Economic Determinants of the *Timing* of Preferential Trade Agreement Formations and Enlargements*

Jeffrey H. Bergstrand†, Peter Egger‡ and Mario Larch§

December 20, 2011

Abstract

Although preferential trade agreements (PTAs) have proliferated over the past 60 years, the formation or enlargement of a PTA remains a rare event. Among 10,585 country-pairs for 57 years (463,289 observations), there have been only 1,560 such events. While a small number of recent studies have examined empirically the economic determinants of the likelihood of a pair of countries having a PTA, this study explains empirically the *timing* of all PTA formations and enlargements from 1950 through 2006 using an econometric duration analysis. Our main and novel goal is to predict (in-sample) a substantive share of these 1,560 PTA events using a parsimonious model with only *economic* variables. Our analysis reveals that we can predict correctly the actual year of entry into force for 21 percent of the 1,560 bilateral PTA formations/enlargements in the period 1950-2006 among the 10,585 pairings of 146 countries using only a few economic variables. Moreover, we can predict correctly 48 percent of these PTA events within a 10-year window using this parsimonious model.

**Key words:** Free Trade Agreements; International Trade; Duration Analysis  
**JEL classification:** F14; F15

---

*Acknowledgements:* To be added.

†*Affiliation:* Department of Finance, Mendoza College of Business, University of Notre Dame, Kellogg Institute for International Studies, and CESifo Munich. *Address:* Department of Finance, Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556 USA. *E-mail:* bergstrand.1@nd.edu.

‡*Affiliation:* ETH Zürich, CEPR, WIFO, CESifo, ifo, GEP at University of Nottingham, and OUCBT at Oxford University. *Address:* ETH Zürich, WEH E6, Weinbergstrasse 35, 8092 Zürich, Switzerland. *E-mail:* egger@kof.ethz.ch

§*Affiliation:* University of Bayreuth, CESifo, Ifo Institute, and GEP at the University of Nottingham. *Address:* University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany. *E-mail:* mario.larch@uni-bayreuth.de
1 Introduction

One of the most notable economic events since World War II is the proliferation of preferential trade agreements (PTAs), including free trade agreements (FTAs) and customs unions (CUs). The study of such agreements has followed fundamentally two paths, one normative and one positive. The normative path is whether or not PTAs are welfare-improving (see Bagwell and Staiger, 1997, 2005; Bond, Riezman, and Syropoulos, 2004). A full survey of this literature is beyond the scope of this paper, but see Baldwin (2007) for an excellent recent survey.

The second path, which is “positive,” examines what factors explain and predict which pairs of countries have PTAs. Building on the work of Krugman (1991a, 1991b) and Frankel (1997), Baier and Bergstrand (2004) introduced asymmetric absolute and relative factor endowments into a Krugman-type increasing-returns/monopolistic-competition model to show theoretically that the net utility gains from a bilateral PTA depend on two countries’ economic sizes and their economic similarity, bilateral distance, and relative factor endowments. Using a single cross-section for 1996, Baier and Bergstrand (2004), or BB, employed a probit analysis to show that these economic factors that tend to improve on net a country-pair’s utility from a PTA also tend to increase the pair’s probability of having a PTA. Egger and Larch (2008), Baldwin and Jaimovich (2009), and Chen and Joshi (2010) confirmed BB’s findings and showed that additional variables—such as pre-existing PTAs—also tended to increase the likelihood of a pair of countries having an agreement.

Despite the proliferation of PTAs over the past 60 years, the formation or enlargement of a PTA is a rare event. Among 10,585 pairings of 146 countries for 57 years from 1950 through 2006 (463,289 observations), there have been only 1,560 such events. This study attempts to explain and predict the actual timing of formations and enlargements of all PTAs among virtually all countries in the world in the last 57 years.

In econometrics, the analysis of the time elapsed until a certain event occurs falls under “duration analysis.” Duration analysis has its origin in survival analysis, which refers to the survival time of a subject in a particular state. In our context, this refers to the survival of

\[\text{Egger and Larch (2008) and Baldwin and Jaimovich (2009) examined the influence on the probabilities of PTAs of existing nearby agreements in a previous period using spatial econometrics, providing broad empirical support for potential trade diversion inducing nonmembers to either join existing PTAs (supporting Baldwin’s domino theory) or form new ones. A similar analysis motivated by network formation was undertaken by Chen and Joshi (2010). However, these studies did not examine the effects of existing PTAs on the timing of new PTAs and enlargements.}\]
a country-pair in the state of “No-PTA.” Central to such analyses is the hazard rate, which in our context emphasizes the conditional probability of a country-pair leaving the state of No-PTA *conditional upon* having been in this state for a particular duration. So, the key difference of this paper from the existing literature is predicting the *specific year* of a PTA formation/enlargement event (or a window of years leading up to the event), not just the PTA’s existence, using a parsimonious economic model.

In this paper, we offer three potential contributions. First, we address how one might interpret the decision of a pair of countries to form/enlarge a PTA in some year $T$ *conditional upon* not having had a PTA until year $T$. This discussion informs us about the determinants of the “hazard rate” (without economic covariates), which is the probability of a country-pair leaving the No-PTA state in year $T$ conditional upon having been without a PTA up until $T$. Classic distributions determining hazard rates include the Weibull and log-logistic distributions, which yield that such hazard rates are functions of time trends. However, a simple time trend in the absence of economic covariates can explain only 11 percent of the variation in the PTA events. Second, most economic duration analyses are concerned with the influence of time-invariant or time-varying economic covariates that “shift” the hazard rate in any year $t$. Drawing upon the recent literature on economic determinants of PTAs noted above, we motivate the inclusion of a set of 11 economic covariates that likely influence the probability of two countries forming/enlarging a PTA in any particular year, conditional upon not having a PTA until that year. These covariates include time-invariant bilateral distance and a dummy for sharing a land border and time-varying economic size, economic size similarity, remoteness, per capita GDP differences (and their square), distance of a country-pair to the nearest PTA, number of members in the nearest PTA (and its square), and the overall “degree of competitive liberalization” a pair faces. These economic covariates increase explanatory power to nearly 30 percent. Third, our analysis reveals that we can predict correctly the *actual year* of entry into force for 21 percent of the 1,560 bilateral PTA formations/enlargements in the period 1950-2006 among 10,585 pairings of 146 countries (463,289 observations) using our parsimonious duration model. Moreover, we can predict correctly 48 percent of these PTA events within a 10-year window leading up to the date of entry using the same model. Including time-invariant bilateral fixed effects to control for political and cultural factors increases the share of correctly predicted
PTA events within a 10-year window to 72 percent.\footnote{Only one other study has examined determinants of the timing of PTA events, Liu (2008). However, that study focused instead on the political economy determinants of PTA timings (specifically, income-inequality’s interactive effects with relative capital-labor endowment ratios) to test a “median-voter” model of PTA timing versus a “lobbying” model. That study’s finding in favor of the median-voter model provides support for our economic determinants of the timing of PTA approach and our alternative focus on domino effects, competitive liberalization, and PTA interdependence. Moreover, Liu (2008) does not provide any predicted probabilities of the PTA-formation year, which is the main goal of our paper.}

The remainder of the paper is as follows. Section 2 motivates the use of an econometric duration model to analyze the timing of PTA (formation/enlargement) “events.” Section 3 motivates the time-invariant and time-varying economic determinants of the hazard rate. Section 4 describes the data and Section 5 provides the empirical results. Section 6 concludes.

