practical falsification test for this assumption, we utilize what we will call the "first stage residual" from the years leading up to each agreement, defined as

$$\ln \text{resid}_{t_0-m} = \ln(X_{ij,t_0-m}/\widehat{X}_{ij,t_0-m}),$$

where "X_{ij,t_0-m}" is the volume of trade sent from i to j in year t_0-m before an agreement beginning in year t_0 and "$\widehat{X}_{ij,t_0-m}$" is its fitted value from (8). In principle, by including these residuals directly in the *second* stage, as we do in Table 5, we can verify to what extent our first stage estimates are picking up pre-existing changes in trade patterns as well as whether these pre-existing changes affect the identification of our second stage estimates. In practice, since earlier work by Magee (2008) has shown that a substantial portion of trade creation occurs in the 4 years leading up to an FTA - an "anticipation effect" - we look to see if we can draw a reasonable line between where we would expect to see only noise versus where we might expect to see anticipation when we add these residuals to our second stage.

The estimates from Table 5 show that such a line clearly exists between the 3rd and 4th year leading up to an FTA: when we include residuals from years $t_0 - 6$ to $t_0 - 4$ directly in our second stage, we observe no statistical relationship with our first stage estimates across all of our key specifications, whereas residuals from the two-to-three year period immediately preceding FTAs tend to be significantly correlated. Importantly, all our earlier results maintain the same signs and significance as in columns 2, 3, and 4 of Table 3 and their magnitudes are only modestly affected. Since we are now allowing *actual* changes in trade from just before an FTA to explain our estimates, this last set of results confirms that our second stage coefficients are primarily being identified by the trade growth that occurs after each FTA goes into effect.[40]

6 Out-of-sample Predictions & Ex Ante Analysis

In this section, we develop a simple, parsimonious prediction model for the effects of FTAs, using candidate predictors drawn from the main empirical results above. We then apply our model to the task of predicting the effects of a hypothetical U.S./E.U. FTA ("TTIP") on the trade volumes and welfare of each all potential TTIP members and non-members in our sample. Notably, we find very large differences in predictions depending on whether we apply an average effect for TTIP versus specific predictions for TTIP's effects on all possible trade flows between members.

6.1 Developing a Prediction Model

Our out-of-sample analysis proceeds by dropping one agreement at a time and then trying to predict its effects based on a model fitted using the other agreements in the sample. For our criteria, we aimed for a set of predictors which: (i) performs well against other alternatives in terms of

[40] Another strategy, which we take up in our Online Appendix, is to treat the estimated standard errors of each individual estimate from the first stage as reasonable indicators of this error and weight accordingly. However, we prefer the above approach because it requires us to explicitly examine whether the movements in trade which identify our second stage coefficients are specifically tied to the timing of the agreements. We have also experimented with examining these residuals individually by year as well as with regressing these residuals (as our dependent variable) on our second stage regressors. All results support the conclusion that the relationships we are identifying in the second stage do not begin to emerge until the three-year period leading up to the FTA and that most of the identification comes afterwards.

R^2; (ii) can be naturally motivated based on our empirical results from Section 5; (iii) contains variables that were shown to be important for "within-agreement" and (if applicable) "within-pair" heterogeneity. We also focus only on the 5th through 95th percentiles of our estimates.[41]

Based on these criteria, our preferred model is the following simple prediction specification:

$$\widetilde{\beta}_{A:d} = \widetilde{\alpha}_0 + \widetilde{\alpha}_1 \ln \text{DIST}_{ij} + \widetilde{\alpha}_2 \text{First stage pair FE}_{ij} + \widetilde{\alpha}_3 \text{GDP}_i + \widetilde{\alpha}_4 \text{GDP}_j + \widetilde{\alpha}_7 \ln \widehat{ToT}_{A:j} + \nu_{ij}. \quad (15)$$

All the predictors used in (15) were shown to be consistently significant throughout the preceding section, especially when we zeroed in on heterogeneity within agreements and (if applicable) within pairs. FTA effects should be smaller for countries that have a higher first stage pair fixed effect (indicating lower *ex ante* trade frictions) and/or are further apart geographically, and/or if they already have an existing agreement in place. In addition, we also allow for asymmetries in FTA effects within pairs by including the GDPs of both partners and our (inverse) measure of the importing country's market power, $\ln \widehat{ToT}_{A:j}$. Fig. 4 offers a visualization of the resulting predictive fit. Each data point in Fig. 4 represents a predicted FTA effect (indexed by the horizontal axis), which we compare with the actual estimate obtained from the first stage (indexed by the vertical axis). The coefficient on our fitted regression line, $\rho_1 = 0.773$, is positive and highly significant. The constant from the fitted regression line, $\rho_0 = 0.091$, is also positive and highly significant.

Before turning to TTIP, we acknowledge two limitations of this approach. First, the predictive fit of our preferred model ($R^2 = 0.178$) indicates that we are able to predict a significant but modest amount of the overall heterogeneity in the effects of FTAs with our simple model. Obviously, if the R^2 were our sole criterion, we could easily inflate the fit of our predictive model by adding many more variables on the righthand-side. However, this runs the risk of "overfitting" the model and, furthermore, leads us away from being able to provide an intuitive understanding of what factors are driving our predictions.[42] A second caveat is that, because not all the directional FTA effects

[41] Models perform similarly in terms of out-of-sample predictions regardless of whether we include outliers. What this buys us is, naturally, less likelihood of extreme values when we go to predict the effects of TTIP. The Online Appendix provides details on the performance of various prediction models that we experimented with.

[42] An alternative approach would be to follow the second stage analysis more closely by running separate out-of-sample predictions with agreement- and/or agreement-by-pair fixed effects to attempt to isolate each dimension of interest. However, this would add considerable complexity to the analysis, whereas we would prefer to stick to a

we computed in our first stage are positive, our methodology can and will predict negative partial effects for at least some TTIP pairs. We do not take a stand on why we observe negative FTA effects in our estimates. Our prior would ordinarily be that TTIP should generally lead to trade creation between all pairs. Where negative values are encountered, we take this to mean that trade creation is likely to be small. Despite these limitations, we believe that our approach is able to shed light on some novel and meaningful sources of predictive power.

6.2 Predicting the Effects of TTIP

Our task in this section is to predict the effects of a U.S./E.U. FTA, given different assumptions about how it will affect trade barriers between the U.S. and its prospective partners in the E.U. In particular, we will explore two main scenarios. Under the "average" scenario, we base the change in trade barriers for all U.S./E.U. pairs on our estimate of the overall average partial effect using (5). In other words, $\widehat{\tau}_{ij}^{-\theta} = e^{\beta_{avg}} = e^{0.293}$, for all U.S./E.U. pairs. Under our second, "heterogeneous" scenario, we predict *direction-pair-specific* partial effects for TTIP using the insights from our prior analysis. In particular, we let $\widehat{\tau}_{ij}^{-\theta} = e^{\beta_{ij}}$, where β_{ij} is computed using both the coefficients estimated from our preferred second stage model as well as the ρ's shown in Fig. 4. In other words,

$$\beta_{ij} = 0.091 + 0.773 \cdot \widetilde{\beta}_{TTIP:d}, \tag{16}$$

where $\widetilde{\beta}_{TTIP:d}$ is the fitted value for each directional pair d within TTIP computed from our second stage estimates. Essentially, we are using the predictive fit from our out-of-sample validation analysis to determine how much weight we should place on our ability to predict heterogeneity in partial effects ($\rho_1 = 0.773$), versus using a common average component ($\rho_0 = 0.091$). We use this information in the interest of providing additional conservatism. The underlying regression coefficients used to compute $\widetilde{\beta}_{TTIP:d}$ are:

$$\begin{aligned}\widetilde{\beta}_{TTIP:d} = {} & 2.909 - 0.311 \cdot \ln \text{DIST}_{ij} - 0.221 \cdot \text{First stage pair FE}_{ij} \\ & + 0.145 \cdot \ln \text{GDP}_i + 0.114 \cdot \ln \text{GDP}_j - 10.861 \cdot \ln \widehat{ToT}_{A:j},\end{aligned} \tag{17}$$

relatively simple design.

which are computed using all agreements in the sample, instead of excluding one at a time. All estimates shown in (17) are statistically different from zero at the $p < 0.01$ significance level.

The predicted *partial* effects for TTIP are shown in Table 6, along with their standard errors. Notably, they are highly heterogeneous. The overall mean—0.256 if weighted by trade, 0.302 if not—is in the same ballpark as the overall average partial effect we estimated from the data (0.293). Consistent with what we saw previously with E.U. accession effects, the largest partial effects, in excess of 0.400 in each direction, involve the Eastern European E.U. members Bulgaria, Romania, and Poland (with fellow Eastern European member Hungary not far behind). The smallest values involve U.S. trade with both Ireland and Malta.

What explains the diversity in predictions across the various pairs? It cannot be variation in bilateral distance, for instance, since all European countries are collectively separated from the U.S. by the Atlantic. Actually, the major source of heterogeneity across pairs is the first stage pair fixed effect, recovered from (10), representing *ex ante* trade frictions between the different potential TTIP pairings. Ireland, for example, is already very tightly integrated with the U.S. in trade: for Ireland-U.S. trade, we obtain an *ex ante* level of trade integration (i.e., iceberg frictions raised to the minus θ) between the U.S. and Ireland of $d_{US,IR}^{-\theta} = 0.027$. While this may not, on the surface, seem like a large number, it is actually the largest of any U.S.-E.U. pair.[43] The smallest trade cost index (0.001) is for the U.S.'s trade with Cyprus, along with, unsurprisingly, its trade with Bulgaria and Romania, followed closely by its trade with Poland (0.002).[44]

Turning to asymmetries, one might expect that, as the largest participant, the U.S. should have more "market power" *ex ante* with respect to its potential E.U. partners. Actually, because the E.U. countries are very tightly integrated with one another, their terms of trade are significantly less sensitive to liberalizing with the U.S. than this logic would indicate. When we simulate the agreement using a common average effect, the $\ln \widehat{ToT}_{A:j}$ we obtain for the U.S. is 0.0036, whereas

[43] As has been observed by several authors—see, e.g., Anderson & van Wincoop (2004) and Head & Mayer (2013)—bilateral trade costs are still surprisingly large in the present day, even for nominally well-integrated countries.

[44] For a useful comparison using simple trade data: in 2006, trade with the U.S. made up only 3.9% of Bulgaria's manufacturing exports and 2.9% of its imports. For Ireland, meanwhile, the U.S. took in fully 19.7% of its exports and provided 11.6% of its imports.

the mean value we obtain for this index for the E.U. countries (0.0023) is actually lower, indicating a symmetric liberalization between the U.S. and E.U. actually tends to benefit the U.S. more than the E.U. As such, we infer that an "efficient agreement" would involve the E.U. countries granting relatively more concessions in this scenario. In addition, the fact that our coefficient on exporter GDP is slightly larger than that of importer GDP would also tend to favor the exports of the U.S., since it is the larger country in each pair.

For our general equilibrium predictions, we will stick with the simple trade model implied by (1) and (2). In particular, we will maintain that labor is the only factor of production and that trade takes place in final goods only. As shown in Head & Mayer (2014), imposing market clearing on a model of this type then delivers a standard general equilibrium system that generalizes across a wide range of different models. While this simple framework omits several factors that have been shown to be important for delivering larger gains from trade (e.g., multiple industries, trade in intermediates, dynamic effects, etc.), it is widely accepted in the literature as a benchmark for computing the general equilibrium effects of trade policies. Furthermore, it will allow us to capture the basic point that, even when an agreement has the same partial effect on all trade flows, general equilibrium outcomes can still be quite heterogeneous. The competitive equilibrium in such a model can be described by the following system of equations,

$$w_i L_i = \sum_j \pi_{ij} \cdot \left(w_j L_j + D_j\right) \quad \forall i, \tag{18}$$

where $\pi_{ij} \equiv A_i w_i^{-\theta} \tau_{ij}^{-\theta} / P_j^{-\theta}$ is the share of j's total expenditure on goods produced in origin country i. Note that we allow trade to be unbalanced. Total expenditure in j is therefore comprised of an (endogenous) labor income term, $w_j L_j$, and an (exogenous) trade balance term D_j. According to (18), the total amount of output produced in origin i, $w_i L_i$, must be equal to the sum of expenditure on goods produced in i across all destinations j.

Equation (18) can be solved (in changes) to predict *general equilibrium* effects of an FTA on both welfare and trade as a result of an FTA. To see this, first let $\widehat{x} = x'/x$ denote the equilibrium change in a variable from an initial level x to a new equilibrium level x' (i.e., the now-standard "hat algebra” notation of Dekle *et al.*, 2007). The *equilibrium in changes* version of (18) is therefore:

$$Y_i \widehat{w}_i = \widehat{w}_i^{-\theta} \sum_j \frac{\pi_{ij} \cdot \widehat{\tau}_{ij}^{-\theta}}{\widehat{P}_j^{-\theta}} \cdot \left(Y_j \widehat{w}_j + D_j\right) \quad \forall i, \tag{19}$$

where $\widehat{P}_j^{-\theta}$ can be computed as:

$$\widehat{P}_j^{-\theta} = \sum_i \pi_{ij} \widehat{w}_i^{-\theta} \widehat{\tau}_{ij}^{-\theta}. \tag{20}$$

Given initial trade shares $\{\pi_{ij}\}$, output levels $\{Y_i\}$, expenditure levels $\{E_j\}$, and a set of changes in trade barrier levels, $\{\widehat{\tau}_{ij}^{-\theta}\}$, one can solve the system defined by (19) and (20) for the resulting changes in wages $\{\widehat{w}_i\}$. With wages in hand, we then obtain the following expressions for the associated general equilibrium changes in both welfare levels and trade flows:

$$\textbf{GE Welfare Impact}: \quad \widehat{W}_i = \widehat{E}_i / \widehat{P}_i \tag{21}$$

$$\textbf{GE Trade Impact}: \quad \widehat{X}_{ij} = \frac{\widehat{w}_i^{-\theta} \widehat{\tau}_{ij}^{-\theta}}{\widehat{P}_j^{-\theta}} \cdot \widehat{E}_j, \tag{22}$$

where the change in national expenditure, $\widehat{E}_i$, is computed as $(Y_i \widehat{w}_i + D_i)/E_i$.[45] We will use the year 2006, the last year in our data, to compute the initial trade levels and trade balances.[46] Finally, since (19) is non-linear in $\widehat{w}_i$, we require an assumption regarding the trade elasticity, θ. Following the recommendations of Simonovska & Waugh (2014), we assume $\theta = 4$.

Table 7 lists the predicted *general equilibrium* effects of TTIP, both for trade and for welfare, under the two noted scenarios. As is standard in this class of models, FTAs have a larger effect on trade flows than they do welfare, as the implied welfare cost of substituting to one's own suppliers is usually relatively small.[47] The U.S., for example, experiences a large change in trade volumes - including a 7.32% increase in the value of its exports - but only a 0.38% increase in its welfare, as

[45]With balanced trade and/or multiplicative trade balances, welfare and real wages are one and the same: $\widehat{W}_i = \widehat{w}_i / \widehat{P}_i$. Naturally, the computed changes in welfare and real wages are usually similar, although they may differ noticeably for countries with large trade imbalances.