2 Motivation for a Duration Analysis of Time-to-PTA

2.1 Analyzing PTA Status versus Timing of PTAs

Figure 1 illustrates the years in which PTA events occurred, summarizing the (cumulative) number of bilateral trade agreement “membership events” for all years from 1950 to 2006 in which some new “membership” occurred (either a new agreement or an enlargement), according to information mainly provided by the World Trade Organization (WTO).\footnote{The cumulative number of all PTA events corresponds to the number of country-pairs in the sample which liberalized trade preferentially since 1950. We had to amend the data to capture agreements not included in the WTO data base. For instance, most of the members of the former communist bloc were engaged in agreements outside the WTO (or, prior to the foundation of the WTO, the General Agreement on Tariffs and Trade, GATT). Note that an event requires that explanatory variables employed in the subsequent empirical analysis are not missing for a country-pair. Accordingly, the memberships associated with this figure are the same ones used in estimation later. Figure 1 presents these cumulative events as a percentage of 10,585 country-pairs covered.} We focus on memberships at the country-pair level and avoid redundant observations by counting events such as the membership of France and (West) Germany in the original European Economic Community (EEC) as a single new membership “event” (instead of two events) and the enlargement of the original EEC to include the United Kingdom as five new membership events instead of ten events.\footnote{For instance, the 10 events in 1958 correspond to the formation of the EEC in that year. The corresponding number of memberships is 10, because Belgium and Luxembourg are counted as a single country (as often done in economic studies), so that there are 5 founding members and the number of unique dyads is } Accordingly, the events should not be interpreted as just new PTAs that have been formed, since we also count as new memberships ones that are brought about through enlargements of existing PTAs. In this study, we do not separate our empirical sample into “new PTAs” versus “enlarged PTAs.” But we also show in Figure 1 the subset of PTA
events that include only FTAs and CUs.

The figure suggests that there have been years with strong and weak membership activity over time. The number of all PTA membership “events” concluded since 1950 rose to 1,560 until the end of 2006, i.e., 0.34 percent of the 469,283 total observations in the panel. From this data, we can create a variable representing the “Time-to-PTA Event,” as done in duration analysis. Our focus will then be to find the economic determinants that explain the “Time-to-PTA Event,” meaning the timing of the formation of a new agreement or an enlargement of an existing PTA agreement.

Our goal in this paper is to predict the duration (in years) before a country-pair formed (or enlarged) a PTA using a duration model with a parsimonious set of time-invariant and time-varying economic variables. This contrasts with the goal of Baier and Bergstrand (2004), or BB, that was to explain which country-pairs had a PTA in a given year (specifically, 1996). The econometric framework employed there was the qualitative choice model of McFadden (1975, 1976). In their framework, the probability of a PTA was linked heuristically to an underlying latent variable, denoted $\Delta U_{ij}$ here. In that study’s context, $\Delta U_{ij}$ represented the difference in utility levels from an action (formation of a PTA), where

$$
\Delta U_{ij} = x_{ij}\beta + e_{ij}, \tag{1}
$$

and $x_{ij}$ denoted a vector of explanatory variables (i.e., economic characteristics) of country-pair $ij$ including a constant, $\beta$ was a vector of parameters, and error term $e_{ij}$ was assumed to be independent of $x_{ij}$ and to have a standard normal distribution. In the context of the model formally, $\Delta U_{ij} = \min(\Delta U_i, \Delta U_j)$ where $\Delta U_i(\Delta U_j)$ denoted the change in utility for the

---

5 The PTAs other than FTAs and CUs are PTAs that provide either one-way (e.g., GSP) or two-way preferential trade, but not free trade.

6 Note that 510 “events” correspond to FTAs and CUs only.

7 Figure 1 illustrates the three apparent “waves of regionalism” since 1950. The first wave (beginning in 1958) was initiated by the formation of the European Economic Community (EEC) and then the European Free Trade Agreement (EFTA). The second wave (beginning in 1973) included several enlargements of the EEC and the introduction of the Generalized System of Preferences for developed-developing countries’ trade. The third wave (beginning in 1989) included the formations of the Canada-U.S. FTA, NAFTA, MERCOSUR, and numerous bilateral agreements.

8 Egger and Larch (2008), Baldwin and Jaimovich (2009), and Chen and Joshi (2010) – the three existing papers in this literature – all followed the methodology in BB in predicting the existence of a PTA.
representative consumer in \( i (j) \). Hence, both countries’ representative consumers needed to benefit from a PTA for their governments to form one.\(^9\) The latent variable \( \Delta U_{ij} \) was assumed to generate the binary indicator variable of PTA membership, \( PTA_{ij} \), which was unity if two countries had a PTA and zero otherwise. The response probability for a PTA, \( P \), was then:

\[
P(PTA_{ij} = 1) = P(\Delta U_{ij} > 0) = G(x_{ij}\beta),
\]

where \( G(\cdot) \) was the standard normal cumulative distribution function, which ensured that \( P(PTA_{ij} = 1) \) was between 0 and 1. This literature assumed that \( P(PTA_{ij} = 1) > 0.5 \) “indicated” \( \Delta U_{ij} > 0 \) and \( P(PTA_{ij} = 1) \leq 0.5 \) indicated \( \Delta U_{ij} \leq 0 \).

Rather than focusing on the explanation of PTAs in a cross section of data, this paper aims at examining the determinants of the timing of PTA events. The empirical analysis of the length of time elapsed until certain events occur is referred to as “duration analysis.” Duration models fall within the class of “limited dependent variable” models in general and censored regression models in particular (cf. Wooldridge, 2010). Duration analysis has been used increasingly in the economics literature since 1980. The most common application is in labor economics evaluating empirically the determinants of the length of a spell of unemployment of an individual, cf., Heckman and Singer (1984) and Kiefer (1988). There is only a small number of studies which have applied this framework in international trade.\(^10\)

While there are now a few papers examining empirically determinants of the probability of two countries’ having a PTA, this is the first paper to motivate an economic analysis of time-to-PTA events in a systematic way and pursue an associated empirical assessment. For this, two issues have to be addressed. First, we have to rationalize an empirical model of timing of PTAs in the absence of time-invariant and time-varying economic covariates. Second, we address how the hazard rate interacts with fundamental economic variables that are known to shift a country-pair’s probability of forming an agreement at any point in time. We address these issues separately below.\(^11\)

---

9Hence, side (or compensation) payments are ruled out.
10Joyce (2005) and Conway (2007) studied determinants of the lengths of spells of IMF programs. Besedes (2008), Besedes and Prusa (2006a,b), Fugazza and Molina (2009), Nitsch (2009), and Hess and Persson (2010) all have studied the determinants of bilateral trade-flow durations.
11The vast number of pairs of countries that form PTAs in our sample (1,560 events) do so permanently; our sample includes only 48 events of pairs ending agreements. By contrast, the labor economics duration literature on unemployment spells and the international economics literature on IMF program spells deal with macroeconomic policies/environments, where entering and exiting spells of unemployment or programs, respectively, is


2.2 Economic Motivation for a Discrete Time Duration Model

In this section, we discuss a simple economic motivation for a discrete time duration model for analyzing time-to-PTA events. Suppose that, at the country-pair $ij$ level in any year $t$, each of two governments choose between two states, entering a bilateral PTA or not. Since all gradual tariff reductions are governed by multilateral tariff law in compliance with WTO rules, assume that they can not change multilateral tariff rates vis-à-vis each other, except by abolishing tariffs through a PTA. The decision of interest is the duration of years $T_{ij}$ after which governments $i$ and $j$ will adopt a bilateral PTA. Data on elapsed time since some fixed year until the inception of a PTA are only available by year. Hence, we can not portray time-to-PTA events by a continuous process econometrically, but need to resort to a discrete time representation.

In the No-PTA state and year $t$, assume the governments of $i$ and $j$ to receive utility $U'_{ij}(t)$ associated with bilateral trade flows. These countries may receive the possibility in any year $t$ to form a PTA. From that, they would realize cum-PTA utility (associated with cum-PTA trade flows) drawn from a continuous distribution with density $f(U_{ij}(t))$ at a constant rate $g$ in every year $t$. The probability of realizing the benefits $U_{ij}(t)$ from concluding a PTA (and the associated trade flows) in a time interval of length $T$ is $gT$. Suppose that the possibility to conclude PTAs is drawn independently of $f(U_{ij}(t))$, and governments know the density function $f(\cdot)$ but not the utility $U_{ij}(t)$ from a given PTA. Moreover, suppose for simplicity that reservation utility $U'_{ij}(t)$ from staying outside of a PTA is independent of the gain in trade flows induced by the conclusion of the PTA, while the change in utility through PTA formation is a function of the change in trade flows but not of the functional form $f(\cdot)$. Upon receiving the possibility to conclude a PTA at random intervals, governments then decide about when to form a PTA. The decision about when to enter a PTA will then depend on the comparison of the expected gains from PTA membership as captured by $U_{ij}(t)$ with the reservation gains captured by $U'_{ij}(t)$. Accordingly, the probability that a PTA is acceptable can be written as:

$$P_{ij}(t) = \int_{U'_{ij}(t)}^{\infty} f(U_{ij}(t))dU_{ij}(t).$$

(3)

Country-pair $ij$’s transition rate from No-PTA to PTA in year $t$ is reflected in the product of frequent. Due to the insignificant number of PTA exits, we do not address the latter events in this study.
the constant rate at which PTAs become possible in any year \((g)\) and the probability that they are acceptable in year \(t\) \((P_{ij}(t))\). This transition rate is the probability of leaving the No-PTA state in \(t\) given that governments \(i\) and \(j\) did not conclude a PTA until then, and it may be referred to as the hazard rate for the distribution of duration until PTA formation:

\[
\lambda_{ij}(t) = P(T_{ij} = t | T_{ij} \geq t) = gP_{ij}(t).
\]

(4)

There are several possible distributions for modeling the hazard rate. For a discrete time multivariate model as the one proposed here, the complementary log-log distribution is a common choice (see Jenkins, 1995, 2005). In comparison to other distributions, this specification has the advantage of allowing for time-varying covariates – an essential element of our study discussed in the next section. Figure 1 suggests that, apart from the covariates in vector \(x_{ijt}\), we may want to allow for a general time trend in a specification of time-to-PTA events. The latter can be easily done with the complementary log-log distribution when specifying \(\lambda_{ij}(t)\) as:

\[
\lambda_{ij}(t) = 1 - \exp \left[ - \exp \left( x_{ijt} \beta + \gamma_t \right) \right],
\]

(5)

where \(x_{ijt} \beta = \beta_0 + \beta_1 x_{1,ijt} + \beta_2 x_{2,ijt} + \ldots + \beta_K x_{K,ijt}\) and \(\gamma_t\) captures the general time trend.

We model the general time trend as \(\gamma_t = r \ln(t)\), which implies that the shape of the hazard monotonically increases if \(r > 0\), decreases if \(r < 0\), or is constant if \(r = 0\). If concluding PTAs becomes generally easier as time marches on, we would expect \(r > 0\).

Figure 2 illustrates the relationship between the hazard rate and time \(t\) with a complementary log-log distribution function. The slope of the hazard rate depends on the coefficient of the general time trend \(r\) as well as the explanatory variables. Consistent with estimates reported later on for the data underlying this study, we assume that \(r = 0.61\) in Figure 2.\(^{12}\) If other explanatory variables do not matter and the hazard rate \(\lambda_{ij}(t)\) of leaving the initial No-PTA state increases with larger \(t\), we obtain a functional relationship as shown by continuous concave locus labeled “Time trend only” in Figure 2 with \(\lambda_{ij}(t) = 1 - \exp \left[ - \exp \left( 0.62 \ln(t) \right) \right]\). One may

\(^{12}\)See Jenkins (2005).

\(^{13}\)This result is consistent with a time trend parameter of 0.61 in Specification 10 in Table 2 below.
think of many factors underlying a rising hazard rate with time per se. For instance, falling tariffs due to multilateral trade liberalization under the General Agreement on Tariffs and Trade or generally declining political costs of political and economic cooperation after World War II could be mentioned here.

Beyond a trend, the timing of PTAs is likely influenced by political, cultural, and economic factors. Clearly, variables that have been found to increase the probability of concluding a PTA in previous research will lead to a reduction of the time-to-PTA events with processes as specified in (3) and (5). The reason is that everything that influences the net utility gain of country-pair \(ij\) from participating in a PTA in any year \(t\) will also raise the hazard rate \(\lambda_{ij}(t)\). Hence, obvious candidate variables in \(x_{ijt}\) to predict (in-sample) a substantive share of the 1,560 PTA events underlying Figure 1 are the determinants of PTA memberships in a cross section of data in BB. For instance, having a common land border (a time-invariant variable) or two countries being jointly economically larger (a time-variant variable) are strong partial predictors of PTA membership in a cross section. Either of these factors raises the probability that a PTA is acceptable, \(P_{ij}(t)\), ceteris paribus. Hence, either of these factors raises the hazard rate \(\lambda_{ij}(t)\). However, time-invariant and time-variant elements in \(x_{ijt}\) affect the hazard rate in functionally different ways. This is illustrated by the two broken loci in Figure 2. For the locus labeled “Including positive time-invariant regressor”, we added 0.52 to the the time trend so that \(\lambda_{ij}(t) = 1 - \exp \left(-\exp (0.52 + 0.62 \ln(t))\right)\). For the locus labeled “Including positive time-variant regressor”, we added \(0.10 \varepsilon_{ijt}\) to the the time trend, where \(\varepsilon_{ijt}\) is drawn randomly from a normal distribution with mean and standard deviation of one. Then, \(\lambda_{ij}(t) = 1 - \exp \left(-\exp (0.10 \varepsilon_{ijt} + 0.62 \ln(t))\right)\). As can be seen from Figure 2, there is a tendency for both time-invariant and time-variant shifters of \(P_{ij}(t)\) to raise the hazard rate. However, time-variant shifters of \(P_{ij}(t)\) render the hazard rate a potentially non-monotonic function of time, which is not the case for time-invariant shifters.

\[A\] parameter on common borders of 0.52 is also consistent with Specification 10 in Table 2 below. \[B\] parameter of 0.10 is, e.g., consistent with the coefficient on two countries’ joint economic size in Specification 10 in Table 2 below. However, we do not draw \(\varepsilon_{ijt}\) from a normal distribution with mean and variance as of two countries’ joint economic size for reasons of illustration.
3 Economic Factors Shifting the Probabilities of PTAs

The purpose of this section is to identify economic variables that potentially “shift” the hazard rate, $\lambda_{ij}(t)$, in any year $t$ ($X_t$), thus increasing or decreasing the likelihood that a PTA occurs sooner. In the spirit of the extant duration analysis literature, we consider determinants of the probability of PTAs suggested in recent studies by BB and Egger and Larch (2008), as such variables are expected to alter in any period $t$ the latent variable $U_{ij}(t)$ defined earlier.

That such a notion is plausible is illustrated readily in Figure 3. While charts such as Figure 1 are typical, Figure 3 provides a novel graph. It takes the information associated with Figure 1 and combines the Time-to-PTA event data with three particular economic characteristics associated with the members of PTAs relative to those of nonmembers. One economic characteristic is the “proximity” of PTA members relative to nonmembers. We measure this using the average distance between the economic centers of members of PTAs relative to the average distance between economic centers of nonmembers. A second economic characteristic is the average economic size of PTA members relative to that of nonmembers; economic size is measured using countries’ GDPs. The third economic characteristic is the average difference between country pairs’ GDPs for PTA members relative to that of nonmembers.

Figure 3 illustrates several profoundly systematic relationships between distance, economic size, economic similarity, and the timing of PTA events. The bottom line indicates two phenomena. The earliest PTA events (1958-1961) were between members whose average distance between members relative to nonmembers was the smallest. As time passed, the average distance between members relative to nonmembers rose systematically. This line suggests that PTAs formed or enlarged sooner among closer countries. The middle line indicates two phenomena related to economic size and PTA-event timing. The earliest PTA events were also between countries whose average economic size was the largest relative to nonmembers. Then, as time passed, the average relative economic size of members declined. This line suggests that PTAs formed or enlarged sooner among economically larger countries. The top line indicates

16 In short, period $t$ economic variables influence $P_{ij}(t)$ in the equation $\lambda_{ij}(t) = gP_{ij}(t)$, because $P_{ij}(t)$ is a function of period $t$ utility $U_{ij}(t)$.

17 “Similarity” is measured (as traditionally) using the product of country $i$’s share of both countries’ real GDPs with country $j$’s share.
two phenomena related to economic-size similarity and PTA-event timing. The earliest PTA events were also among countries with very similar GDP sizes. As time passed, the degree of size similarity declined in general. This line suggests that PTAs formed or enlarged sooner among countries with more similar economic sizes.