[46]The limiting factor here is data on gross output for later years, especially for the U.S. Alternatively, we could use GDP to construct internal trade for a more recent year, as has been done in other studies. The theory calls for a measure of gross sales, however (i.e., gross output, since GDP measures value added).

[47]This result is an artifact of assuming a single differentiated good with a trade elasticity of 4. If θ differs across industries, changes in trade for goods with lower values of θ can have very large welfare effects, c.f. Ossa (2015).

buyer prices in the U.S. (i.e., P_{US}) rise at more or less the same rate as U.S. wages.

Table 7 reveals several key insights we wish to focus on. First, even in the "average" scenario, where all TTIP pairs enjoy a common partial effect, the general equilibrium implications of TTIP introduce their own layer of heterogeneity. Usefully, this heterogeneity can largely be related back to a key aspect of our analysis, the level of *ex ante* trade frictions between countries. For example, the largest welfare effect is for Ireland, who enjoys a 1.35% increase in its welfare thanks to closer trade ties with the U.S. Intuitively, since Ireland already has the lowest *ex ante* trade barriers with the U.S., using a common partial effect for TTIP would eliminate a relatively larger portion of Ireland's remaining trade frictions with the U.S. than those of other E.U. members.[48] Similarly, the lowest welfare gainers under the average scenario include Cyprus, Bulgaria, Greece, Poland, and Romania—countries with which the U.S.'s *ex ante* trade relations are not as strong.

Of course, these same close relations between the U.S. and Ireland also led us to predict much smaller partial effects under the "heterogeneous" scenario. In turn, the subsequent welfare effect for Ireland is likewise predicted to be much smaller than that of the "average" scenario and is indistinguishable from zero based on the accompanying bootstrapped 95% confidence interval. The case of Ireland thus illustrates the following simple, powerful conclusion: low *ex ante* trade frictions are associated with both small partial effects *ex post* as well as larger welfare effects *ex post.* Therefore, using a common average partial effect will tend to systematically overestimate welfare gains for country-pairs who are already well-integrated in trade. A similar principle also applies in reverse. The Eastern European E.U. members Bulgaria, Hungary, Poland, and Romania are among those that see the largest improvements in welfare from introducing heterogeneous partial effects, reflecting the large partial effects we predicted for these countries in Table 6. For non-TTIP countries, general equilibrium effects are relatively similar across scenarios. As one would expect, these countries all experience mild trade diversion and most experience small welfare losses. The largest losers notably major regional trade partners on either side of the Atlantic not included in the agreement, such as the EFTA countries Norway, Iceland, and Switzerland, the U.S.'s NAFTA

[48] This echoes an observation formalized by Baier *et al.* (2018): for the same (absolute) reduction in trade frictions, countries who start out with already-close trade relations gain more in terms of welfare.

partners Canada and Mexico, as as well as the United Kingdom (owing to its presumed "Brexit" from the E.U.)

7 Conclusion

How do free trade agreements actually affect trade between member countries? And can we predict the impact of future agreements on member and non-member countries? This paper introduces a novel methodology intended to push forward our ability to answer each of these questions. Our approach not only allows us to shed light on several useful, intuitive determinants of the partial effects of FTAs, but also directly lends itself towards developing and validating an *ex ante* prediction model for predicting the effects of future agreements. Several notable aspects of the analysis include a novel set of theory-guided indices for predicting the magnitude of FTA partial effects, the ability to consider a wide variety of other possible sources of heterogeneity, and the opportunity to specifically examine determinants of heterogeneous partial effects within the same agreement.

Still, many relevant questions remain just beyond our current reach. For example, there remains only so much we can say about which FTA provisions work in favor of creating trade versus inhibiting trade. It is also widely acknowledged that economic integration agreements have consequences for investment as well as trade. The consequences for investment, too, are likely very heterogeneous across agreements and may interact with the trade-creating effects of FTAs in ways we cannot capture in our current study. Furthermore, as shown in Anderson & Yotov (2016), FTAs can have very different effects across industries and these industry-level differences in turn have important consequences for quantifying the welfare impact of FTAs. Adapting our two stage procedure to a similar industry-level perspective would be a natural extension of our methods. Including trade in agricultural products and services would make for similar improvements, especially trade in services, since services are an increasingly important component of both world trade and the objectives of new trade agreements. As new data on trade in services as well as FDI are becoming increasingly available, incorporating these various important elements will make for valuable new avenues for future research.

References

Aichele, Rahel, Felbermayr, Gabriel J., & Heiland, Inga. 2014. Going deep: The trade and welfare effects of TTIP.

Anderson, James E. 1979. A Theoretical Foundation for the Gravity Equation. *American Economic Review*, **69**(1), 106–116.

Anderson, James E., & Marcouiller, Douglas. 2002. Insecurity and the Pattern of Trade: An Empirical Investigation. *Review of Economics and Statistics*, **84**(2), 342–352.

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

Anderson, James E., & van Wincoop, Eric. 2004. Trade Costs. *Journal of Economic Literature*, **42**(3), 691–751.

Anderson, James E., & Yotov, Yoto V. 2016. Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002. *Journal of International Economics*, **99**, 279–288.

Anderson, James E., Larch, Mario, & Yotov, Yoto V. 2015. *On the Effects of the Transatlantic Trade and Investment Partnership on Trade and Growth*. Unpublished Manuscript.

Arkolakis, Costas, Costinot, Arnaud, & Rodríguez-Clare, Andrés. 2012. New Trade Models, Same Old Gains? *The American Economic Review*, **102**(1), 94–130.

Armington, Paul S. 1969. A Theory of Demand for Products Distinguished by Place of Production. *Staff Papers (International Monetary Fund)*, **16**(1), 159–178.

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

Bagwell, Kyle, & Staiger, Robert W. 1999. An Economic Theory of GATT. *The American Economic Review*, **89**(1), 215–248.

Bagwell, Kyle, & Staiger, Robert W. 2005. Multilateral Trade Negotiations, Bilateral Opportunism and the Rules of GATT/WTO. *Journal of International Economics*, **67**(1), 268–294.

Bagwell, Kyle, & Staiger, Robert W. 2011. What Do Trade Negotiators Negotiate About? Empirical Evidence from the World Trade Organization. *The American Economic Review*, **101**(4), 1238–1273.

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

Baier, Scott L., Bergstrand, Jeffrey H., & Feng, Michael. 2014. Economic Integration Agreements and the Margins of International Trade. *Journal of International Economics*, **93**(2), 339–350.

Baier, Scott L., Bergstrand, Jeffrey H., & Clance, Matthew. 2018. Heterogeneous Effects of Economic Integration Agreements. *Journal of Development Economics*, **135**.

Bergstrand, Jeffrey H., Larch, Mario, & Yotov, Yoto V. 2015. Economic Integration Agreements, Border Effects, and Distance Elasticities in the Gravity Equation. *European Economic Review*, **78**(June), 307–327.

Broda, Christian, Greenfield, Joshua, & Weinstein, David. 2006 (Sept.). *From Groundnuts to Globaliza-*

tion: A Structural Estimate of Trade and Growth. Working Paper 12512. National Bureau of Economic Research.

Broda, Christian, Limao, Nuno, & Weinstein, David E. 2008. Optimal Tariffs and Market Power: the Evidence. *American Economic Review*, **98**(5), 2032–65.

Caliendo, Lorenzo, & Parro, Fernando. 2015. Estimates of the Trade and Welfare Effects of NAFTA. *Review of Economic Studies*, **82**(1), 1–44.

Carrère, Cèline. 2006. Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model. *European Economic Review*, **50**(2), 223–247.

Carrère, Céline, Grujovic, Anja, & Robert-Nicoud, Frédéric. 2015. Trade and frictional unemployment in the global economy.

Cheng, I-hui, & Wall, Howard J. 2005. Controlling for Heterogeneity in Gravity Models of Trade and Integration. *Federal Reserve Bank of St. Louis Review*, **87**(1), 49–64.

Costinot, Arnaud, & Rodríguez-Clare, Andrés. 2014. Trade Theory with Numbers: Quantifying the Consequences of Globalization. *Pages 197–261 of:* Gopinath, Gita, Helpman, Elhanan, & Rogoff, Kenneth (eds), *Handbook of International Economics*, 4 edn., vol. 4. North Holland.

Dai, Mian, Yotov, Yoto V., & Zylkin, Thomas. 2014. On the trade-diversion effects of free trade agreements. *Economics Letters*, **122**(2), 321–325.

Dekle, Robert, Eaton, Jonathan, & Kortum, Samuel. 2007. Unbalanced Trade. *American Economic Review*, **97**(2), 351–355.

Eaton, Jonathan, & Kortum, Samuel. 2002. Technology, Geography, and Trade. *Econometrica*, **70**(5), 1741–1779.

Egger, Peter H., & Staub, Kevin E. 2015. GLM estimation of trade gravity models with fixed effects. *Empirical Economics*, **50**(1), 137–175.

Egger, Peter H., & Tarlea, Filip. 2015. Multi-Way Clustering Estimation of Standard Errors in Gravity Models. *Economics Letters*, **134**, 144–147.

Eicher, Theo S., & Henn, Christian. 2011. In search of WTO trade effects: Preferential trade agreements promote trade strongly, but unevenly. *Journal of International Economics*, **83**(2), 137–153.

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

Felbermayr, Gabriel, Heid, Benedikt, Larch, Mario, & Yalcin, Erdal. 2015. Macroeconomic Potentials of Transatlantic Free Trade: a High Resolution Perspective for Europe and the World. *Economic Policy*, **30**(83), 491–537.

Frankel, Jeffrey A., & Wei, Shang-Jin. 1997. Regionalization of World Trade and Currencies: Economics and Politics. *Pages 189–226 of: The Regionalization of the World Economy*. University of Chicago Press.

Frankel, Jeffrey A., Stein, Ernesto, & Wei, Shang-Jin. 1997. *Regional Trading Blocs in the World Economic System*. Peterson Institute.

Fugazza, Marco, & Maur, Jean-Christophe. 2008. Non-tariff barriers in CGE models: How useful for policy? *Journal of Policy Modeling*, **30**(3), 475–490.

Ghosh, Sucharita, & Yamarik, Steven. 2004. Are regional trading arrangements trade creating?: An application of extreme bounds analysis. *Journal of International Economics*, **63**(2), 369–395.

Hanushek, Eric A. 1974. Efficient Estimators for Regressing Regression Coefficients. *The American Statistician*, **28**(2), 66–67.

Head, Keith, & Mayer, Thierry. 2013. What separates us? Sources of resistance to globalization. *Canadian Journal of Economics/Revue canadienne d'économique*, **46**(4), 1196–1231.

Head, Keith, & Mayer, Thierry. 2014. Gravity Equations: Workhorse, Toolkit, and Cookbook. *Pages 131–195 of:* Gopinath, Gita, Helpman, Elhanan, & Rogoff, Kenneth (eds), *Handbook of International Economics*, 4 edn., vol. 4. North Holland.

Hofmann, Claudia, Osnago, Alberto, & Ruta, Michele. 2017. Horizontal depth: a new database on the content of preferential trade agreements.

Hummels, David, & Klenow, Peter J. 2005. The Variety and Quality of a Nation's Exports. *The American Economic Review*, **95**(3), 704–723.

Kee, Hiau Looi, Nicita, Alessandro, & Olarreaga, Marcelo. 2009. Estimating Trade Restrictiveness Indices. *The Economic Journal*, 172–199.

Kehoe, Timothy J., & Ruhl, Kim J. 2013. How Important Is the New Goods Margin in International Trade? *Journal of Political Economy*, **121**(2), 358–392.

Kehoe, Timothy J., Rossbach, Jack, & Ruhl, Kim J. 2015. Using the new products margin to predict the industry-level impact of trade reform. *Journal of International Economics*, **96**(2), 289–297.

Kohl, Tristan. 2014. Do we really know that trade agreements increase trade? *Review of World Economics*, **150**(3), 443–469.

Kohl, Tristan, Brakman, Steven, & Garretsen, Harry. 2016. Do Trade Agreements Stimulate International Trade Differently? Evidence from 296 Trade Agreements. *The World Economy*, **39**(1), 97–131.

Krugman, Paul. 1980. Scale Economies, Product Differentiation, and the Pattern of Trade. *The American Economic Review*, **70**(5), 950–959.

Larch, Mario, Wanner, Joschka, Yotov, Yoto, & Zylkin, Thomas. 2017. The Currency Union Effect: A PPML Re-assessment with High-Dimensional Fixed Effects. Apr.

Lewis, Jeffrey B., & Linzer, Drew A. 2005. Estimating regression models in which the dependent variable is based on estimates. *Political Analysis*, **13**(4), 345–364.

Limão, Nuno. 2007. Are Preferential Trade Agreements with Non-trade Objectives a Stumbling Block for Multilateral Liberalization? *The Review of Economic Studies*, **74**(3), 821–855.

Limão, Nuno. 2016. Preferential Trade Agreements. *Pages 279–367 of: Handbook of Commercial Policy*, vol. 1. Elsevier.

Ludema, Rodney D., & Mayda, Anna Maria. 2013. Do terms-of-trade effects matter for trade agreements? Theory and evidence from WTO Countries*. *The Quarterly Journal of Economics*, **128**(4), 1837–1893.

Magee, Christopher S. P. 2008. New measures of trade creation and trade diversion. *Journal of International Economics*, **75**(2), 349–362.

Manova, Kalina. 2013. Credit Constraints, Heterogeneous Firms, and International Trade. *Review of Economic Studies*, **80**(2), 711–744.

Mansfield, Edward D., Milner, Helen V., & Pevehouse, Jon C. 2007. Vetoing Co-operation: The Impact of Veto Players on Preferential Trading Arrangements. *British Journal of Political Science*, **37**(03), 403–432.

Melitz, Marc J. 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, **71**(6), 1695–1725.

Melitz, Marc J., & Ottaviano, Gianmarco I. P. 2008. Market Size, Trade, and Productivity. *Review of Economic Studies*, **75**(1), 295–316.

Nicita, Alessandro, & Olarreaga, Marcelo. 2007. Trade, Production, and Protection Database, 1976–2004. *World Bank Econ Rev*, **21**(1), 165–171.

Nicita, Alessandro, Olarreaga, Marcelo, & Silva, Peri. 2018. Cooperation in WTO's Tariff Waters? *Journal of Political Economy*, Jan.

Ossa, Ralph. 2015. Why trade matters after all. *Journal of International Economics*, **97**(2), 266–277.

Piermartini, Roberta, & Yotov, Yoto. 2016. *Estimating Trade Policy Effects with Structural Gravity*. Working Paper ERSD-2016-10. WTO.

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

Santos Silva, J. M. C., & Tenreyro, Silvana. 2011. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, **112**(2), 220–222.

Simonovska, Ina, & Waugh, Michael E. 2014. The elasticity of trade: Estimates and evidence. *Journal of International Economics*, **92**(1), 34–50.

Soloaga, Isidro, & Winters, L. Alan. 2001. Regionalism in the nineties: What effect on trade? *The North American Journal of Economics and Finance*, **12**(1), 1–29.

Staiger, Robert W. 2012 (Jan.). *Non-Tariff Measures and the WTO*. SSRN Scholarly Paper ID 1998738. Social Science Research Network, Rochester, NY.

Staiger, Robert W., & Sykes, Alan O. 2011. International Trade, National Treatment, and Domestic Regulation. *The Journal of Legal Studies*, **40**(1), 149–203.

Tinbergen, Jan. 1962. An analysis of world trade flows. *Pages 1–117 of: Shaping the World Economy*. New York: Twentieth Century Fund.

White, Halbert. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, **48**(4), 817–838.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yotov, Yoto V. 2012. A Simple Solution to the Distance Puzzle in International Trade. *Economics Letters*, **117**(3), 794–798.

Zylkin, Thomas. 2015. *Beyond Tariffs: Quantifying Heterogeneity in the Effects of Free Trade Agreements*. mimeo.

Figures

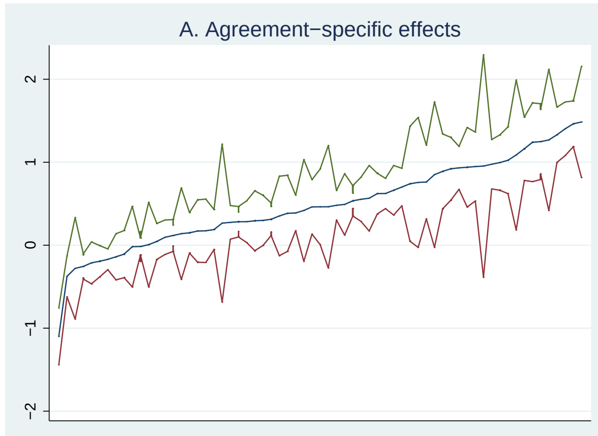

Figure 1: Distribution of Agreement-level FTA Partial Effects, with 95% CIs

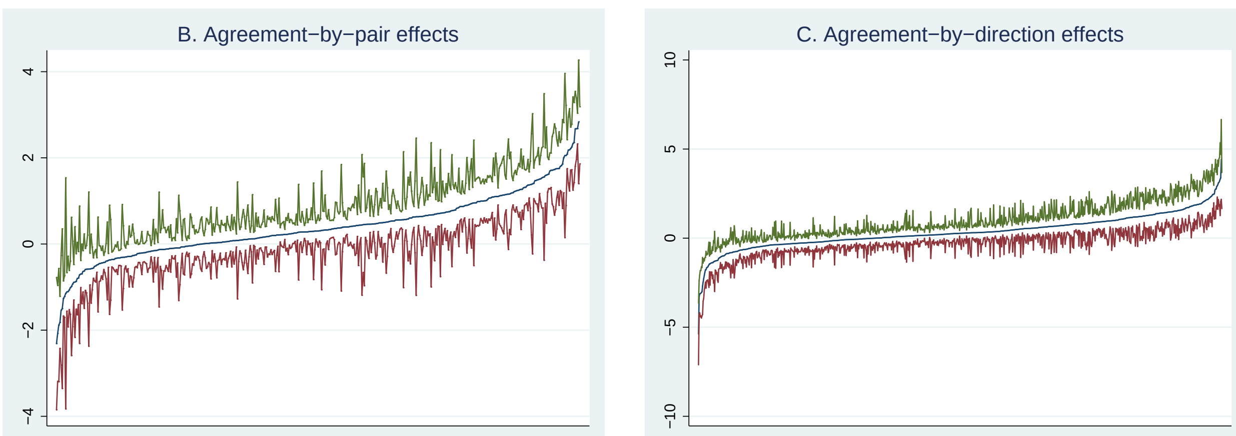

Figure 2: Distributions of Agreement-pair-specific (left) and Direction-specific (right) FTA Effects, with 95% CIs

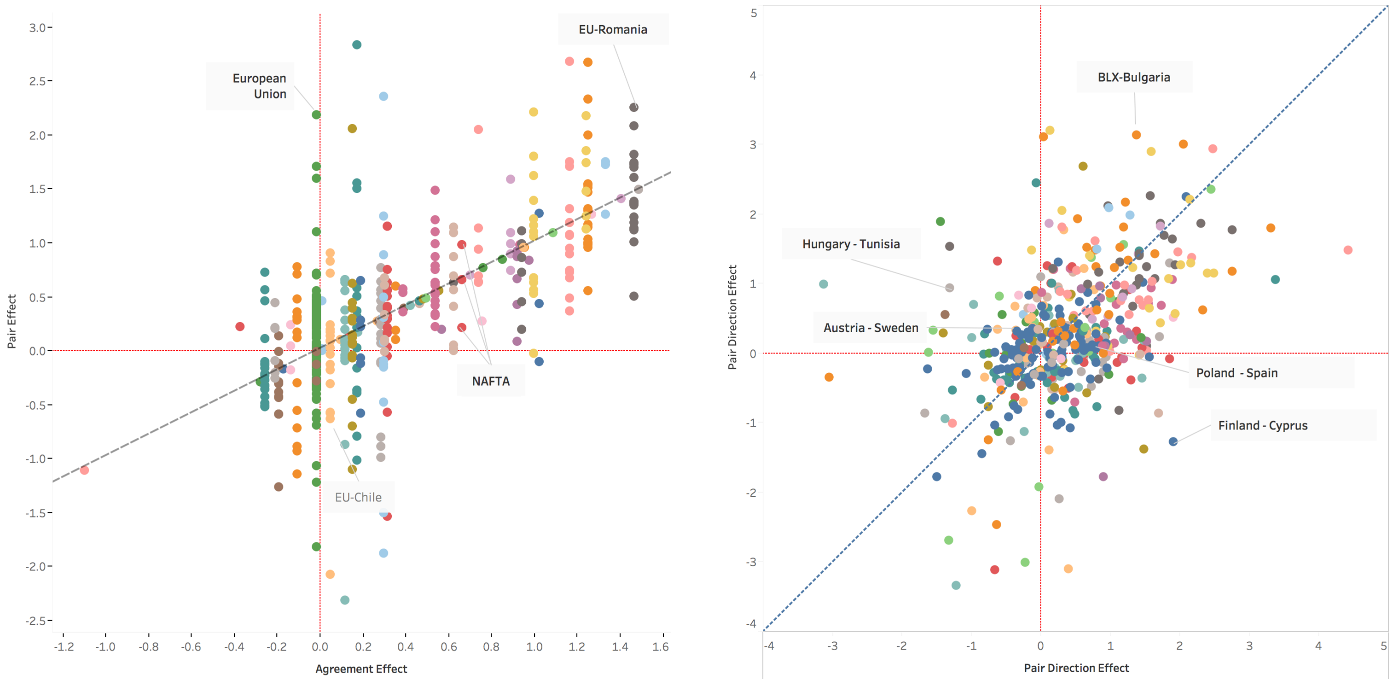

Figure 3: Pair-wise Heterogeneity within Agreements (left); Directional Heterogeneity within Pairs (right)

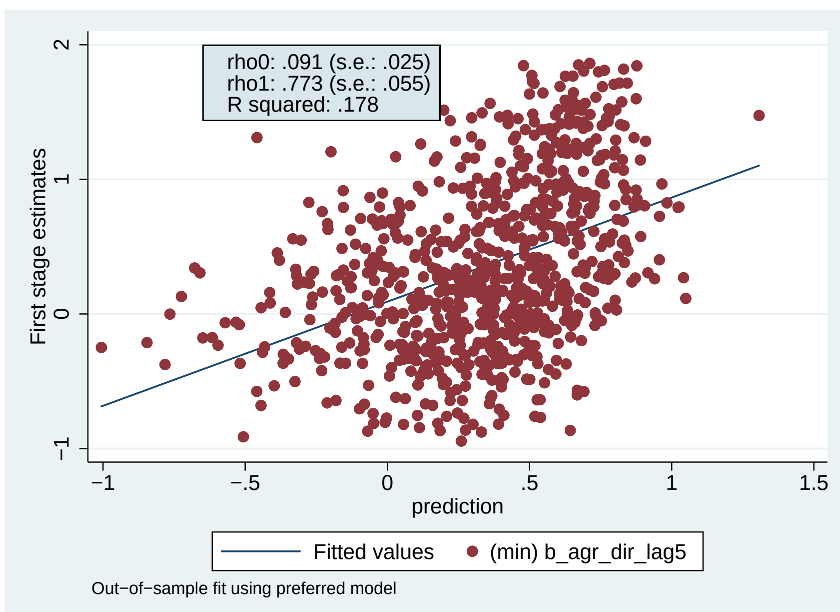

Figure 4: Out-of-sample Validation Results

Tables

Table 1: Estimates of Agreement-Specific FTA Effects

Agreement	β_A	$s.e.$	Agreement	β_A	$s.e.$	Agreement	β_A	$s.e.$
Positive effects:			(*cont'd*)			(*cont'd*)		
Bulgaria-Turkey†	1.485	0.342	Israel-Mexico	0.553	0.136	Israel-Romania	0.174	0.195
EU-Romania†	1.463	0.142	EU-Turkey†	0.535	0.083	Pan Arab FTA	0.171	0.192
Romania-Turkey†	1.403	0.165	Canada-Costa Rica	0.492	0.189	EU-Egypt	0.149	0.125
Andean Community†	1.331	0.170	Canada-Israel†	0.481	0.091	Australia-Singapore	0.139	0.282
Israel-Turkey†	1.269	0.434	Egypt-Turkey	0.463	0.232	EU-Morocco	0.117	0.090
EU-Bulgaria†	1.248	0.225	Chile-China	0.462	0.167	Morocco-US	0.096	0.106
CEFTA†	1.240	0.242	Tunisia-Turkey	0.389	0.109	EU-Chile†	0.045	0.111
EU-Poland†	1.162	0.195	EU-Mexico	0.313	0.095	EU†	-0.016	0.066
Costa Rica-Mexico	1.087	0.461	Chile-US	0.283	0.128	Mercosur-Bolivia	0.007	0.260
Mercosur†	1.024	0.205	EU-Tunisia	0.283	0.086	EFTA-Singapore	-0.018	0.248
EU-Hungary†	0.996	0.170	Chile-South Korea	0.275	0.103	ASEAN†	-0.107	0.145
Poland-Turkey†	0.976	0.152				EFTA-Mexico†	-0.140	0.142
Bulgaria-Israel†	0.948	0.212	**Insignificant effects** ($p > .05$):			EFTA-Israel†	-0.213	0.129
EFTA-Hungary†	0.939	0.244	Jordan-US	0.954	0.684	Singapore-US	-0.279	0.312
Hungary-Turkey†	0.932	0.132	Canada-Chile	0.851	0.447			
EFTA-Poland†	0.921	0.193	Hungary-Israel	0.757	0.400	**Negative effects**:		
EFTA-Romania†	0.890	0.230	Mexico-Uruguay	0.463	0.377	Australia-US†	-0.170	0.064
Colombia-Mexico†	0.762	0.226	Chile-Costa Rica	0.419	0.313	EU-Cyprus†	-0.194	0.096
EFTA-Bulgaria	0.740	0.353	EFTA-Morocco	0.384	0.234	EU-Israel†	-0.256	0.080
Japan-Mexico†	0.701	0.115	Mercosur-Chile	0.353	0.244	Canada-US†	-0.375	0.126
NAFTA†	0.662	0.152	EFTA-Turkey	0.299	0.154	Chile-Singapore†	-1.099	0.174
Australia-Thailand†	0.623	0.093	EU-EFTA	0.294	0.184			
Mercosur-Andean†	0.622	0.125	Chile-Mexico	0.266	0.486			
Israel-Poland	0.566	0.202	Agadir Agreement	0.188	0.123			

Summary statistics:			
Simple		*Weighted Averages*	
Median β^A estimate:	0..463	by inverse variance:	0.293
Mean β^A estimate:	0.491	by number of country-pairs	0.382
Variance of estimates:	0.261	by (#pairs×inv. var):	0.200

This table reports estimates of the partial FTA effects for all agreements in our sample. Standard errors are "three-way" clustered by exporter, importer, and year. † denotes estimates that are statistically different from the overall average estimate of $\beta = 0.293$. There are 33 such estimates, or 50.8%. See text for further details.

Table 2: Heterogeneity in EU Accession Effects

Pair	$\beta_{EU:p}$	*s.e.*	Pair	$\beta_{EU:p}$	*s.e.*	Pair	$\beta_{EU:p}$	*s.e.*
Largest EU Accession Effects (*by pair*):								
Hungary-Poland*†	2.186	0.487	Cyprus-Hungary*	0.503	0.251	Spain-Poland	0.412	0.228
Cyprus-Finland*†	1.711	0.399	BLX-Cyprus*	0.493	0.176	U.K.-Hungary	0.400	0.218
Hungary-Malta*†	1.600	0.571	Finland-Hungary	0.470	0.418	Austria-BLX*	0.394	0.086
Austria-Malta*	1.101	0.514	U.K.-Poland*	0.469	0.225	Austria-Spain*	0.375	0.191
Cyprus-Netherlands*†	0.716	0.135	Cyprus-Greece*	0.457	0.196	Italy-Poland*	0.370	0.168
Cyprus-U.K.	0.703	0.370	Cyprus-Germany*	0.456	0.153	BLX-Finland*	0.352	0.100
Cyprus-Italy*	0.555	0.139	Denmark-Hungary*	0.437	0.149	Austria-Poland*	0.352	0.105
France-Poland*	0.517	0.147	BLX-Sweden*	0.431	0.129	Germany-Poland*	0.334	0.110
Small and Negative EU Accession Effects (*by pair*):								
Austria-Sweden†	-0.202	0.107	Cyprus-Malta	-0.307	0.371	Denmark-Finland*†	-0.443	0.150
Germany-Malta†	-0.205	0.112	Finland-Sweden*†	-0.312	0.102	Cyprus-Denmark*†	-0.455	0.213
Greece-Sweden†	-0.210	0.255	Denmark-Malta†	-0.327	0.193	Italy-Malta†	-0.584	0.403
U.K.-Sweden†	-0.213	0.143	Finland-U.K.*†	-0.331	0.133	Finland-Portugal†	-0.630	0.441
Germany-Sweden*†	-0.220	0.107	Cyprus-Ireland*†	-0.334	0.159	Portugal-Sweden*†	-0.694	0.353
Finland-Italy†	-0.256	0.136	Finland-Ireland*†	-0.356	0.125	Ireland-Malta*†	-1.069	0.232
Hungary-Ireland†	-0.269	0.264	Italy-Sweden*†	-0.360	0.087	Cyprus-Poland*†	-1.220	0.247
Ireland-Sweden*†	-0.291	0.144	BLX-Hungary†	-0.399	0.210	Greece-Malta*†	-1.819	0.308
Pair	$\beta_{EU:d}$	*s.e.*	Pair	$\beta_{EU:d}$	*s.e.*	Pair	$\beta_{EU:d}$	*s.e.*
Examples of Asymmetric EU Accession Effects *(arrows indicate direction of trade)*:								
Netherlands →Austria*	0.418	0.158	Spain →Austria*†	0.734	0.212	Poland →Austria*	0.575	0.186
Austria →Netherlands*†	-0.486	0.162	Austria →Spain	0.110	0.187	Austria →Poland	0.157	0.152
Poland →Spain*	0.795	0.258	Poland →Sweden*	0.549	0.223	Poland →Netherlands	0.364	0.193
Spain →Poland	0.057	0.183	Sweden →Poland	0.040	0.156	Netherlands →Poland	0.157	0.158
Poland →U.K.*†	0.825	0.266	U.K. →Sweden	0.048	0.197	Sweden →Austria	-0.105	0.224
U.K. →Poland	0.100	0.184	Sweden →U.K.*†	-0.431	0.182	Austria →Sweden†	-0.293	0.239
Netherlands →Sweden	0.317	0.199	Spain →Sweden	0.285	0.209	U.K. →Austria*	0.342	0.141
Sweden →Netherlands†	-0.353	0.200	Sweden →Spain†	-0.152	0.207	Austria →U.K.	0.197	0.147

Summary of within-EU estimates

Pairwise estimates ($\beta_{EU:p}$)			*Directional estimates* ($\beta_{EU:d}$)		
Mean: 0.047	Median: 0.046	s.d.: 0.514	Mean: 0.085	Median: 0.048	s.d.: 0574
# positive and significant:		27/98 (27.6%)	# positive and significant:		37/196 (18.9%)
# negative and significant:		13/98 (13.3%)	# negative and significant:		26/196 (13.3%)
# statistically different from $\beta = 0.293$:		41/98 (41.8%)	# statistically different from $\beta = 0.293$:		82/196 (41.8%)

This table reports examples of pair-specific and asymmetric estimated partial effects for the EU accessions in our sample. * denotes estimates that are statistically different from 0. † denotes estimates that are statistically different from the overall average estimate of $\beta = 0.293$. See text for further details.