BB provided a formal theoretical economic foundation for the relationships between a country-pair’s bilateral proximity, economic size, and economic similarity for influencing the probability of an PTA, based upon a Krugman-type model of trade. First, two countries’ governments will form a PTA the smaller the distance between them. For a given distance between the country-pair and the ROW, the closer are the two countries, the lower their transport costs of international trade and consequently the higher is their trade volume. Elimination of the ad valorem tariff between close (and analogously adjacent) members alleviates price distortion on a large amount of trade, improving utility of consumers more. Second, a PTA between two large economies increases the volume of trade in more varieties than a PTA between two small economies, and reduces trade in fewer varieties from nonmembers than two small economies, improving utility more in large GDP countries relative to small GDP countries. Third, two economies that are more similar in economic size (due to more similarly sized absolute factor endowments) are more likely to trade. For a given economic size of two countries, if one of the countries has virtually all of the pair’s factor endowments, formation of a PTA provides little welfare gain to the large economy because it trades little with the small economy because the latter produces few varieties.

BB also addressed theoretically and empirically two other economic variables influencing the probability of a PTA. First, two countries were more likely to benefit from a PTA the more remote the pair was from the ROW. As the distance between a pair of countries with the ROW increases, the volume of trade with remote countries decreases and that with near countries increases. With less trade with remote countries, the relative tariff distortion with remote countries from a PTA between a pair has less impact on utility; in other words, there is less “trade diversion.” With more trade with near countries, the elimination of the tariff distortion has greater utility gains. Second, BB showed (because it allowed for two industries and two factors) that the net welfare gain from a PTA between a country-pair increases with relative capital-labor ratios, but eventually might decline due to increased specialization. As relative factor endowments widen, both countries in the pair specialize more in the industries where
they have comparative advantages and enjoy more net welfare gains from a PTA. However, BB also showed that – at low intercontinental transport costs – the net gains from a PTA increase at first but then eventually fall with widening relative factor endowments. Initially, each country gains from specialization in one of the two industries. Yet at high levels of specialization, each country relies more on intra- and inter-continental trade to meet its demand for varieties of the other industry. With increasing specialization, the net welfare gains from inter-industry trade are eventually offset by the net trade diversion associated with intra-industry trade. Thus, we have a theoretical motivation for seven economic variables influencing the hazard rate: bilateral distance, common land border, sum of the pair’s GDPs, similarity of the pair’s GDPs, remoteness of the pair from the ROW, per capita GDP differences of the pair (capturing the pair’s relative capital-labor differences, in a large sample), and the square of per capita GDP differences of the pair.

Egger and Larch (2008), or EL, enriched the BB framework by examining the role of existing PTAs for influencing the likelihood of formation of new or enlargement of existing PTAs, motivating four more variables for our duration analysis. Drawing on the domino theory in Baldwin (1995), EL argued that current PTAs’ existence was influenced by “interdependence” with (in panel) other existing PTAs. Given a PTA forms, outsiders will lose in utility due to trade diversion. This creates, under some conditions, an incentive for them to join an existing PTA, or under alternative conditions form a new PTA.

EL applied spatial econometrics to introduce a “spatial lag” into the BB framework. This approach is beyond the scope of this paper. However, we introduce four new variables to this literature to capture the influences of existing agreements on new or enlarged PTAs. First, we include a measure of distance of a pair of countries to the “nearest PTA” ($DISTPTA$). Intuitively, the closer are two countries to an existing PTA, the greater is the trade diversion they have incurred from that PTA. This implies a greater economic incentive to form/enlarge a PTA because of the potentially offsetting trade creation. Second, a country-pair’s utility is influenced by the “degree of regionalism” (or “competitive liberalization”) in the ROW. The greater the number of PTAs in the ROW, the more trade diversion and loss of utility a country-pair experiences. We include a variable measuring the “degree of regionalism” in the ROW for every pair which is a spatially-weighted average of all the PTAs that countries $i$ and $j$ face in ROW, denoted $WPTA$. Third, the variable $DISTPTA$ influences – in the
terminology of Baldwin (1995) – the “demand for membership” of outsiders into an existing PTA or a new PTA. However, in Baldwin (1995), the “supply of membership” was purposely assumed to be infinitely elastic. In reality, PTA membership is also constrained potentially by existing members and potential new PTA partners; that is, supply of membership may have finite elasticity. In a formal theoretical model, we are able to show that the likelihood of a PTA between a country-pair may be constrained by the number of members in the “nearest PTA,” as some members of the existing agreement suffer sufficient trade diversion from other existing members as a result of a potential new entrant that these “marginally worse-off” members prevent entry. We capture this new influence with a variable \( NPTA \) (and its squared value, \( SQNPTA \)), which is the actual number of members of the nearest existing PTA (and the square of that number). These four variables alongside the seven variables motivated by BB suggest 11 economic covariates to examine in our duration analysis.\(^{18}\)

4 Data

The data set for the timing-of-PTA events was compiled for the period 1950-2006 using information from notifications to the World Trade Organization (WTO), the CIA World Fact Book, and individual web pages of countries. Regarding explanatory variables, to capture whether two countries are “natural” trading partners, we include two variables. \( DIST_{ij} \) is the (natural) logarithm of the great circle distance between the capitals of countries \( i \) and \( j \). The CIA World Fact Book was used to generate the dummy variable for two countries having a common land border \( (BORDER_{ij}) \). Countries’ economic sizes are measured using real GDPs from Maddison (2003). \( GDPSUM_{ij,t} \) is the log of the sum of the two countries’ real GDPs. \( GDPSIM_{ij,t} \) is the log of the similarity of two countries’ RGDPs, where “similarity” is measured (as traditionally) using the product of country \( i \)’s share of both countries’ real GDPs with country \( j \)’s share \( (s_is_j) \). Measurement of minimum distances of countries to the nearest existing PTAs uses great circle distances also; \( DISTPTA_{ij,t-5} \) measures the minimum distance of a country pair to the closest PTA at time \( t - 5 \). \( WPTA_{ij,t-5} \) was defined in detail earlier. \( NPTA_{ij,t-5} \) measures the number of member countries in the closest PTA at time \( t - 5 \) and \( SQNPTA_{ij,t-5} \) is the square

\(^{18}\)All of these economic variables (with the exception of relative factor endowments) can be motivated simultaneously in a simple one-sector Krugman-type formal theoretical model available in a Theoretical Supplement on the authors’ websites (and appended to the end of this manuscript).
of \( NPTA_{ij,t-5} \).

5 Empirical Results

We separate the discussion of empirical results into three parts. First, we discuss the main empirical results for a time trend only and then sequential inclusion of the seven economic covariates. Second, we address the sensitivity of the results by adding four additional exogenous variables and then by allowing for unobserved heterogeneity. Third, we examine the ability of the model for predicting the actual year of the formation of a PTA between each pair of countries, as well as for various “windows” preceding PTA events.

5.1 Main Empirical Results

Table 1 provides the results from estimating economic determinants of the instantaneous probability of leaving the initial state (No-PTA) in the interval \( [t, t+dt) \) given survival up until time \( T \) (i.e., the hazard rate). Specification 1 provides the results of estimating the hazard rate on the time trend. We find evidence of positive duration dependence. In the absence of economic covariates, this result suggests that the longer a country-pair has had no PTA, the higher the probability in any period \( t \) that it will enter a PTA. This could be attributable readily to the effects of “competitive liberalization” but also to numerous other economic variables that have been omitted and cause instantaneous shocks to the probability of forming/enlarging a PTA, among them the “degree of regionalism (or competitive liberalization)” a pair faces, which we examine.

The next set of specifications we discuss (2-4) correspond to Figure 3. Specification 2 confirms our expectation about the relationship between distance and the hazard rate. A lower distance between a country-pair and sharing a land border likely reduces trade costs, encouraging more trade. With a larger trade volume, the utility gains for pair \( ij \) from a PTA increase, raising the hazard rate for a PTA. Yet, the adjacency dummy variable has an unexpected negative (though statistically insignificant) effect on the hazard rate, since sharing a common land border is also likely to reduce trade costs. However, we will find in a fuller specification shortly that this sign is reversed. Specification 3 confirms that larger economic
(GDP) size has a positive effect on $\lambda$. Two larger economies exchange a larger variety of goods. Consequently, the gains from an PTA are higher, increasing $U_{ij}$, and thus $\lambda$. Specification 4 confirms that more similar economic sizes of the two countries – other things equal – increases the hazard rate also. If one of the countries has most of the varieties consumed, there is little incentive for trade, so the gains to an PTA are smaller.