Table 3: Second Stage Estimates: Baseline

	Dependent variable: First stage directional FTA estimates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First stage pair FE†	-0.195***	-0.372***	-0.261***		-0.154***	-0.346***	-0.246***	
	(0.034)	(0.041)	(0.057)		(0.036)	(0.041)	(0.056)	
ln $\widehat{ToT}_{A:j}$	-17.424***	-10.786***	-7.616**	-9.423***	-19.392***	-13.557***	-5.881*	-7.978**
	(2.483)	(2.995)	(3.193)	(3.284)	(2.871)	(3.120)	(3.245)	(3.241)
Extensive margin of trade	1.167***	0.555**	0.349	-0.573*	0.932***	0.402	0.249	-0.697**
	(0.170)	(0.263)	(0.246)	(0.303)	(0.174)	(0.268)	(0.253)	(0.334)
Count Enf. Provisions					0.484*	0.242		
					(0.256)	(0.324)		
Institutional Quality					0.493*	0.631*		
					(0.296)	(0.328)		
Customs Union					0.005	0.053		
					(0.087)	(0.073)		
ln 1+Applied Tariff					1.200**	-0.228	-0.736	-1.008
					(0.529)	(0.572)	(0.710)	(0.647)
ln DIST		-0.456***	-0.241***			-0.474***	-0.257***	
		(0.044)	(0.081)			(0.057)	(0.085)	
CONTIG		0.188**	-0.003			0.144	-0.044	
		(0.093)	(0.103)			(0.104)	(0.104)	
COLONY		0.014	0.076			-0.050	0.120	
		(0.100)	(0.111)			(0.102)	(0.115)	
LANG		0.112	0.073			0.264***	0.032	
		(0.082)	(0.093)			(0.091)	(0.098)	
LEGAL		0.161**	0.262***			0.190**	0.248***	
		(0.075)	(0.080)			(0.076)	(0.080)	
Prior Agreement		-0.262***	-0.000			-0.235***	0.028	
		(0.055)	(0.085)			(0.072)	(0.086)	
Exporter (log) Real GDP		0.202***	0.152***			0.205***	0.169***	
		(0.035)	(0.038)			(0.037)	(0.040)	
Importer (log) Real GDP		0.146***	0.091***	-0.139***		0.111***	0.069**	-0.182***
		(0.026)	(0.030)	(0.043)		(0.028)	(0.030)	(0.048)
Exporter (log) GDP per capita		-0.067	0.002			-0.059	0.071	
		(0.083)	(0.108)			(0.099)	(0.136)	
Importer (log) GDP per capita		0.172***	0.235***	0.117		0.145**	0.271**	0.071
		(0.048)	(0.086)	(0.079)		(0.070)	(0.116)	(0.098)
Constant	0.349***	0.349***			0.349***	0.363***		
	(0.029)	(0.026)			(0.032)	(0.028)		
Agreement FEs			x				x	
Agr.×pair FEs				x				x
Observations	908	908	908	908	826	826	852	806
R^2	0.078	0.261	0.424	0.729	0.099	0.270	0.427	0.740
Within R^2			0.107	0.086			0.104	0.118

This table reports second stage OLS estimates with robust standard errors (reported in parentheses). The dependent variable is $\beta_{A:d}$, an estimated direction-specific FTA partial effect which we have estimated in a prior stage. Observation counts vary in columns 5-8 versus columns 1-4 because of the limited availability of tariff data (from TRAINS) and information on the number of enforceable provisions and institutional quality for each FTA (from Kohl *et al.*, 2016). In columns 1, 2, 4, and 5, all variables are de-meaned with respect to their within-sample mean. This allows us to interpret the regression constant as reflecting the overall average FTA estimate after netting out the average effects of each of the included covariates. * $p < 0.10$, ** $p < .05$, *** $p < .01$. † Also accounts for "globalization" effects. See text for further details.

Table 4: Second Stage Estimates: Institutions and Factor Endowments

	Dependent variable: First stage directional FTA estimates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First stage pair FE†	-0.160***	-0.296***	-0.263***		-0.156***	-0.257***	-0.233***	
	(0.038)	(0.047)	(0.072)		(0.036)	(0.042)	(0.060)	
$\ln \widehat{ToT}_{A:j}$	-17.127***	-12.921***	-8.662***	-9.347***	-5.373**	-5.988**	-6.404**	-7.939***
	(3.175)	(3.186)	(3.297)	(3.547)	(2.530)	(2.640)	(2.974)	(2.805)
Extensive margin of trade	1.383***	0.725**	0.252	-0.362	1.103***	0.268	0.301	-0.253
	(0.182)	(0.281)	(0.274)	(0.346)	(0.185)	(0.266)	(0.282)	(0.315)
Exporter Democracy	0.046***	0.054***	0.055					
	(0.015)	(0.015)	(0.035)					
Importer Democracy	0.050***	0.049***	0.049	-0.007				
	(0.012)	(0.012)	(0.032)	(0.019)				
Exporter Bureaucracy Quality	-0.459***	-0.261***	-0.023					
	(0.045)	(0.056)	(0.067)					
Importer Bureaucracy Quality	-0.268***	-0.170***	0.060	0.074				
	(0.044)	(0.062)	(0.070)	(0.064)				
Exporter Checks and Balances	0.057***	0.045***	0.020					
	(0.018)	(0.017)	(0.018)					
Importer Checks and Balances	0.038**	0.016	0.001	-0.018				
	(0.019)	(0.018)	(0.019)	(0.019)				
Exporter Rule of Law	0.089***	0.060**	-0.019					
	(0.027)	(0.029)	(0.042)					
Importer Rule of Law	0.191***	0.139***	0.058	0.066				
	(0.030)	(0.029)	(0.044)	(0.040)				
Exp. (log) Physical Capital / Labor ratio					-0.461***	-0.260***	-0.082	
					(0.054)	(0.088)	(0.152)	
Imp. (log) Physical Capital / Labor ratio					-0.188***	-0.151	0.027	0.088
					(0.050)	(0.098)	(0.140)	(0.106)
Exp. (log) Human Capital / Labor ratio					0.423*	0.228	0.240	
					(0.235)	(0.250)	(0.282)	
Imp. (log) Human Capital / Labor ratio					0.908***	0.525**	0.547*	0.313
					(0.241)	(0.250)	(0.308)	(0.246)
$\lvert\Delta$(log) Physical Capital / Labor ratio$\rvert$					0.419***	0.175**	-0.205	
					(0.073)	(0.075)	(0.141)	
$\lvert\Delta$(log) Human Capital / Labor ratio$\rvert$					0.424*	0.064	-0.188	
					(0.240)	(0.244)	(0.366)	
$\lvert\Delta$(log) GDP per capita$\rvert$					-0.429***	-0.125	0.334	
					(0.098)	(0.100)	(0.206)	
Constant	0.351***	0.325***			0.321***	0.415***		
	(0.028)	(0.029)			(0.034)	(0.044)		
Gravity controls		x	x	x		x	x	x
Agreement FEs			x				x	
Agr.×pair FEs				x				x
Observations	826	826	826	826	716	716	716	716
R^2	0.268	0.349	0.472	0.733	0.332	0.416	0.502	0.774
Within R^2			0.139	0.095			0.152	0.132

This table reports second stage OLS estimates with robust standard errors (reported in parentheses). The dependent variable is $\beta_{A:d}$, an estimated direction-specific FTA partial effect which we have estimated in a prior stage. Observation counts vary due to the limited availability of factor endowment information (from PWT 9.0) and data on country institutions (from ICRG). In columns 1, 2, 4, and 5, all variables are de-meaned with respect to their within-sample mean. This allows us to interpret the regression constant as reflecting the overall average FTA estimate after netting out the average effects of each of the included covariates. "Gravity variables" (suppressed for brevity) include ln DIST, CONTIG, COLONY, LANG, LEGAL, Prior Agreement, GDPs, and per capita GDPs. Full results available on request. * $p < 0.10$, ** $p < .05$, *** $p < .01$. † Also accounts for "globalization" effects. See text for further details.

Table 5: Second Stage Estimates: Anticipation Effects

	Dependent variable: First stage directional FTA estimates								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First stage pair FE †	-0.350***	-0.280***	-0.289***	-0.224***	-0.221***	-0.234***			
	(0.035)	(0.032)	(0.030)	(0.047)	(0.044)	(0.039)			
$\ln \widehat{ToT}_{A:j}$	-11.286***	-9.082***	-8.989***	-7.525**	-5.769**	-5.745**	-9.586***	-6.894***	-6.940***
	(2.983)	(2.478)	(2.471)	(3.288)	(2.680)	(2.709)	(3.387)	(2.560)	(2.609)
Extensive margin of trade	0.491**	0.347*	0.365*	0.248	0.152	0.179	-0.618**	-0.460*	-0.457*
	(0.248)	(0.208)	(0.201)	(0.242)	(0.212)	(0.214)	(0.294)	(0.247)	(0.249)
ln DIST	-0.451***	-0.316***	-0.317***	-0.246***	-0.250***	-0.248***			
	(0.044)	(0.036)	(0.036)	(0.081)	(0.080)	(0.079)			
CONTIG	0.191**	0.105	0.126*	-0.046	-0.056	-0.033			
	(0.097)	(0.074)	(0.077)	(0.105)	(0.090)	(0.090)			
COLONY	0.021	-0.024	-0.028	0.102	0.154	0.161			
	(0.103)	(0.085)	(0.085)	(0.111)	(0.110)	(0.108)			
LANG	0.097	0.054	0.065	0.041	-0.053	-0.052			
	(0.085)	(0.071)	(0.071)	(0.094)	(0.088)	(0.085)			
LEGAL	0.147**	0.136**	0.131**	0.245***	0.209***	0.207***			
	(0.072)	(0.061)	(0.059)	(0.079)	(0.070)	(0.069)			
Prior Agreement	-0.291***	-0.207***	-0.222***	0.005	-0.008	-0.020			
	(0.060)	(0.049)	(0.052)	(0.083)	(0.072)	(0.071)			
Exporter (log) Real GDP	0.212***	0.112***	0.110***	0.164***	0.093***	0.088***			
	(0.036)	(0.027)	(0.027)	(0.039)	(0.032)	(0.032)			
Importer (log) Real GDP	0.124***	0.083***	0.086***	0.071**	0.055**	0.057**	-0.172***	-0.086***	-0.087***
	(0.026)	(0.022)	(0.022)	(0.029)	(0.025)	(0.024)	(0.043)	(0.032)	(0.032)
Exporter (log) GDP per capita	-0.060	-0.025	-0.027	-0.030	0.033	0.037			
	(0.082)	(0.059)	(0.053)	(0.106)	(0.094)	(0.090)			
Importer (log) GDP per capita	0.193***	0.158***	0.157***	0.214**	0.217***	0.221***	0.138*	0.102*	0.103*
	(0.048)	(0.036)	(0.035)	(0.086)	(0.076)	(0.074)	(0.079)	(0.060)	(0.060)
ln resid_{t_0-6}	-0.007		0.082	-0.007		0.077	-0.095		-0.023
	(0.087)		(0.066)	(0.087)		(0.069)	(0.120)		(0.114)
ln resid_{t_0-5}	0.074		0.046	0.127		0.086	0.068		0.001
	(0.164)		(0.098)	(0.161)		(0.099)	(0.203)		(0.149)
ln resid_{t_0-4}	0.024		-0.156	-0.017		-0.125	-0.019		0.010
	(0.166)		(0.126)	(0.163)		(0.124)	(0.197)		(0.151)
ln resid_{t_0-3}		0.165**	0.227***		0.147**	0.185**		0.103	0.095
		(0.068)	(0.073)		(0.065)	(0.073)		(0.099)	(0.131)
ln resid_{t_0-2}		0.267***	0.269***		0.259***	0.262***		0.262**	0.263***
		(0.082)	(0.085)		(0.082)	(0.085)		(0.104)	(0.097)
ln resid_{t_0-1}		0.654***	0.642***		0.584***	0.577***		0.656***	0.658***
		(0.103)	(0.098)		(0.108)	(0.100)		(0.118)	(0.105)
Constant	0.355***	0.356***	0.356***						
	(0.027)	(0.022)	(0.022)						
Agreement FEs		x			x			x	
Agr.×pair FEs			x			x			x
Observations	874	876	874	874	876	874	874	876	874
R^2	0.262	0.508	0.515	0.433	0.606	0.612	0.742	0.824	0.824
Within R^2				0.114	0.385	0.394	0.119	0.399	0.399

This table reports second stage OLS estimates with robust standard errors (in parentheses). "ln resid_{t_0-m}" is defined as $\ln(X_{ij,t_0-m}/\widehat{X}_{ij,t_0-m})$, where "$X_{ij,t_0-m}$" is the volume of trade sent from i to j in the year t_0-m leading up to an agreement which begins in year t_0 and "$\widehat{X}_{ij,t_0-m}$" is its fitted value from the first stage estimation. Observation counts vary because we do not observe these residuals for the earliest agreements in the sample. In columns 1-3, all variables are de-meaned with respect to their within-sample mean. This allows us to interpret the regression constant as reflecting the overall average FTA estimate after netting out the average effects of each of the included covariates. * $p < 0.10$, ** $p < .05$, *** $p < .01$. † Also accounts for "globalization" effects. See text for further details.

Table 6: Predicting the Partial Effects of TTIP

Exporter	Importer	$\beta_{TTIP:d}$	Exporter	Importer	$\beta_{TTIP:d}$
Predicted partial effects of TTIP (*by TTIP pair*)					
United States	Austria	0.319 (0.220, 0.417)	Austria	United States	0.219 (0.126, 0.313)
United States	Belgium-Lux.	0.249 (0.147, 0.350)	Belgium-Lux.	United States	0.147 (0.055, 0.239)
United States	Bulgaria	0.536 (0.428, 0.645)	Bulgaria	United States	0.397 (0.284, 0.510)
United States	Cyprus	0.399 (0.277, 0.522)	Cyprus	United States	0.218 (0.088, 0.348)
United States	Denmark	0.315 (0.216, 0.414)	Denmark	United States	0.207 (0.113, 0.300)
United States	Finland	0.335 (0.235, 0.435)	Finland	United States	0.218 (0.123, 0.313)
United States	France	0.346 (0.245, 0.448)	France	United States	0.290 (0.195, 0.386)
United States	Germany	0.305 (0.201, 0.410)	Germany	United States	0.264 (0.165, 0.364)
United States	Greece	0.467 (0.365, 0.569)	Greece	United States	0.352 (0.252, 0.452)
United States	Hungary	0.399 (0.298, 0.501)	Hungary	United States	0.276 (0.178, 0.374)
United States	Ireland	0.051 (-0.049, 0.152)	Ireland	United States	0.002 (-0.095, 0.100)
United States	Italy	0.368 (0.266, 0.471)	Italy	United States	0.303 (0.206, 0.399)
United States	Malta	0.104 (-0.027, 0.235)	Malta	United States	-0.067 (-0.202, 0.069)
United States	Netherlands	0.198 (0.099, 0.297)	Netherlands	United States	0.140 (0.048, 0.233)
United States	Poland	0.587 (0.485, 0.689)	Poland	United States	0.488 (0.388, 0.588)
United States	Portugal	0.456 (0.358, 0.554)	Portugal	United States	0.342 (0.248, 0.436)
United States	Romania	0.539 (0.435, 0.642)	Romania	United States	0.421 (0.317, 0.525)
United States	Spain	0.479 (0.379, 0.579)	Spain	United States	0.402 (0.308, 0.496)
United States	Sweden	0.248 (0.148, 0.348)	Sweden	United States	0.153 (0.060, 0.245)
Summary statistics:					
Simple			*Trade-weighted*		
Median $\beta_{TTIP:d}$ prediction:		0.310	Median $\beta_{TTIP:d}$ prediction:		0.264
Mean $\beta_{TTIP:d}$ prediction:		0.302	Mean $\beta_{TTIP:d}$ prediction:		0.256
Standard deviation:		0.149	Standard deviation:		0.103

This table reports the predicted partial TTIP effects, $\beta_{TTIP:d}$, for all pairs of countries affected by TTIP. The United Kingdom is not included in TTIP. Trade frictions between EU countries are assumed to not be affected. The numbers shown in parentheses are 95% confidence intervals. See text for further details.

Table 7: General Equilibrium Predictions for the Effects of an E.U./U.S. FTA

	Percentage changes in trade and welfare, by country					
	"Average" Scenario			*"Heterogeneous" Scenario*		
	Δ% Exports	Δ% Imports	Δ% Welfare	Δ% Exports	Δ% Imports	Δ% Welfare
Austria	1.18	1.14	0.24	1.28 (0.93, 1.86)	1.23 (0.90, 1.79)	0.22 (0.17, 0.31)
Belgium-Luxembourg	1.58	1.59	0.20	1.19 (0.80, 1.88)	1.20 (0.81, 1.89)	0.12 (0.08, 0.20)
Bulgaria	0.81	0.59	0.06	1.71 (1.41, 2.33)	1.25 (1.03, 1.70)	0.14 (0.11, 0.20)
Canada	-0.69	-0.67	-0.10	-0.81 (-1.24, -0.52)	-0.78 (-1.20, -0.50)	-0.10 (-0.15, -0.07)
Cyprus	0.74	0.20	0.02	0.88 (0.69, 1.26)	0.24 (0.19, 0.35)	-0.02 (-0.08, 0.05)
Denmark	1.34	1.21	0.23	1.42 (1.04, 2.07)	1.28 (0.93, 1.86)	0.21 (0.17, 0.30)
Finland	1.34	1.60	0.20	1.48 (1.14, 2.11)	1.77 (1.36, 2.53)	0.21 (0.17, 0.29)
France	1.86	1.76	0.22	2.22 (1.69, 3.22)	2.10 (1.60, 3.04)	0.25 (0.19, 0.37)
Germany	1.90	2.59	0.36	1.98 (1.40, 3.03)	2.70 (1.90, 4.14)	0.37 (0.27, 0.57)
Greece	1.82	0.54	0.02	3.20 (2.64, 4.36)	0.96 (0.79, 1.30)	0.02 (-0.02, 0.08)
Hungary	0.72	0.70	0.11	1.07 (0.81, 1.53)	1.05 (0.80, 1.50)	0.15 (0.13, 0.20)
Iceland	-0.45	-0.17	-0.06	-0.15 (-0.41, 0.07)	-0.05 (-0.15, 0.03)	-0.07 (-0.11, -0.04)
Ireland	2.62	4.54	1.35	0.26 (-0.47, 1.28)	0.46 (-0.81, 2.22)	0.13 (-0.24, 0.65)
Israel	-0.57	-0.63	-0.06	-0.65 (-1.00, -0.42)	-0.73 (-1.11, -0.47)	-0.07 (-0.11, -0.05)
Italy	1.69	1.88	0.15	2.21 (1.71, 3.17)	2.44 (1.89, 3.51)	0.19 (0.15, 0.27)
Malta	2.06	1.23	0.29	0.26 (-0.43, 1.19)	0.15 (-0.26, 0.71)	-0.09 (-0.18, 0.03)
Mexico	-0.75	-0.75	-0.11	-0.91 (-1.41, -0.58)	-0.90 (-1.39, -0.58)	-0.12 (-0.18, -0.08)
Morocco	-0.15	-0.08	-0.05	-0.05 (-0.21, 0.10)	-0.03 (-0.11, 0.05)	-0.07 (-0.10, -0.04)
Netherlands	1.58	1.60	0.54	1.02 (0.58, 1.70)	1.03 (0.59, 1.71)	0.29 (0.15, 0.52)
Norway	-0.30	-0.18	-0.08	-0.18 (-0.34, -0.07)	-0.11 (-0.21, -0.04)	-0.08 (-0.13, -0.05)
Poland	0.71	0.58	0.04	1.78 (1.47, 2.40)	1.44 (1.19, 1.94)	0.13 (0.11, 0.18)
Portugal	1.35	0.94	0.09	2.33 (1.93, 3.17)	1.63 (1.35, 2.22)	0.15 (0.13, 0.20)
Romania	0.91	0.56	0.04	1.93 (1.59, 2.62)	1.20 (0.99, 1.63)	0.10 (0.07, 0.15)
Spain	1.27	0.88	0.07	2.39 (1.98, 3.24)	1.64 (1.36, 2.23)	0.13 (0.10, 0.18)
Sweden	1.58	1.99	0.33	1.21 (0.77, 1.93)	1.52 (0.97, 2.43)	0.23 (0.15, 0.38)
Switzerland	-0.31	-0.32	-0.05	-0.25 (-0.40, -0.17)	-0.26 (-0.42, -0.18)	-0.05 (-0.07, -0.03)
Tunisia	-0.07	-0.05	-0.06	0.05 (-0.13, 0.22)	0.03 (-0.09, 0.15)	-0.08 (-0.12, -0.05)
Turkey	-0.11	-0.08	-0.03	0.01 (-0.16, 0.16)	0.00 (-0.11, 0.11)	-0.03 (-0.05, -0.02)
United Kingdom	-0.28	-0.21	-0.05	-0.15 (-0.31, -0.03)	-0.12 (-0.24, -0.02)	-0.05 (-0.08, -0.03)
United States	7.32	4.33	0.38	6.49 (4.63, 9.96)	3.84 (2.74, 5.89)	0.37 (0.27, 0.55)
Other Non-TTIP	-0.23	-0.29	-0.02	-0.23 (-0.34, -0.16)	-0.29 (-0.43, -0.20)	-0.02 (-0.03, -0.01)
EU	1.48	1.51	0.22	1.55 (1.09, 2.36)	1.58 (1.11, 2.40)	0.20 (0.13, 0.31)
TTIP	2.60	2.32	0.28	2.50 (1.77, 3.81)	2.23 (1.58, 3.41)	0.26 (0.19, 0.41)
Non-TTIP	-0.30	-0.35	-0.03	-0.31 (-0.48, -0.21)	-0.36 (-0.56, -0.24)	-0.03 (-0.05, -0.02)
World	1.28	1.28	0.14	1.22 (0.75, 1.98)	1.22 (0.75, 1.98)	0.13 (0.08, 0.21)

This table compares the results from a general equilibrium simulation of the effects of a hypothetical U.S./E.U. FTA ("TTIP") under two scenarios: (i) an "average" scenario, where all GE effects are predicted based off of a common average partial effect being applied equally to all TTIP pairs, and (ii) a "heterogeneous" scenario where we use heterogeneous partial effects predicted by our out-of-sample prediction model. For the latter set of results, we include bootstrapped 95% confidence intervals in parentheses. See text for further details.

Data Appendix

Table A.1: Included Countries

Main sample *(52 countries/regions)*: Argentina, Australia, Austria, Bulgaria, Belgium-Luxembourg, Bolivia, Brazil, Canada, Switzerland, Chile, China, Colombia, Costa Rica, Cyprus, Germany, Denmark, Ecuador, Egypt, Spain, Finland, France, United Kingdom, Greece, Hungary, Indonesia, Ireland, Iceland, Israel, Italy, Jordan, Japan, South Korea, Kuwait, Morocco, Mexico, Malta, Myanmar, Malaysia, Netherlands, Norway, Philippines, Poland, Portugal, Qatar, Romania, Singapore, Sweden, Thailand, Tunisia, Turkey, Uruguay, United States

"Rest of World" *(17 countries/regions)*: Cameroon, Hong Kong, India, Iran, Kenya, Sri Lanka, Macau, Mauritius, Malawi, Niger, Nigeria, Nepal, Panama, Senegal, Trinidad & Tobago, Tanzania, South Africa

Table A.2: Included Agreements

Multilateral Trade Blocs

Agreement	Year	Member Countries
ASEAN*	2000	Indonesia, Malaysia, Myanmar, Philippines, Singapore, Thailand
Agadir	2006	Egypt, Jordan, Morocco, Tunisia
Andean Community†	1993	Bolivia, Colombia, Ecuador
CEFTA	1993	Poland (1993-2004), Hungary (1993-2004), Romania (1997-2004), Bulgaria (1998-2004)
EFTA	1960	Norway, Switzerland, Iceland (1970), Portugal (1960-1986), Austria (1960-1995), Sweden (1960-1995) Finland (1986-1995).
EU†	1958	Belgium-Luxembourg, France, Italy, Germany, Netherlands, Denmark (1973), Ireland (1973), United Kingdom (1973), Greece (1981), Portugal (1986), Spain (1986), Austria (1995), Finland (1995), Sweden (1995), Cyprus (2004), Malta (2004), Hungary (2004), Poland (2004)
Mercosur*†	1995	Argentina, Brazil, Uruguay
NAFTA	1994	Canada, Mexico, U.S.
Pan Arab Free Trade Area	1998	Egypt, Kuwait, Jordan, Morocco, Qatar, Tunisia

EFTA's outside agreements: Turkey (1992), Bulgaria (1993), Hungary (1993), Israel (1993), Poland (1993), Romania (1993), Mexico (2000), Morocco (2000), Singapore (2003)

EU's outside agreements: EFTA (1973), Cyprus (1988), Hungary (1994), Poland (1994), Bulgaria (1995), Romania (1995), Turkey (1996)†, Tunisia (1998), Israel (2000), Mexico (2000), Morocco (2000),Chile (2003), Egypt (2004)

Other agreements: Australia-Singapore (2003), Australia-Thailand (2005), Australia-U.S. (2005), Bulgaria-Israel (2002), Bulgaria-Turkey (1998), Canada-Chile (1997), , Canada-Costa Rica (2003), Canada-Israel (1997), Canada-U.S. (1989), Chile-China (2006), Chile-Costa Rica (2002), Chile-Mexico (1999), Chile-Singapore (2006), Chile-South Korea (2004), Chile-U.S. (2004), Colombia-Mexico (1995), Costa Rica-Mexico (1995), Egypt-Turkey (2006), Hungary-Israel (1998), Hungary-Turkey (1998), Israel-Mexico (2000), Israel-Poland (1998), Israel-Romania (2001), Israel-Turkey (2001), Japan-Mexico (2005), Jordan-U.S. (2002), Mercosur-Andean (2005), Mercosur-Bolivia (1996), Mercosur-Chile (1996), Mexico-Uruguay (2005), Morocco-U.S. (2006), Poland-Turkey (2000), Romania-Turkey (1998), Singapore-U.S. (2004), Tunisia-Turkey (2006)

*For these two blocs, we follow the NSF-Kellogg Database in using, respectively, the date at which ASEAN "moved toward" becoming a free trade area and the date at which Mercosur became a customs union.
†Denotes a deeper level of agreement (e.g., a customs union).

Supplementary Appendix. (*Not for publication.*)

The purpose of this Appendix is to supplement the analysis of Baier, Yotov, and Zylkin (2018) with some additional documentation supporting our preferred methods as well as to provide an extensive series of sensitivity checks. We begin by adding more details on the use of our "terms of trade sensitivity" (TOT) indices as proxies for market power.

1. More details on TOT indices

In connection with the use of our "$\ln \widehat{ToT}_{A:j}$" indices as a key source of variation in the analysis, we provide more details on how these indices vary across countries as well as how they correlate with alternative external measures of "market power". Table B.1 shows the average $\ln \widehat{ToT}_{A:j}$ values for each country, sorted from least to greatest. Noting that, by construction, $\ln \widehat{ToT}_{A:j}$ is an inverse measure of market power, the countries listed in the top-left would be expected to grant the largest concessions when signing agreements, whereas countries at the bottom-right would be likely to be already close to their optimal policies.

The estimates from Table B.1 reveal that while larger countries are generally ranked higher up in the table than smaller countries, our TOT indices are clearly not just a function of market size. The degree to which a country is "networked" with its FTA partners-to-be also appears to matters here. The countries we infer as having the most market power are large Asian economies such as China, Japan, and South Korea, each of which has pursued agreements with countries outside of their local geographic areas (Chile in the case of China and South Korea, Mexico in the case of Japan). We also see that EU members, even relatively small ones such as Belgium-Luxembourg, are well-represented in the left and middle columns of Table B.1. As discussed in Section 6, this is because tight trade networks within the EU make outside agreements relatively less welfare-enhancing for individual EU members versus if they had each pursued bilateral agreements with these outside countries. By the same token, relatively small European countries such as Hungary or Malta, who had each depended much more heavily on the EU for trade than vice versa before signing their respective EU agreements, are associated with some of the largest TOT values that we obtain.