The next set of specifications (5-7) in Table 1 refer to models that include a role for “existing PTAs,” or “interdependence” in EL’s terms or “contagion” in terms of Baldwin and Jaimovich (2009). Specification 5 adds the distance of the country-pair $ij$ to its nearest PTA. In line with expectations, the closer a country-pair is to its nearest PTA the larger the trade diversion it experienced by being non-members. Hence, there is a larger potential gain from a PTA to offset this trade diversion. Specification 6 adds our measure of the “degree of regionalism (or competitive liberalization)” in the ROW. In principle, the greater the degree of regionalism in the $ROW$ measured by $WPTA$, the larger the economic incentive to form/enlarge a PTA. This is confirmed empirically with the positive and statistically significant coefficient estimate for $WPTA$. Specification 6 also adds the number of members in the nearest PTA of the pair ($NPTA$), but that coefficient is insignificant; we will discuss that in more detail shortly.

An important result to especially note in Specification 6 is the substantive reduction in the coefficient estimate for the time trend (and its z-statistic). It appears that the time trend $t$ was implicitly picking up the effect of the “degree of regionalism” in the $ROW$. With this economic covariate included explicitly, the time trend is much less important.

While Specification 6 suggests that the number of members in the nearest PTA has no significant effect on the hazard rate, Specification 7 informs us that it does have a positive – but quadratic – effect on the hazard rate. The economic intuition for this result is subtle. In the context of the “domino theory” in Baldwin (1995), an increase in the number of members of the nearest PTA causes a rise in demand for membership of non-members in this agreement, tending to increase the hazard rate for $ij$. However, Baldwin’s domino theory assumes an infinitely elastic supply of membership by a PTA. As the number of members of a PTA increases, there are incumbent members whose utility falls when a new member is added, especially the members of the PTA most distant from the core. Everytime a new member is added, the utility from the PTA of the (marginal) “worse-off” member declines. It can be shown in a simple Krugman-like model like that in BB that at some point the marginal worse-off member’s utility declines from
new member, dampening the average utility gain of members in the PTA. This finite-elasticity-of-membership supply implies theoretically a quadratic relationship between the number of members of the nearest PTA ($NPTA$) and the hazard rate for pair $ij$. Furthermore, note that the effect of the time trend is further diminished by the inclusion of $NPTA$ and its square ($SQNPTA$).

5.2 Sensitivity Analysis

In the spirit of determining the robustness of the empirical results, Table 2 presents several alternative specifications. For ease of comparison, we include Specification 7 from Table 1 as the first specification in Table 2.

5.2.1 FTAs and CUs Only

We have also estimated the model using a different definition of the dependent variable, where only events for FTAs and CUs were taken into account. Since much of the extant literature on economic determinants of PTAs focuses only on FTAs and CUs, we thought it useful to see if the model works for this alternative definition of PTA. Specification 8 reproduces Specification 7 with the new endogenous variable restricted to only FTAs and CUs. As Specification 8 in Table 2 reveals, the results are robust to this modification.

5.2.2 Remoteness from the ROW

Following the motivation from Figure 3, we focused first on the BB economic covariates bilateral distance, economic size, and economic similarity, as well as the EL variables for $DISTPTA$, $WPTA$, $NPTA$, and $SQNPTA$. However, it is possible that the results are sensitive to the omission of the remoteness of a country pair relative to the rest of the world, as BB found for the probability of the existence of a PTA in a single year. The remoteness of natural trading partners is measured as in BB. This variable – labeled $REMOTE_{ij}$ – is the interaction of a dummy variable of 1 (0) for two countries on the same (different) continents, $DCONT_{ij}$, and a measure of “remoteness”:

$$REMOTE_{ij} = DCONT_{ij} \frac{1}{2} \left[ \log \left( \sum_{k=1, k\neq j}^{N} \frac{Distance_{ik}}{N-1} \right) + \log \left( \sum_{k=1, k\neq i}^{N} \frac{Distance_{jk}}{N-1} \right) \right].$$

\[6\]

\[19\]See our Theoretical Supplement.
Specification 9 (based on the original broad definition of the dependent variable) reveals that the results in Specification 7 are insensitive to the presence or absence of this variable. However, the statistical significance of its positive coefficient estimate suggests plausibly that the more remote a pair of countries is, the higher the likelihood that a PTA will form sooner.

5.2.3 Relative Factor Endowments

In reality, the world is not composed of a single sector nor only one factor. The world is composed of many sectors and factors, and trade and trade-policy relationships may be sensitive to relative factor endowments, as suggested in BB. To account for this, we included in Specification 10 two variables reflecting relative factor endowments (in the spirit of B-B). One variable is differences in per capita incomes ($PCY_{DIFF}^{ij,t}$) to proxy for differences in capital-labor ratios. We also included the square of $PCY_{DIFF}^{ij,t}$, referred to as $SQPCY_{DIFF}^{ij,t}$, as Figure 5 in BB (a model with two sectors and two factors) and associated discussion suggests that the probability of a PTA may be a quadratic function of relative factor endowments.

Computing relative capital-labor ratios for a panel of 146 countries from 1950-2006 is virtually impossible due to the absence of a good measure of capital stocks for such a large sample of countries, typically calculated using the perpetual-inventory-based method using capital formation data. Following Egger and Larch (2008), we use per capita real GDPs to proxy for the capital-labor ratios, as the correlation coefficient between capital-labor ratios in B-B and per capita real GDPs is 0.975. $PCY_{DIFF}^{ij,t}$ is calculated as the absolute value of the difference in the logarithm of real per capita incomes of countries $i$ and $j$. Population data are from Maddison (2003).

Specification 10 reveals that the basic results (coefficient estimates) are insensitive to the presence or absence of these variables. However, the statistical significance of the two variables provides evidence that relative factor endowments can affect the timing of PTA events. In general, their coefficient signs suggest that inter-industry (Heckscher-Ohlin) trade likely provides an explanation for the timing of PTAs. The coefficient estimates are consistent with the notion that larger potential Heckscher-Ohlin trade increases the likelihood that a PTA forms sooner, up to a point.

We note that in Table 2, for the most part, the explanatory power is 27 percent, about 16 percent higher than that using the time trend alone. In the next section, we consider the
influence of unobserved heterogeneity.

5.2.4 Controlling for Unobserved Heterogeneity

The model estimated in previous sections explains about one-quarter of the variation in the timing of PTA events. In predicting PTA events, we will use Specification 10 as our primary model. Thus, predictions will be based upon a standard duration model, with a time trend (owing to the assumption of a log-logistic distribution) and the 11 economic covariates.

However, it is useful to see whether the results in Specification 10 are biased by unobserved heterogeneity. The range of possible omitted variables is vast. In order to allow for a set of unobserved effects which may be correlated with $x_{ij}(t)$, we should consider introducing fixed country-pair effects. However, the introduction of fixed effects in a non-linear function is not a trivial endeavor, since the number of incidental parameters is increasing with sample size. Chamberlain (1980) shows that, for a fixed time dimension, maximum likelihood estimates of $\beta$ will be inconsistent as the number of cross-sectional observations goes to infinity. Chamberlain (1980) provides an approach that eliminates the incidental parameters problem. Essentially, the unobserved effects can be eliminated by an appropriate differencing transformation. Chamberlain (1980) suggests including averages of all time-varying explanatory variables along with the original variables in the empirical models (see also Wooldridge, 2010).

Specification 11 reports the results using the Chamberlain-Wooldridge-type model. For brevity, we do not report coefficient estimates for fixed effects; moreover, recall that $DIST$ and $BORDER$ are time-invariant (and so are not de-meaned). First, unsurprisingly the pseudo-$R^2$ increases, from 27 to 38 percent. Second, Specification 11 reveals that most of the coefficient estimates maintain the same qualitative effects as found in the previous specification. In particular, PTA determinants $DISTPTA$, $BORDER$, $GDPSUM$, $GDPSIM$, $DISTPTA$, $WPTA$, and $REMOTE$ all have the expected coefficient signs and their coefficient estimates are statistically significant. However, we find that $NPTA$, $SQNPTA$, $PCFYDIFF$, and $SQPCYDIFF$ do change signs unexpectedly.