To further examine the external validity of our TOT indices, Table B.2 shows raw correlations between the $\ln \widehat{ToT}_{A:j}$'s and external estimates of export supply elasticities facing each importer taken from Nicita *et al.* (2018) and Broda *et al.* (2008). More precisely, for the Nicita *et al.* (2018) measure, we use the the simple average export supply elasticity reported in their Table 1. For the Broda *et al.* (2008) measure, we create our own weighted average elasticities by combining their HS4-level estimates with data on HS4 imports for the years 1994-2002. We also show correlations between these measures and applied MFN tariffs, which are weighted-average MFN tariffs taken from TRAINS. As the export supply elasticities and our $\ln \widehat{ToT}_{A:j}$ indices are both inverse measures of market power, the positive correlation between these three variables is as expected. For comparison's sake, it is also interesting to note that our $\ln \widehat{ToT}_{A:j}$ indices are more closely aligned with these external estimates of market power than the importer's MFN tariff, which actually enters with the "wrong" sign in each case. The signs of these latter correlations likely reflect the WTO's successful efforts at getting high-market power countries to reduce their MFN tariffs.[49]

2. Out-of-sample validation procedure

As noted in Section 6, our preferred model for the out-of-sample prediction analysis involves country GDPs, geographical distance, the "pair fixed effect" term from the first stage, and the importer's $\ln \widehat{ToT}_{A:j}$ index. Here, we describe the procedure that we used to select this model:

1. As noted in Section 6.1, we start with a selection of variables that were shown to be significant throughout our second stage analysis, especially when we zeroed in on heterogeneity within agreements and (if applicable) within pairs. The variables meeting this criteria are: Exporter (log) GDP, Importer (log) GDP, ln DIST, LEGAL, First stage pair FE, and

[49]Note that the number of observations for most of the correlations shown in this table is only 655. The reason for this is that the export supply elasticity estimates from Nicita *et al.* (2018) are not available for all countries in our sample. For the estimates from Broda *et al.* (2008), only 9 countries in our data set overlap with theirs. The correlations involving these latter estimates are based on only 17 observations.

$\ln \widehat{ToT}_{A:j}$.

2. Following the procedure described in the text, we repeatedly estimate our second stage model using this initial set of candidate variables after dropping one agreement at a time and then record the out-of-sample predictions for the different pairs of country within the excluded agreement. Noting that we first drop outliers, dropping and predicting each agreement in this way then gives us a set of 817 out-of-sample predicted values we can compare with our original first stage estimates. The fit between these predicted values and the original first stage estimates is summarized as "Baseline" in the accompanying Table B.3. Specifically, we compute a simple linear fit from the following regression:

$$\beta_{A:d} = \rho_0 + \rho_1 \cdot \widetilde{\beta}_{A:d} + e, \tag{23}$$

where $\beta_{A:d}$ is the first stage estimate for a directional pair d within agreement A and $\widetilde{\beta}_{A:d}$ is the corresponding out-of-sample prediction. We use the R^2 from this regression to judge predictive fit.

3. Next, we experiment with dropping each of these regressors one at a time to examine how the goodness of fit changes. As we can see in the case of sharing a common legal system ("LEGAL"), the R^2 of the predictive fit increases slightly when we do not include LEGAL in the model. Thus, we conclude LEGAL does not have sufficient predictive power out-of-sample to be included in our prediction model. The R^2 does suffer, however, when any of the other key variables are dropped, as can be seen from the other results in the Table B.3. Based on this analysis, we concluded that the model described as "drop LEGAL" should be our preferred model for predicting the effects of a hypothetical future agreement. We have also verified that dropping variables in any other order using this procedure leads to the same conclusion.

3. Supplementary aggregate results

To supplement the "aggregate" results reported in the paper for an overall average treatment effect of FTAs, in this section we document how these estimates vary under different specifications of the timing of the FTA effects. The motivation for this analysis was twofold. First, we wanted to confirm the representativeness of our sample and analysis in relation to existing studies that have used lead and lag FTA dummies to assess the timing of FTA effects (c.f., Baier & Bergstrand, 2007; Anderson & Yotov, 2016.) Second, we used the analysis in this section to guide the main estimations with heterogeneous effects. Table B.4 presents the results from a series of specifications with common lead and lag coefficients. Column 1 reproduces the average FTA estimate we would have estimated without any lags or leads. Column 2 introduces 2 year lead and 2 and 4 year lags. Then, column 3 uses 3 year lead and 3 and 6 year lags, and so forth, until we reach to column 6, which has 6 year lead and 6 and 12 year lags.

Several findings stand out from these supplementary estimates: (i) None of the leads are statistically or economically significant. (ii) We see that when we move further and further away from the year in which the agreement entered into force, the estimates of the leads decrease monotonically and move from positive to negative. (iii) As expected, the estimates of the lags reveal phasing-in effects. (iv) Also as expected, we do see that the phasing-in is non-monotonic. After entering into force, the agreements have stronger effects, then they fade away; (v) The majority of the trade impact of FTAs seems to occur within the first 5-6 years. (vi) Finally, consistent with our expectations, we found that the longer the second lags, the weaker were their estimates both in terms of magnitude and significance. We conclude based on this analysis that using a single 5 year lag in the first stage provides a reasonable approximation of the overall timing of effects. Later in this Appendix, we also document that our main results in the second stage are robust to the choice of lag intervals in the first stage.

4. Additional sensitivity experiments

To supplement the results reported in the paper, in this section we explore some additional exercises demonstrating the robustness of our findings to alternative choices we could have made at various points in the analysis. Our three main focuses here are different ways to incorporate lagged effects of our first stage estimates, different ways of computing our $\ln \widehat{ToT}_{A:j}$ indices, and different ways of weighting our first stage estimates to account for the fact they are measured with error. We also show some additional results for when we include estimated demand elasticities from Broda *et al.* (2006) as an additional explanatory variable in the second stage.

Varying the first stage treatment of lags. Tables B.5 and B.6 show second stage results based on various alternative methods for how to incorporate lagged effects of FTAs into the analysis (including one alternative where we ignore lagged effects altogether):

- *Drop later agreements*. First, we examine what happens when, instead of imputing lagged effects for later agreements based on their estimated initial effects, we simply drop these agreements from the second stage. As we can see in columns 1-4 of Table B.5 - which correspond to the specifications used columns 1-4 of Table 3 in the main text - the main effect of dropping later agreements in this way is to significantly reduce the number of estimates we can work with in the second stage, from 908 to 560. Nonetheless, despite losing more than a third of our sample in the second stage, the results are broadly robust. The only (minor) differences relative to the corresponding results in Table 3 are: (i) ln DIST becomes only marginally significant when we examine variation within agreements (column 3); (ii) the importer's GDP term and our extensive margin measure change significance slightly when we examine variation within pairs (column 4).

- *Ignore lags*. Next, another way of dealing with the issue of not being able to estimate lags for later agreements might to be to ignore lags altogether in the first stage. Accordingly, columns 5-8 of Table B.5 show what second stage results look like using an alternative first stage where, in place of (8), we instead estimate

$$X_{ij,t} = \exp\left(\eta_{i,t} + \psi_{j,t} + \gamma_{\overrightarrow{ij}} + \sum_{A}\sum_{d\in A}\beta_{A:d}FTA_{ij,t} + \sum_{t} b_t\right) + \varepsilon_{ij,t},$$

where we are now implicitly assuming that the effects of an agreement on trade should occur all at once and that, for each directional pair, this effect is given by a single parameter "$\beta_{A:d}$". We then re-produce our second stage analysis using these alternative $\beta_{A:d}$'s, for which we do not need to impute any lags. As can be seen from columns 5-8 of Table B.5, our second stage results still remain similar even if we ignore lags in the first stage in this way.

- *Use different lags*. Finally, Table B.6 displays three further sets of alternative results based on if we had used 3, 4, or 6 year lags in estimating our first stage instead of 5 year lags. Each set of three results corresponds to columns 2-4 of Table 3 in the main text.[50] The results are highly stable across these alternatives and consistent with those shown in the paper for 5 year lags.

Along with our analysis of how the aggregate results vary with different lags (in Table B.4), the additional experiments with alternative lags that we presented here give us confidence that the approach we have taken in the paper gives us a reasonable picture of the timing of FTA effects. In further experiments, omitted for brevity, we have also found all other second stage results reported in the paper are robust to these alternative choices over how to allow for lagged effects.

Different ways of computing TOT indices and/or assessing market power. In this set of experiments, we demonstrate the robustness of our results to several reasonable alternatives for computing our "terms of trade sensitivity" indices and/or assessing market power by other means:

- *Restrict* $\ln \widehat{ToT}_{A:j}$ *to always be the same within customs unions*. As can be seen from Table A.2 of our Data Appendix, a significant number of the agreements in our sample involve the members of a customs union collectively signing an FTA with an outside country (or the

[50]That is, we skip results from column 1 of of Table 3 without added gravity controls (though these can be shown to be similar as well).

addition of a new member to a customs union, as is the case with the EU accessions in our sample). As such, a reasonable alternative to how we construct our TOT indices would be to restrict them to be the same for existing members of a customs union. We perform this experiment the first four columns of Table B.7, where we use a new version of our TOT index - "$\ln \widehat{ToT}^{CU}_{A:j}$ - which always takes a common value for the existing members of a customs union. Specifically, whenever j belongs to a customs union, we use a weighted average of the underlying $\ln \widehat{ToT}_{A:j}$'s for each member of the customs union, using each country's share of world manufacturing expenditure as weights. As columns 1-4 of Table B.7 show, the results from this experiment are exactly in line with those reported in the paper.

- *Use the bilateral difference in terms of trade sensitivity*. Our decision to use only the importer's terms of trade index follows from our interpretation of the theory, which suggests that importers with more market power all else equal should have higher trade barriers *ex ante*. However, it is fair to also point out that our preferred second stage specification also includes an "inclusive" measure of the average level of trade barriers within any given pair (i.e., our "First stage pair FE" variable), which could plausibly capture the average level of policy barriers. Therefore, another reasonable alternative for implementing our $\ln \widehat{ToT}_{A:j}$ indices would be to instead take the bilateral difference between the $\ln \widehat{ToT}_{A:j}$'s of the two countries. That is, we define

$$\Delta \ln \widehat{ToT}_{A:ij} = \ln \widehat{ToT}_{A:i} - \ln \widehat{ToT}_{A:j}$$

as the bilateral difference in terms of trade sensitivity and examine whether this term specifically predicts asymmetries in trade creation within pairs across each of our prior second stage specifications. As can be seen from columns 5-8 of Table B.7, taking this alternative perspective again leads to our results being as expected: the sign on $\Delta \ln \widehat{ToT}_{A:ij}$ is positive and significant across each of our prior second stage specifications, confirming again that asymmetries in FTA effects consistently reflect relatively larger trade creation flowing from countries with high terms of trade sensitivity to countries with low terms of trade sensitivity.[51]

- *Use the ratio of export prices to import prices.* Another reasonable criticism of how we construct "terms of trade" in the paper is that it involves the ratio of producer prices to consumer prices, following the formulation used in Anderson & Yotov (2016). However, readers more familiar with traditional presentations of the theory may be more used to thinking of "terms of trade" as a ratio of export prices to import prices. Following the argumentation of Anderson & Yotov (2016), we maintain this ratio represents the relevant concept to focus on in a multi-country gravity context (especially in a world with unbalanced trade).[52] Nonetheless, drawing on the derivations Caliendo & Parro (2015) include in their Appendix, it is possible to derive the following, alternative index for the change in the terms of trade:

$$\ln \widehat{ToT}^{alt}_{A:j} = \frac{EX_j}{E_j} \ln \widehat{w}_j - \sum_{i \neq j} \frac{X_{ij}}{E_j} \ln \widehat{w}_i,$$

  which can be calculated using the exact same simulation procedure as our original index, only now multiplying the change in j's wage by its total exports EX_j and also approximating a CES index for the change in import prices (as opposed to using the change in buyer prices more generally, which would also include domestic producer prices and changes in trade costs). However, an interesting issue that arises with this index is that it is sensitive to the choice of numeraire when trade is not balanced.[53] Therefore, before computing $\ln \widehat{ToT}^{alt}_{A:j}$, we first replicate Dekle *et al.* (2007)'s famous exercise of solving for a counterfactual world

[51] Indeed, if one carefully examines column 8 of Table B.7, it is apparent that the coefficient on $\Delta \ln \widehat{ToT}_{A:ij}$ is exactly -2 times the corresponding estimate shown in column 4 of our original Table 3. The reason for this is because both specifications include agreement-by-pair fixed effects, such that identification can only be obtained off of asymmetries in $\ln \widehat{ToT}_{A:j}$ in either case.

[52] More to the point, the main consideration here is that the notion of "terms of trade" we rely on should satisfy the condition stated in equation (3) of Bagwell & Staiger (1999) (p. 220). Specifically, a country's terms of trade and welfare should be positively related (after all, it is the welfare inefficiency generated by the terms of trade externality that ultimately matters in the theory.) Using the ratio of producer prices to buyer prices is a straightforward way of satisfying this consideration, is easy to compute in a multi-country setting, and does not require trade to be balanced.

[53] To see this, note that in a world with unbalanced trade, j's total imports $\sum_{i \neq j} X_{ij}$ will not be equal to its total exports EX_j. Any uniform increase in wages across all countries would therefore raise the terms of trade of any net exporter country and lower the terms of trade of any net importer country. Our preferred index does not depend on choice of numeraire when solving the model.

with balanced trade.[54] Using this "balanced trade" world as the new baseline, we then proceed to simulate a common partial effect for each agreement (as in the text), only now taking $\ln \widehat{ToT}^{alt}_{A:j}$ as our notion of "terms of trade sensitivity".

The results from this alternative formulation of our TOT index are shown in the first 4 columns of B.8, which again correspond to our baseline results from Table 3. The most obvious difference of note from our prior results is that the magnitudes of the coefficients computed for $\ln \widehat{ToT}^{alt}_{A:j}$ are much larger than those found using our preferred approach. This is mainly a matter of scaling differences across the two indices; our original index $\ln \widehat{ToT}_{A:j}$ is more sensitive to changes in trade costs than $\ln \widehat{ToT}^{alt}_{A:j}$. We do note $\ln \widehat{ToT}^{alt}_{A:j}$ becomes statistically insignificant when agreement fixed effects are used (in column 3). However, our results for other specifications are qualitatively no different from before, including in column 4 where we narrow the focus to asymmetries within pairs by using agreement-pair fixed effects.

- *Use import shares*. Finally, one other alternative way of assessing "market power" in this context would be to simply use each country's share of world import demand. As columns 5-8 of B.8 show, our results for import shares enter positively and significantly as expected across each of our main specifications.

As with our experiments involving lags, we have verified that all second stage results reported in the paper are robust to these alternative choices.

Heterogeneous trade elasticities. Since the general framework we used to motivate some of our key variables relied on the typical assumption of a constant trade elasticity, we wanted to examine the robustness our second stage results to introducing external estimates of country-specific trade elasticities. Specifically, we obtained country-specific import demand elasticities at the HS3 level from Broda *et al.* (2006). We then constructed country-specific aggregate import demand

[54] That is to say, we start from a world where a set of exogenous transfers $\{D_j\}$ satisfying $\sum_j D_j = 0$ allows us to rationalize observed imbalances between output Y_j and expenditure E_j as $E_j = Y_j + D_j$. We then solve for a counterfactual world with all D_j's set equal to zero, using the same equilibrium conditions described in Section 6.2.

elasticities by weighting these HS3-level elasticities by the weight of each HS3 industry in a country's overall manufacturing imports over the years 1994-2002. For countries in our main sample for which Broda *et al.* (2006) do not explicitly provide elasticity estimates, we first constructed "world-level" elasticities at the HS3 level using the share of each HS3 industry in world trade. We then combined these world-level elasticities with our country-specific import shares to construct aggregate elasticities for the remaining countries. Specifically, Broda *et al.* (2006) provide estimates of the elasticity of substitution between foreign varieties within the same industry. Taking the elasticity parameters estimated by Broda *et al.* (2006) to be "σ", the trade elasticity we use in our analysis can be expressed as $\theta = \sigma - 1$. Because Broda *et al.* (2006)'s estimated elasticities feature some extreme values (ranging from a minimum of 1.06 to values well over 1,000), we first take the precaution of bounding these estimates to be between 2 and 15 before constructing aggregate elasticities. The end result is very similar even if we do not take this precaution, however.