5.3 Predicting the Actual Years of PTA Formation or Enlargements

In this section, we examine the predicted timings – specifically, the predicted year – of all PTA events. Because this is a demanding objective, we also consider predicting a “window”
of up to 10 years prior to the “event.” Table 3 provides a summary of the accuracy of the model for predicting the year or window of each bilateral PTA event using Specifications 10 and 11 from the previous section. It is important to note that our preferred specification for predicting events is Specification 10; this is the simplest log-logistic hazard-rate model with only 11 economic variables. We note three conclusions.

First, Specification 10 predicts 320 out of the 1,560 PTA events in the world from 1950-2006 in the actual year of the formation/enlargement. This is 21 percent of the events.

Second, unsurprisingly as we open up a larger and larger “window” of prediction, the number of correct predictions increases. With a 5-year window, Specification 10 predicts 640, or 41 percent, of the 1,560 events. With a 10-year window, the model predicts 752 of the 1,560 PTA timings. This is 48 percent of all PTA events among 10,585 pairings of 146 countries for 57 years. This is a remarkable result in light of the rarity of bilateral PTA events.

Third, for completeness, we also report predictions using Specification 11, which adjusts effectively for “pair fixed effects” by employing the econometrically appropriate Chamberlain time-averaged fixed effects for each time-varying variable. Specification 11 accounts for unobserved heterogeneity. This specification predicts 551 of 1,560 events in the actual year of the event, or 35 percent. With a 5-year window, 62 percent of events are predicted. With a 10-year window, 72 percent of events are predicted.

In sum, even the simplest specification with no fixed effects or year dummies performs quite well, with nearly 50 percent of all PTA events predicted using a 10-year window.

6 Conclusion

Despite the proliferation of PTAs in the last 60 years, there have been only 1,560 bilateral formations/enlargements among 10,585 pairings of 146 countries from 1950-2006. We used an econometric “duration analysis” to determine the economic factors explaining the instantaneous probability at time $T$ of leaving the initial state of “No-PTA” to form or enter a PTA (given survival of the state No-PTA up until time $T$). We found that numerous bilateral economic factors of country pairs – bilateral distance, joint economic size, similarity in economic size – had statistically significant effects on the timing of country-pairs’ PTA “events.” Moreover, we found evidence that PTAs themselves influenced the timing of subsequent PTA events; that
is, we found evidence of “contagion in regionalism,” consistent with recent work by Egger and Larch (2008) on interdependence of PTAs. Our preferred specification explains 27 percent of the variation in the timing of virtually all post-1950 PTAs. Using this specification, our simple duration model with 11 economic covariates can predict 21 percent of all PTA events in the actual year of the formation/enlargement and – using a 10-year window instead – predicts 48 percent of all PTA events.

In reality, political, historical, and cultural factors also matter for explaining the timing of PTA formations and enlargements. Nevertheless, this study provides a benchmark for understanding the full set of determinants of PTA timings, having shown that economic factors can be very influential in explaining a substantive portion of all PTA events since World War II.

References


endogenous trade bloc formation theory on EU data, CEPR Discussion Paper 6389.


Figure 1
Years to Event for Different Groups for the Years 1950-2006

The graph shows the percentage of 10,585 country-pairs for various years since 1950. The data includes all events and CU- and FTA-memberships only.
Figure 2

Illustration of hazard rate as a function of time

- Time trend only
- Including positive time-invariant regressor
- Including positive time-variant regressor

\[ \lambda(t) \]

Years since 1950

[Graph showing the hazard rate as a function of time with different lines representing different scenarios]
Figure 3
Characteristics of RTA Insiders versus Outsiders for the Years 1950-2006

GDP similarity of members relative to non-members

GDP of members relative to non-members

Average distance among members relative to distance among non-members

Index ratio PTA insiders vs. outsiders

Year

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Theory</th>
<th>Spec. 1</th>
<th>Spec. 2</th>
<th>Spec. 3</th>
<th>Spec. 4</th>
<th>Spec. 5</th>
<th>Spec. 6</th>
<th>Spec. 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>+</td>
<td>2.05 (74.56)</td>
<td>2.07 (74.99)</td>
<td>1.93 (69.33)</td>
<td>1.89 (68.27)</td>
<td>1.91 (76.81)</td>
<td>0.77 (43.22)</td>
<td>0.61 (34.14)</td>
</tr>
<tr>
<td><strong>Bilateral Determinants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIST</td>
<td>—</td>
<td>-0.21 (-14.69)</td>
<td>-0.27 (-19.40)</td>
<td>-0.26 (-18.89)</td>
<td>-0.12 (-7.61)</td>
<td>-0.16 (-10.50)</td>
<td>-0.17 (-11.32)</td>
<td></td>
</tr>
<tr>
<td>BORDER</td>
<td>+</td>
<td>-0.02 (-0.27)</td>
<td>-0.08 (-1.08)</td>
<td>-0.12 (5.80)</td>
<td>0.44 (6.47)</td>
<td>0.54 (7.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDPSUM</td>
<td>+</td>
<td>0.17 (27.67)</td>
<td>0.25 (36.67)</td>
<td>0.16 (21.29)</td>
<td>0.17 (23.59)</td>
<td>0.12 (16.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDPSIM</td>
<td>+</td>
<td>0.21 (24.67)</td>
<td>0.15 (17.63)</td>
<td>0.16 (18.87)</td>
<td>0.16 (18.59)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PTA Determinants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISTPTA</td>
<td>—</td>
<td>-0.47 (-47.23)</td>
<td>-0.20 (-16.42)</td>
<td>-0.16 (-13.51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WPTA</td>
<td>+</td>
<td>0.33 (8.56)</td>
<td>5.20 (41.92)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPTA</td>
<td>+</td>
<td>0.001 (0.47)</td>
<td>0.41 (10.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQNPTA</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.004 (-2.72)</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>-10.82 (-106.25)</td>
<td>-9.10 (-58.29)</td>
<td>-9.92 (-63.08)</td>
<td>-10.48 (-66.53)</td>
<td>-9.43 (-58.61)</td>
<td>-10.96 (-38.46)</td>
<td>-10.50 (-36.70)</td>
</tr>
<tr>
<td>Pseudo-R^2</td>
<td></td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.15</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>Log-likelihood (model)</td>
<td></td>
<td>-44,840</td>
<td>-44,720</td>
<td>-44,340</td>
<td>-44,008</td>
<td>-42,670</td>
<td>-37,407</td>
<td>-36,551</td>
</tr>
</tbody>
</table>

**Notes:** *t*-statistics in parentheses. There are 463,289 observations, 10,585 country-pairs, and 1,560 events. All coefficient estimates are correctly signed and statistically significant at the 1 percent level in one-tail t-tests, except BORDER (correct and significant in Specifications 5 and 6). The likelihood value of the constant only model is -50,207. The p-value of the likelihood ratio statistics on the model is 0.
Table 2: Sensitivity Analysis of Economic Determinants of Hazard Rates for Country-Pairs*