Table B.9 shows the effects of adding country-specific import demand elasticities to our second stage. To set the stage, what we might expect to observe here is that importers with higher import demand elasticities should be associated with a larger increase in trade following the signing of an FTA; thus, we would expect these country-specific elasticities should be positively related to our first stage FTA estimates. This is indeed what we observe in column 1 of Table B.9, where we include the elasticity estimates alongside our initial specification with only key covariates. However, this result is not robust to the inclusion of standard gravity controls in column 2, where the estimate of the coefficient on the import demand elasticity is not longer statistically significant. We also observe that the new elasticity variable actually becomes negative and significant when we include agreement fixed effects (in column 3) as well as when we include agreement-pair fixed effects (in column 4). Importantly, we note that our prior results from Table 3 are robust to the introduction of the elasticity estimates.

Different weighting methods. We finish our robustness analysis by documenting what happens when we vary the weighting method used to compute our second stage estimates. Following the recommendations of Lewis & Linzer (2005), in the main text, we considered only (unweighted) estimates using OLS and heteroskedasticity-robust standard errors (using the standard error correction of White, 1980). Here, for completeness, we consider the two other alternatives for weighting the first stage, which also were examined in Lewis & Linzer (2005): (i) a standard WLS ("Weighted Least Squares") estimator - which weights each observation by the inverse of the first stage standard error - and (ii) a special FGLS ("Feasible Generalized Least Squares") estimator proposed by Hanushek (1974) for problems of this type.

For concreteness, we will refer to the (lack of) weighting method associated with OLS as "**W1**". A second, widely used weighting method for two stage estimation is the standard WLS estimator, or "**W2**":

$$\mathbf{W2}: \; weight_{A:d} = \frac{1}{\sqrt{\sigma^2_{I,A:d}}},$$

where $\sigma_{I,A:d}$ is the standard error associated with each $\beta_{A:d}$ estimated in the first stage. **W2** has the desirable property that more precisely estimated $\beta_{A:d}$'s from the first stage are given more weight in determining second stage estimates. Unfortunately, this weighting method has the drawback of assuming *all* model uncertainty in the second stage is due to the error associated with $\beta_{A:d}$. Accordingly, a third alternative, first suggested by Hanushek (1974), is

$$\mathbf{W3}: \; weight_{A:d} = \frac{1}{\sqrt{\sigma^2_{I,A:d} + \hat{\sigma}^2_{II}}},$$

where $\hat{\sigma}^2_{II}$ is an unbiased estimate of the the second stage error variance, assuming homoskedastic errors. Weighting using **W3** is an example of a "Feasible Generalized Least Squares" estimator, which we will abbreviate as "FGLS". Intuitively, FGLS varies the degree of weighting depending on the relative magnitudes of the (individual) first stage error variances vs. the (total) second stage error variance; it therefore nests both WLS and OLS as extreme cases.

As we see from columns 1-4 of Tables B.10, our FGLS estimates are a close match for our original OLS estimates, implying that - exactly as Lewis & Linzer (2005) would caution with regards to WLS - the degree of second stage error is large relative to the standard errors we estimate in the first stage. Indeed, the WLS results we compute using **W2** (which we include for instructive reasons only) are significantly less similar to our original results using OLS. For completeness, we also show "analysis of variance" results for each of these three weighting methods, focusing on how much variation in our first stage directional estimates represents heterogeneity that we observe across agreements versus within agreements (and within pairs). As Table B.11 shows, all three weighting methods suggest very similar decompositions in this regard.

Table B.1: Comparison of TOT indices across countries

Country	$\ln \widehat{ToT}_{A:j}$	Country	$\ln \widehat{ToT}_{A:j}$	Country	$\ln \widehat{ToT}_{A:j}$
China	0.0001	Germany	0.0007	Egypt	0.0072
Japan	0.0001	Colombia	0.0009	Singapore	0.0080
South Korea	0.0001	Jordan	0.0009	Finland	0.0088
Iceland	0.0002	United States	0.0010	Cyprus	0.0091
Spain	0.0002	Argentina	0.0012	Morocco	0.0098
Switzerland	0.0002	Denmark	0.0012	Bulgaria	0.0104
France	0.0002	Ecuador	0.0014	Poland	0.0113
Portugal	0.0002	Australia	0.0020	Tunisia	0.0122
Norway	0.0003	Myanmar	0.0027	Austria	0.0122
Belgium-Luxembourg	0.0003	Uruguay	0.0037	Sweden	0.0126
United Kingdom	0.0003	Mexico	0.0040	Canada	0.0137
Qatar	0.0003	Indonesia	0.0041	Hungary	0.0155
Ireland	0.0003	Chile	0.0042	Malaysia	0.0184
Italy	0.0003	Turkey	0.0049	Malta	0.0242
Greece	0.0003	Bolivia	0.0053		
Brazil	0.0003	Romania	0.0057		
Kuwait	0.0004	Philippines	0.0063		
Costa Rica	0.0005	Israel	0.0065		
Netherlands	0.0005	Thailand	0.0068		

This table shows a cross-country comparison of the values we compute for $\ln \widehat{ToT}_{A:j}$, our "terms of trade sensitivity" index. Note that this variable is agreement-specific. For countries that join multiple FTAs during the sample, we show the mean value across the different agreements.

Table B.2: Correlations between TOT indices, export supply elasticities, and MFN tariffs

	Raw correlations (# obs. = 655)			
	$\ln \widehat{ToT}_{A:j}$	Exp. supply elasticity (Nicita et al)	Exp. supply elasticity (Broda et al)	MFN tariff
$\ln \widehat{ToT}_{A:j}$	1			
Exp. supply elasticity (Nicita et al)	0.460	1		
Exp. supply elasticity (Broda et al)	0.300	0.732	1	
MFN tariff	0.396	0.247	0.572	1

$\ln \widehat{ToT}_{A:j}$ is our "terms of trade sensitivity" index, described in Section 3.2. "Export supply elasticity (Nicita et al)" and "Export supply elasticity (Broda et al)" respectively refer to external estimates of export supply elasticities facing each importer taken from Nicita *et al.* (2018) and Broda *et al.* (2008). As these first three variables are intended as inverse measures of market power, we would expect them to be positively correlated with one another and with $\ln \widehat{ToT}_{A:j}$. "MFN tariff" is the log of "(1 + MFN tariff)". We would expect it to be negatively correlated with these other variables. There are 655 observations for which $\ln \widehat{ToT}_{A:j}$, MFN tariff, and the Nicita *et al.* (2018) export supply elasticity measure are jointly available. Correlations between the Broda *et al.* (2008) export supply elasticity measure and the other variables are computed using only 17 observations.

Table B.3: Out-of-sample Validation Results

Selected Prediction Models and Model Fit Results				
Model	*Included variables*	ρ_0	ρ_1	R^2
Baseline	First Stage Pair FEs, ln $\widehat{ToT}_{A:j}$, lnDIST, log GDP$_i$, log GDP$_j$, LEGAL	0.092***	0.764***	0.174
Drop LEGAL†	First Stage Pair FEs, ln $\widehat{ToT}_{A:j}$, lnDIST, log GDP$_i$, log GDP$_j$	0.091***	0.771***	0.178
Drop LEGAL & First Stage Pair FEs	ln $\widehat{ToT}_{A:j}$, lnDIST, log GDP$_i$, log GDP$_j$	0.150***	0.577***	0.051
Drop LEGAL & ln $\widehat{ToT}_{A:j}$	First Stage Pair FEs, lnDIST, log GDP$_i$, log GDP$_j$	0.040	0.855***	0.158
Drop LEGAL & lnDIST	First Stage Pair FEs, ln $\widehat{ToT}_{A:j}$ log GDP$_i$, log GDP$_j$	0.105**	0.712***	0.071
Drop LEGAL & GDPs	First Stage Pair FEs, ln $\widehat{ToT}_{A:j}$, lnDIST	0.065**	0.799***	0.085

This table compares the predictive power of alternative models that can be used to make out-of-sample FTA predictions. To predict FTA effects out-of-sample, we drop one agreement at a time, then fit the indicated model based on the remaining agreements. The reported coefficients are from a regression of our first stage estimates on the predicted FTA effects obtained from out-of-sample prediction. ρ_1 is the correlation from this regression and ρ_0 is the constant. The R^2 is used to judge goodness-of-fit. Results across the last 4 models are similar if we continue to include LEGAL. * $p < 0.10$, ** $p < .05$, *** $p < .01$. † denotes our preferred model. See the Supplementary Appendix for further details.

Table B.4: Supplementary Aggregate FTA Estimates with Varying Leads and Lags

	No lags or leads	*2 & 4 year lags; 2 year leads*	*3 & 6 year lags; 3 year leads*	*4 & 8 year lags; 4 year leads*	*5 & 10 year lags; 5 year leads*	*6 & 12 year lags; 6 year leads*
	(1)	(2)	(3)	(4)	(5)	(6)
(*lead effects*)						
$FTA_{ij,t+6}$						-0.064
						(0.096)
$FTA_{ij,t+5}$					-0.036	
					(0.093)	
$FTA_{ij,t+4}$				-0.006		
				(0.086)		
$FTA_{ij,t+3}$			0.034			
			(0.078)			
$FTA_{ij,t+2}$		0.052				
		(0.074)				
(*main effect*)						
$FTA_{ij,t}$	0.188**	-0.001	0.031	0.074	0.107*	0.134**
	(0.091)	(0.036)	(0.044)	(0.051)	(0.060)	(0.068)
(*lagged effects*)						
$FTA_{ij,t-2}$		0.069***				
		(0.027)				
$FTA_{ij,t-4}$		0.183***		0.179***		
		(0.048)		(0.041)		
$FTA_{ij,t-3}$			0.146***			
			(0.041)			
$FTA_{ij,t-6}$			0.113**			0.174***
			(0.044)			(0.051)
$FTA_{ij,t-8}$				0.066		
				(0.046)		
$FTA_{ij,t-5}$					0.182***	
					(0.045)	
$FTA_{ij,t-10}$					0.042	
					(0.035)	
$FTA_{ij,t-12}$						0.025
						(0.043)
N	58,989	58,989	58,989	58,989	58,989	58,989

This table shows supplementary aggregate first stage FTA estimates (based on eq. (5)) with varying lag and lead terms. Standard errors (clustered by exporter, importer, and year) are shown in parentheses. * $p < 0.10$, ** $p < .05$, *** $p < .01$.

Table B.5: Robustness: Varying the first stage

	Dependent variable: First stage directional FTA estimates							
	Drop later agreements				*No lags in the first stage*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First stage pair FE†	-0.234***	-0.478***	-0.300***		-0.174***	-0.348***	-0.262***	
	(0.040)	(0.057)	(0.094)		(0.031)	(0.038)	(0.051)	
ln $\widehat{ToT}_{A:j}$	-21.490***	-11.421***	-9.937**	-10.587**	-22.411***	-11.599***	-8.420*	-13.723***
	(3.543)	(4.020)	(3.932)	(4.356)	(3.551)	(4.126)	(4.412)	(4.497)
Extensive margin of trade	1.052***	0.580	0.388	-0.830**	1.094***	0.578**	0.404*	-0.396
	(0.224)	(0.357)	(0.327)	(0.399)	(0.149)	(0.227)	(0.213)	(0.267)
ln DIST		-0.628***	-0.213*			-0.422***	-0.228***	
		(0.052)	(0.124)			(0.038)	(0.073)	
CONTIG		0.199*	0.077			0.161*	-0.025	
		(0.105)	(0.128)			(0.086)	(0.093)	
COLONY		-0.098	-0.070			-0.028	0.013	
		(0.119)	(0.123)			(0.086)	(0.093)	
LANG		0.293***	0.174			0.119	0.103	
		(0.099)	(0.109)			(0.074)	(0.081)	
LEGAL		0.037	0.107			0.124*	0.230***	
		(0.086)	(0.092)			(0.065)	(0.070)	
Prior Agreement		-0.225**	-0.092			-0.233***	-0.022	
		(0.105)	(0.081)			(0.050)	(0.081)	
Exporter (log) Real GDP		0.259***	0.169***			0.178***	0.148***	
		(0.049)	(0.051)			(0.030)	(0.033)	
Importer (log) Real GDP		0.257***	0.155***	-0.127*		0.157***	0.120***	-0.102***
		(0.032)	(0.047)	(0.065)		(0.022)	(0.026)	(0.038)
Exporter (log) GDP per capita		-0.087	0.090			-0.053	0.004	
		(0.121)	(0.145)			(0.066)	(0.091)	
Importer (log) GDP per capita		0.164**	0.324***	0.084		0.138***	0.186**	0.072
		(0.064)	(0.105)	(0.096)		(0.041)	(0.072)	(0.067)
Constant	0.469***	0.469***			0.271***	0.276***		
	(0.036)	(0.031)			(0.026)	(0.023)		
Agreement FEs			x				x	
Agr.×pair FEs				x				x
Observations	560	560	560	560	908	908	908	908
R^2	0.132	0.373	0.524	0.752	0.076	0.274	0.435	0.735
Within R^2			0.129	0.105			0.124	0.063

Second stage estimates are obtained using OLS with robust standard errors (reported in parentheses). The dependent variable is $\beta_{A:d}$, an estimated direction-specific FTA partial effect which we have estimated in a prior first stage. This table experiments with different ways of addressing the issue that we cannot explicitly estimate lagged effects for the later (post-2001) agreements in our sample. Columns 1-4 demonstrates what our main second stage results look like if we simply drop these later agreements (thereby reducing our second stage sample to 560 observations). Columns 5-8 are based on an alternative first stage where we ignore lags altogether. In columns 1, 2, 5, and 6, all variables are de-meaned with respect to their within-sample mean. This allows us to interpret the regression constant as reflecting the overall average FTA estimate after netting out the average effects of each of the included covariates. * $p < 0.10$, ** $p < .05$, *** $p < .01$. † Also accounts for "globalization" effects. See text for further details.

Table B.6: Robustness: Varying the lag interval used in the first stage

	Dependent variable: First stage directional FTA estimates								
	Use 3 year lags			*Use 4 year lags*			*Use 6 year lags*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First stage pair FE†	-0.359***	-0.265***		-0.362***	-0.261***		-0.384***	-0.271***	
	(0.040)	(0.054)		(0.041)	(0.056)		(0.042)	(0.057)	
ln $\widehat{ToT}_{A:j}$	-8.927***	-7.535**	-9.182***	-10.208***	-7.626**	-9.315***	-10.931***	-7.287**	-9.207***
	(2.784)	(2.991)	(3.028)	(2.891)	(3.094)	(3.126)	(2.998)	(3.206)	(3.278)
Extensive margin of trade	0.569**	0.411*	-0.512*	0.575**	0.393	-0.538*	0.528**	0.329	-0.632**
	(0.248)	(0.232)	(0.285)	(0.255)	(0.239)	(0.294)	(0.265)	(0.249)	(0.303)
ln DIST	-0.439***	-0.235***		-0.439***	-0.236***		-0.472***	-0.254***	
	(0.041)	(0.077)		(0.042)	(0.079)		(0.045)	(0.081)	
CONTIG	0.143	-0.029		0.171*	-0.011		0.187*	0.000	
	(0.090)	(0.102)		(0.092)	(0.103)		(0.096)	(0.105)	
COLONY	0.033	0.077		0.028	0.080		0.012	0.079	
	(0.094)	(0.104)		(0.097)	(0.108)		(0.103)	(0.113)	
LANG	0.092	0.073		0.098	0.070		0.107	0.063	
	(0.079)	(0.092)		(0.081)	(0.093)		(0.082)	(0.096)	
LEGAL	0.160**	0.247***		0.156**	0.248***		0.181**	0.277***	
	(0.071)	(0.075)		(0.073)	(0.078)		(0.076)	(0.081)	
Prior Agreement	-0.193***	0.001		-0.239***	-0.006		-0.266***	0.008	
	(0.054)	(0.085)		(0.055)	(0.086)		(0.057)	(0.085)	
Exporter (log) Real GDP	0.184***	0.140***		0.190***	0.145***		0.214***	0.160***	
	(0.033)	(0.036)		(0.034)	(0.037)		(0.036)	(0.038)	
Importer (log) Real GDP	0.147***	0.095***	-0.123***	0.141***	0.090***	-0.133***	0.152***	0.094***	-0.148***
	(0.024)	(0.028)	(0.040)	(0.025)	(0.029)	(0.041)	(0.026)	(0.030)	(0.044)
Exporter (log) GDP per capita	-0.073	-0.007		-0.067	-0.002		-0.073	0.007	
	(0.076)	(0.102)		(0.080)	(0.106)		(0.082)	(0.107)	
Importer (log) GDP per capita	0.144***	0.200**	0.091	0.157***	0.216**	0.101	0.185***	0.263***	0.135*
	(0.045)	(0.081)	(0.074)	(0.047)	(0.085)	(0.076)	(0.048)	(0.086)	(0.078)
Constant	0.342***			0.349***			0.343***		
	(0.025)			(0.026)			(0.027)		
Agreement FEs		x			x			x	
Agr.×pair FEs			x			x			x
Observations	908	908	908	908	908	908	908	908	908
R^2	0.260	0.417	0.728	0.255	0.416	0.729	0.271	0.432	0.732
Within R^2		0.111	0.075		0.106	0.080		0.114	0.095

Second stage estimates are obtained using OLS with robust standard errors (reported in parentheses). The dependent variable is $\beta_{A:d}$, an estimated direction-specific FTA partial effect which we have estimated in a prior first stage. This table shows how our main second stage results varying when we use different lag intervals in the first stage. In columns 1, 4, and 7, all variables are de-meaned with respect to their within-sample mean. This allows us to interpret the regression constant as reflecting the overall average FTA estimate after netting out the average effects of each of the included covariates. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table B.7: Robustness: Alternative Terms of Trade Sensitivity Measures I

	Dependent variable: First stage directional FTA estimates							
	Use Terms of Trade Sensitivity for CUs as a whole				*Use Bilateral Differences in ToT sensitivity*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First stage pair FE†	-0.196***	-0.372***	-0.262***		-0.224***	-0.389***	-0.261***	
	(0.034)	(0.041)	(0.057)		(0.036)	(0.041)	(0.057)	
$\ln \widehat{ToT}^{CU}_{A:j}$	-17.259***	-10.903***	-7.639**	-9.337***				
	(2.454)	(2.941)	(3.105)	(3.192)				
$\Delta \ln \widehat{ToT}_{A:ij}$					9.739***	5.751***	5.588***	4.711***
					(1.719)	(1.896)	(1.777)	(1.642)
Extensive margin of trade	1.164***	0.557**	0.350	-0.572*	1.131***	0.498*	0.380	-0.573*
	(0.170)	(0.263)	(0.246)	(0.304)	(0.177)	(0.265)	(0.245)	(0.303)
ln DIST		-0.456***	-0.240***			-0.443***	-0.232***	
		(0.044)	(0.081)			(0.043)	(0.080)	
CONTIG		0.185**	-0.003			0.212**	-0.006	
		(0.093)	(0.103)			(0.096)	(0.102)	
COLONY		0.016	0.076			0.018	0.082	
		(0.100)	(0.111)			(0.102)	(0.113)	
LANG		0.112	0.074			0.140*	0.067	
		(0.081)	(0.093)			(0.082)	(0.095)	
LEGAL		0.160**	0.261***			0.167**	0.257***	
		(0.075)	(0.080)			(0.076)	(0.079)	
Prior Agreement		-0.264***	-0.002			-0.283***	0.031	
		(0.055)	(0.085)			(0.056)	(0.086)	
Exporter (log) Real GDP		0.202***	0.152***			0.213***	0.153***	
		(0.035)	(0.038)			(0.036)	(0.038)	
Importer (log) Real GDP		0.146***	0.091***	-0.139***		0.151***	0.083***	-0.139***
		(0.026)	(0.030)	(0.043)		(0.025)	(0.030)	(0.043)
Exporter (log) GDP per capita		-0.067	0.002			-0.074	0.001	
		(0.083)	(0.108)			(0.083)	(0.107)	
Importer (log) GDP per capita		0.173***	0.235***	0.118		0.155***	0.218**	0.117
		(0.048)	(0.086)	(0.078)		(0.047)	(0.085)	(0.079)
Constant	0.349***	0.349***			0.349***	0.349***		
	(0.029)	(0.026)			(0.029)	(0.026)		
Agreement FEs			x				x	
Agr.×pair FEs				x				x
Observations	908	908	908	908	908	908	908	908
R^2	0.078	0.262	0.424	0.729	0.071	0.258	0.427	0.729
Within R^2			0.108	0.086			0.111	0.086

Robust standard errors reported in parentheses. The dependent variable is $\beta_{A:d}$, an estimated direction-specific FTA partial effect which we have estimated in a prior first stage. In columns 1-4 of this table, whenever the importer belongs to a customs union, $\ln \widehat{ToT}^{CU}_{A:j}$ is computed using the terms of trade sensitivity index for the customs union as a whole, rather than for each individual member of the customs union. In columns 5-8, we use the bilateral difference in computed $\ln \widehat{ToT}_{A:j}$ indices within each pair, rather than using only the importer's $\ln \widehat{ToT}_{A:j}$ term. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table B.8: Robustness: Alternate Terms of Trade Sensitivity Measures II

	Dependent variable: First stage directional FTA estimates							
	Use Alt. Terms of Trade Sensitivity				*Use World Import Demand Shares*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First stage pair FE†	-0.191***	-0.374***	-0.261***		-0.234***	-0.392***	-0.269***	
	(0.034)	(0.041)	(0.058)		(0.038)	(0.042)	(0.058)	
$\ln \widehat{ToT}^{alt}_{A:j}$	-81.795***	-52.243***	-24.763	-47.880***				
	(13.336)	(14.969)	(16.168)	(16.986)				
Importer's share of world imports					5.568***	3.859**	5.028***	6.635***
					(1.335)	(1.734)	(1.469)	(1.971)
Extensive margin of trade	1.037***	0.530**	0.326	-0.626**	0.927***	0.408	0.296	-0.454
	(0.167)	(0.265)	(0.247)	(0.305)	(0.166)	(0.267)	(0.242)	(0.309)
ln DIST		-0.455***	-0.246***			-0.459***	-0.255***	
		(0.044)	(0.081)			(0.044)	(0.081)	
CONTIG		0.179*	-0.008			0.209**	-0.005	
		(0.094)	(0.103)			(0.097)	(0.104)	
COLONY		0.019	0.079			0.019	0.069	
		(0.101)	(0.112)			(0.102)	(0.113)	
LANG		0.113	0.071			0.119	0.052	
		(0.082)	(0.094)			(0.084)	(0.096)	
LEGAL		0.166**	0.261***			0.171**	0.258***	
		(0.076)	(0.080)			(0.077)	(0.079)	
Prior Agreement		-0.263***	0.016			-0.272***	0.045	
		(0.055)	(0.085)			(0.055)	(0.085)	
Exporter (log) Real GDP		0.200***	0.146***			0.204***	0.148***	
		(0.035)	(0.038)			(0.035)	(0.037)	
Importer (log) Real GDP		0.151***	0.096***	-0.137***		0.131***	0.055	-0.170***
		(0.025)	(0.030)	(0.043)		(0.029)	(0.034)	(0.044)
Exporter (log) GDP per capita		-0.080	-0.012			-0.091	-0.001	
		(0.081)	(0.107)			(0.080)	(0.107)	
Importer (log) GDP per capita		0.181***	0.243***	0.131*		0.143***	0.206**	0.101
		(0.049)	(0.087)	(0.076)		(0.052)	(0.086)	(0.075)
Constant	0.349***	0.349***			0.349***	0.349***		
	(0.029)	(0.026)			(0.029)	(0.026)		
Agreement FEs			x				x	
Agr.×pair FEs				x				x
Observations	908	908	908	908	908	908	908	908
R^2	0.071	0.260	0.422	0.729	0.067	0.256	0.427	0.732
Within R^2			0.104	0.085			0.112	0.096

Robust standard errors reported in parentheses. The dependent variable is $\beta_{A:d}$, an estimated direction-specific FTA partial effect which we have estimated in a prior first stage. In columns 1-4, we compute $\ln \widehat{ToT}^{alt}_{A:j}$ using an alternate formulation for the change in terms of trade based on Caliendo & Parro (2015) after first imposing balanced trade. In columns 5-8, we use the importer's share of world imports. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table B.9: Second Stage Estimates: Include Import Demand Elasticity

	Dependent variable: First stage directional FTA estimates			
	(1)	(2)	(3)	(4)
First stage pair FE†	-0.190***	-0.370***	-0.262***	
	(0.034)	(0.042)	(0.057)	
$\ln \widehat{ToT}_{A:j}$	-17.364***	-10.791***	-6.807**	-8.816***
	(2.465)	(2.996)	(3.205)	(3.311)
Extensive margin of trade	1.126***	0.558**	0.382	-0.521*
	(0.168)	(0.261)	(0.244)	(0.305)
Import Demand Elasticity	0.088***	0.009	-0.067**	-0.078**
	(0.029)	(0.031)	(0.032)	(0.036)
ln DIST		-0.452***	-0.251***	
		(0.045)	(0.081)	
CONTIG		0.191**	-0.012	
		(0.094)	(0.102)	
COLONY		0.014	0.034	
		(0.100)	(0.111)	
LANG		0.113	0.097	
		(0.082)	(0.093)	
LEGAL		0.161**	0.260***	
		(0.075)	(0.079)	
Prior Agreement		-0.263***	0.023	
		(0.056)	(0.086)	
Exporter (log) Real GDP		0.201***	0.149***	
		(0.035)	(0.038)	
Importer (log) Real GDP		0.144***	0.105***	-0.119***
		(0.028)	(0.030)	(0.045)
Exporter (log) GDP per capita		-0.070	-0.003	
		(0.082)	(0.108)	
Importer (log) GDP per capita		0.173***	0.213**	0.097
		(0.049)	(0.086)	(0.079)
Constant	0.349***	0.349***		
	(0.029)	(0.026)		
Agreement FEs			x	
Agr.×pair FEs				x
Observations	908	908	908	908
R^2	0.088	0.261	0.427	0.732
Within R^2			0.112	0.096

Robust standard errors reported in parentheses. The dependent variable is $\beta_{A:d}$, an estimated direction-specific FTA partial effect which we have estimated in a prior first stage. "Import Demand Elasticity" refers to aggregate (manufacturing) import demand elasticity estimates constructed using country-specific 3 digit HS import demand elasticities taken from Broda *et al.* (2006). * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table B.10: Robustness: Alternate Weighting Methods (FGLS & WLS)

	Dependent variable: First stage directional FTA estimates							
	FGLS				*WLS*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First stage pair FE†	-0.183***	-0.348***	-0.259***		-0.142***	-0.281***	-0.225***	
	(0.030)	(0.036)	(0.055)		(0.023)	(0.029)	(0.036)	
$\ln \widehat{ToT}_{A:j}$	-16.489***	-9.659***	-7.339**	-8.787***	-20.408***	-11.600***	-6.076	-6.313
	(2.320)	(2.675)	(3.142)	(3.184)	(3.091)	(3.447)	(3.797)	(4.081)
Extensive margin of trade	1.061***	0.518**	0.336	-0.555*	0.672***	0.381**	0.208	-0.378
	(0.157)	(0.234)	(0.242)	(0.296)	(0.131)	(0.191)	(0.204)	(0.260)
ln DIST		-0.416***	-0.238***			-0.319***	-0.216***	
		(0.040)	(0.079)			(0.033)	(0.065)	
CONTIG		0.169**	-0.008			0.121*	-0.079	
		(0.086)	(0.101)			(0.072)	(0.083)	
COLONY		0.014	0.078			-0.006	0.097	
		(0.093)	(0.109)			(0.075)	(0.089)	
LANG		0.110	0.073			0.096	0.058	
		(0.075)	(0.092)			(0.061)	(0.079)	
LEGAL		0.122*	0.255***			0.060	0.147**	
		(0.068)	(0.078)			(0.055)	(0.059)	
Prior Agreement		-0.244***	0.000			-0.194***	-0.002	
		(0.052)	(0.083)			(0.044)	(0.058)	
Exporter (log) Real GDP		0.194***	0.152***			0.163***	0.130***	
		(0.032)	(0.037)			(0.025)	(0.030)	
Importer (log) Real GDP		0.142***	0.091***	-0.135***		0.120***	0.087***	-0.094***
		(0.023)	(0.029)	(0.042)		(0.018)	(0.021)	(0.035)
Exporter (log) GDP per capita		-0.104	-0.005			-0.179***	-0.092	
		(0.068)	(0.105)			(0.046)	(0.075)	
Importer (log) GDP per capita		0.150***	0.231***	0.124		0.092***	0.174***	0.192***
		(0.043)	(0.084)	(0.077)		(0.034)	(0.064)	(0.062)
Constant	0.328***	0.334***			0.255***	0.298***		
	(0.028)	(0.025)			(0.024)	(0.023)		
Agreement FEs			x				x	
Agr.×pair FEs				x				x
Observations	908	908	908	908	908	908	908	908
R^2	0.075	0.256	0.423	0.727	0.075	0.256	0.423	0.727
Within R^2			0.108	0.086			0.108	0.086

Second stage estimates in columns 1-4 are obtained using an efficient Feasible Generalized Least Squares (FGLS) method proposed by Hanushek (1974) for cases in which the dependent variable has been estimated with error in a prior stage. Estimates in columns 5-8 are obtained using Weighted Least Squares (WLS), weighted by inverse first stage standard error. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table B.11: Decomposition of Variance in FTA Effects

	Source of variance:		
Estimation:	Across agreements	Pairs within agreements	Within pairs
OLS	0.355	0.349	0.296
FGLS	0.350	0.346	0.304
WLS	0.344	0.340	0.316

This table offers "analysis of variance" results for each of the three weighting methods that we employ, focusing on how much variation in our first stage directional estimates represents heterogeneity that we observe across agreements versus within agreements, and within pairs. Our results reveal that the three weighting methods suggest very similar variance decompositions.