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Theory</th>
<th>Spec. 7</th>
<th>Spec. 8</th>
<th>Spec. 9</th>
<th>Spec. 10</th>
<th>Spec. 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>+</td>
<td>0.61</td>
<td>0.68</td>
<td>0.61</td>
<td>0.61</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(34.14)</td>
<td>(37.47)</td>
<td>(34.11)</td>
<td>(33.52)</td>
<td>(33.14)</td>
</tr>
<tr>
<td>Bilateral Determinants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIST</td>
<td>—</td>
<td>-0.17</td>
<td>-0.13</td>
<td>-0.15</td>
<td>-0.15</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-11.32)</td>
<td>(-8.51)</td>
<td>(-8.57)</td>
<td>(-8.66)</td>
<td>(-13.92)</td>
</tr>
<tr>
<td>BORDER</td>
<td>+</td>
<td>0.54</td>
<td>0.09</td>
<td>0.54</td>
<td>0.52</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.32)</td>
<td>(1.20)</td>
<td>(7.25)</td>
<td>(6.99)</td>
<td>(6.42)</td>
</tr>
<tr>
<td>GDPSUM</td>
<td>+</td>
<td>0.12</td>
<td>0.11</td>
<td>0.13</td>
<td>0.13</td>
<td>2.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(16.66)</td>
<td>(14.68)</td>
<td>(16.80)</td>
<td>(17.12)</td>
<td>(58.86)</td>
</tr>
<tr>
<td>GDPSIM</td>
<td>+</td>
<td>0.16</td>
<td>0.09</td>
<td>0.16</td>
<td>0.16</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(18.59)</td>
<td>(10.73)</td>
<td>(18.55)</td>
<td>(18.36)</td>
<td>(32.09)</td>
</tr>
<tr>
<td>PTA Determinants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISTPTA</td>
<td>—</td>
<td>-0.16</td>
<td>-0.34</td>
<td>-0.16</td>
<td>-0.17</td>
<td>-0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-13.51)</td>
<td>(-25.51)</td>
<td>(-13.64)</td>
<td>(-13.81)</td>
<td>(-22.35)</td>
</tr>
<tr>
<td>WPTA</td>
<td>+</td>
<td>5.20</td>
<td>2.03</td>
<td>5.18</td>
<td>5.25</td>
<td>14.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(41.92)</td>
<td>(18.17)</td>
<td>(41.67)</td>
<td>(42.00)</td>
<td>(71.48)</td>
</tr>
<tr>
<td>NPTA</td>
<td>+</td>
<td>0.41</td>
<td>1.37</td>
<td>0.41</td>
<td>0.41</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(10.45)</td>
<td>(21.19)</td>
<td>(10.55)</td>
<td>(10.40)</td>
<td>(-3.30)</td>
</tr>
<tr>
<td>SQNPTA</td>
<td>—</td>
<td>-0.004</td>
<td>-0.03</td>
<td>-0.004</td>
<td>-0.004</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.72)</td>
<td>(-14.66)</td>
<td>(-2.84)</td>
<td>(-2.71)</td>
<td>(1.83)</td>
</tr>
<tr>
<td>Robustness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REMOTE</td>
<td>+</td>
<td>0.01</td>
<td>0.01</td>
<td>12.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.99)</td>
<td>(2.07)</td>
<td>(12.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCYDIFF</td>
<td>+</td>
<td>0.26</td>
<td>-0.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.44)</td>
<td>(-8.24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQPCYDIFF</td>
<td>—</td>
<td>-0.08</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-7.50)</td>
<td>(9.79)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>-10.50</td>
<td>-18.93</td>
<td>-10.70</td>
<td>-10.51</td>
<td>-1.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-36.70)</td>
<td>(-36.69)</td>
<td>(-35.30)</td>
<td>(-34.54)</td>
<td>(-1.44)</td>
</tr>
<tr>
<td>Pseudo-R²</td>
<td></td>
<td>0.27</td>
<td>0.30</td>
<td>0.27</td>
<td>0.27</td>
<td>0.38</td>
</tr>
<tr>
<td>Log-likelihood (model)</td>
<td></td>
<td>-36,551</td>
<td>-35,187</td>
<td>-36,549</td>
<td>-36,521</td>
<td>-31,019</td>
</tr>
</tbody>
</table>

Notes: *z-statistics in parentheses. There are 463,289 observations, 10,585 country-pairs, and 1,560 events. All coefficient estimates are correctly signed and statistically significant at the 1 percent level in one-tail t-tests, except BORDER, GDPSUM, GDPSIM, and REMOTE in Specification 11. The likelihood value of the constant only model is -50,207. The p-value of the likelihood ratio statistics on the model is 0. "na" denotes not applicable.
Table 3: Predicting the Timing of the 1,560 PTA Events Covered*

<table>
<thead>
<tr>
<th>Predicted events</th>
<th>Specification (10)</th>
<th>Specification (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number PTA events</td>
<td>PTA events % of all</td>
</tr>
<tr>
<td>Total number PTA events</td>
<td>1,560</td>
<td>100</td>
</tr>
<tr>
<td>In the same year as the event occurred</td>
<td>320</td>
<td>21</td>
</tr>
<tr>
<td>In the same year as the event occurred or up to 1 year prior to that</td>
<td>384</td>
<td>25</td>
</tr>
<tr>
<td>In the same year as the event occurred or up to 2 years prior to that</td>
<td>416</td>
<td>27</td>
</tr>
<tr>
<td>In the same year as the event occurred or up to 3 years prior to that</td>
<td>541</td>
<td>35</td>
</tr>
<tr>
<td>In the same year as the event occurred or up to 4 years prior to that</td>
<td>551</td>
<td>35</td>
</tr>
<tr>
<td>In the same year as the event occurred or up to 5 years prior to that</td>
<td>640</td>
<td>41</td>
</tr>
<tr>
<td>In the same year as the event occurred or up to 10 years prior to that</td>
<td>752</td>
<td>48</td>
</tr>
</tbody>
</table>

Notes: Underlying cut-off values minimize a quadratic loss function of predicting binary events by the complementary log-log model.
A THEORETICAL SUPPLEMENT (Not Intended for Publication)

Since the focus of our study is on identifying empirical determinants of the timing of PTA events, we choose a minimal general equilibrium model to motivate the eleven observable economic variables for our empirical specifications. Importantly, our purpose here is simply to motivate economic factors that “shift” the hazard rate \( (\lambda_{ij}(t)) \) in any year \( t \). Consequently, a static (one-period) model is sufficient. Following Baier and Bergstrand (2004), we use a simple one-sector Krugman general equilibrium model. However, under a simple assumption that there are costs increasing in the number of agreements bargained (e.g., negotiation “congestion” costs which are quadratic in the number of partners one deals with per period), to every PTA formation that ensures one PTA formation at a time, our model generates endogenously the determination of the sequencing (not timing per se) of PTA events, implying which PTA events occur before others.

A.1 Theoretical Model

A.1.1 Consumers

The model consists of \( N \) countries and one sector. Each country \( j \) hosts a single representative consumer who derives utility from the consumption of goods. Utility is characterized by a taste for variety which is captured formally by Dixit and Stiglitz (1977) preferences with a constant elasticity of substitution (CES). Let \( c_{ij}(k) \) be the consumption in country \( j \) of the differentiated good produced by firm \( k \) in country \( i \). Let \( \sigma \) denote the elasticity of substitution between varieties of goods (assuming \( \sigma > 1 \)). Finally, \( n_i \) refers to the number of varieties produced (and firms) in country \( i \). The utility function \( U_j \) is given by:

\[
U_j = \left[ \sum_{i=1}^{N} \int_{n_i} c_{ij}(k)^{\frac{1}{\sigma}} dk \right]^{\frac{\sigma}{1-\sigma}}.
\]

(7)

We include Samuelson iceberg-type trade costs that are allowed to be asymmetric among all country pairs. We assume that \( 1 + a_{ij} \) units of a good have to be shipped from country \( i \) to ensure that one unit arrives in country \( j \) (assuming \( a_{ii} = 0 \)); in the section discussing calibration of the model for simulations, we will discuss our novel trade-cost structure in detail. Furthermore, we assume ad valorem import tariff rates on goods and services, where \( t_{ij} \) denotes the tariff rate levied by country \( j \) on goods imported from \( i \) (assuming \( t_{ii} = 0 \)).

We assume one factor of production, labor \( (L) \). Each laborer in each country also represents one household.

20A static model cannot say anything about “time” per se, of course. However, under the simple assumption of one PTA event at a time (e.g., assuming that the U.S. International Trade Commission or Congress can only deal with one or a few PTAs at a time), our static model can deliver endogenously the sequencing of events. If certain events occur sooner than others, the time to those events is necessarily shorter. For instance, if our model generates endogenously that countries 1 and 2 form a PTA before countries 1 and 3, this suggests that two countries with the joint economic characteristics of countries 1 and 2 will form a PTA before two countries with the joint economic characteristics of countries 1 and 3. Since we are only interested in motivating observable economic variables, we omit unobservable (“dynamic”) considerations. For more explicit consideration of dynamic aspects of PTA liberalization, see Bond and Park (2002), Zissimos (2007), and Bond (2008).

21By contrast, Baier and Bergstrand (2004) used a common exogenous intra-continental bilateral trade cost factor and a common exogenous inter-continental bilateral trade cost factor.
The consumer is assumed to maximize equation (7) subject to the budget constraint:
\[
\frac{Y_j}{L_j} = \left[ \sum_{i=1}^{N} \int n_i \ p_{ij}(k) c_{ij}(k) dk \right] = w_j + \frac{T_j}{L_j},
\]
where \(Y_j\) denotes national income in \(j\), \(L_j\) is the number of households in \(j\), \(p_{ij}(k)\) refers to the consumer price of variety \(k\) originating from country \(i\) and purchased in country \(j\) (inclusive of any trade costs and tariffs), \(w_j\) is the wage rate of the representative worker in \(j\), and \(T_j\) denotes \(j\)'s total tariff revenue (redistributed in a lump-sum fashion to \(j\)'s representative household).

Within a country, firms are assumed symmetric and have access to the same technology so that all firms in \(i\) charge an identical (mill) price \(p_i\). Consequently, the price the consumer in \(j\) pays for any product from country \(i\) is \(p_{ij}(k) = p_i(1 + a_{ij} + t_{ij})\) for all varieties (or firms) \(k\). At identical consumer prices \(p_{ij}\), all firms in country \(i\) face identical demand from consumers in \(j\), \(c_{ij}\). Then, maximizing utility subject to the income constraint yields a set of demand equations for economy \(j\) with \(L_j\) households:
\[
X_{ij}^D = \left[ \frac{\sigma p_i(1 + a_{ij} + t_{ij})^{-\sigma}(1 + a_{ij})}{\sum_{i=1}^{N} n_i [p_i(1 + a_{ij} + t_{ij})]^{1-\sigma}} \right] Y_j,
\]
where \(X_{ij}^D\) is demand in country \(j\) for each good from country \(i\).

### A.1.2 Firms

All firms in the industry are assumed to produce under the same technology. The output of goods produced by a firm in country \(i\), denoted by \(g_i\), requires \(l_i\) units of labor, as well as an amount \(\phi\) of fixed costs, expressed in terms of the output of the good produced. The production function – similar to that in Krugman (1980) – is given by \(l_i = \phi + g_i\). Firms maximize profits subject to the technology, given the demand schedule derived in Section (A.1.1). In this model, profit maximization leads to a constant markup over marginal production costs and there are zero profits in equilibrium due to free entry and exit. Profit maximization ensures \(p_i = \frac{\sigma}{\sigma - 1} w_i\). Zero profits in equilibrium ensures \(g_i = (\sigma - 1)\phi\).

### A.1.3 Factor Endowment Constraints

We assume that endowments of labor, \(L_i\), are exogenously given and internationally immobile. Assuming full employment, \(L_i = n_i l_i\) or, equivalently, \(n_i = (\phi \sigma)^{-1} L_i\).

The zero profit conditions and the clearing of goods and factor markets lead to balanced multilateral trade for each economy.

### A.2 Calibration of the Model

Our model can be simulated to motivate several potentially “testable” hypotheses regarding relationships between economic characteristics of pairs of countries and the sequencing of either a PTA formation or enlargement, based upon the demand for and supply of nonmembers and members, respectively. For the utility function, we

\[22\] We assume away heterogeneous productivity for firms to limit the complexity of the model. Notice that, for the question of interest here, heterogeneous firms would not change the insights qualitatively, cf., Arkolakis, Costinot, and Rodriguez-Clare (2011).
have one parameter, the elasticity of substitution in consumption between varieties of goods (σ); this elasticity is set equal to 4 as in related earlier studies. For technology, we set the fixed cost term in the production function to unity (ϕ = 1), without loss of generality. Initially, factor endowments of labor are assumed identical across all countries in the symmetric benchmark equilibrium with values of \( L_i = 100 \) for all countries. In this paper, we focus on one industry, leaving analysis of sectoral differences for subsequent research and our sensitivity analysis.

Assumptions regarding international transport costs depart from those in Baier and Bergstrand (2004), although intrainational trade costs are zero (as there). We consider a structure that is consistent with the location of countries on a circle as indicated in Figure S1 (i.e., the “world”). Each country has two “neighbors.” The iceberg transport (or “trade”) cost factor with the two immediate neighbors is \( a \). For instance this applies for country 1’s trade with countries 2 and \( N \). Trade costs with the pair of “second neighbors” (e.g., countries 3 and \( N - 1 \) for country 1) are \( 2a \), and so on. \(^{23}\) The corresponding trade cost “stair-case” is displayed in Figure S2, from the perspective of either country 1 (top) or country 6 (bottom). However, we retain symmetry in the sense that the same stair-case applies to every country on the circle. Even though trade costs with the most remote counterpart on the circle may be large depending on the number of economies in the “world,” trade flows between all pairs of countries will be positive as long as the number of countries is finite (\( N < \infty \)). An obvious advantage of assuming stair-case-type trade costs rather than trade costs which are symmetric for all cross-border trade flows is that an increasing number of adjacent PTA members is associated with higher average trade costs for representative intra-PTA trade relationships. This contributes ultimately to a finite elasticity of “supply of memberships.” We will assume \( a = 0.05 \).

The number of firms, product varieties, labor employments, wage rates, consumption levels, and price levels in each country can be determined uniquely given the parameters of the model (σ, ϕ) and initial transport costs, tariffs, and labor endowments. In summary, \( σ = 4 \), \( ϕ = 1 \), \( a = 0.05 \), \( L = 100 \), leaving only initial tariff rates to be specified.

As in B-B, we assume the existence in each country of a social planner, which sets tariffs initially at 30 percent (\( t = 0.3 \)). \(^{24}\) Based upon initial parameter values, the social planner in each country considers whether its representative consumer’s utility would be better off or worse off from forming a PTA. For a country’s planner to form a new – or join an existing – PTA, the change in utility from doing so must be positive.

\(^{23}\)Hence, \( a_{13} = a_{1,N-1} = 2a \). We have also considered exponentially increasing trade costs; our theoretical results are robust to this alternative specification.

\(^{24}\)The value of 0.3 was originally chosen in B-B following Frankel (1997). As noted in B-B, the ideal approach would be to consider the Nash equilibrium tariffs; the Nash equilibrium tariffs in a post-integration situation are likely to differ from those in the pre-integration situation. It is interesting to note, however, that the calculation of the Nash equilibrium tariffs in the six-country case of B-B yield a pre-integration tariff rate of approximately 0.3 for all countries (assuming symmetry). Moreover, we also note that Ornelas (2005) finds in a political economy framework strong rationales for governments selecting into PTAs based upon “economic” welfare of their countries. Note that tariff rates of 0.3 seem to be high. However, Anderson and van Wincoop (2004) suggest that trade barriers in a broader sense (including non-tariff barriers) are as high as 1.7 for a representative developed country. As many PTAs reduce trade barriers more broadly than just in terms of ad-valorem tariff rates, it seems justifiable to to work with a trade facilitating effect of PTAs which amounts to 0.3.

---

25
A.3 Bilateral Economic Determinants of the Timing of PTA Events

Figure 3 (in the text) summarized three novel “stylized” facts that our numerical GE model can potentially explain. The data in this figure suggest that: (1) PTA events occurred sooner among pairs of countries that were closer; (2) PTA events occurred sooner among pairs of countries that were economically larger; and (3) PTA events occurred sooner among pairs of countries with more similar economic sizes. This section shows how the GE model can motivate theoretically each of these three stylized facts.

A.3.1 Bilateral Distance

As a benchmark, consider initially the case where countries are identical in factor endowments. Figure S3 illustrates the sequencing of PTA agreements for country 1 (chosen arbitrarily); recall that in the presence of a cost to form (join) a PTA, \( y^* \), countries form PTAs one at a time, implying that the sequencing of PTA formations suggests which PTAs form sooner. The first step – the foundation for a PTA – takes place at a random address on the circle, say, between countries 1 and 2. By assumption, PTA membership will happen only if every member country gains in welfare from the PTA. The foundation and subsequent enlargement process of a single PTA is illustrated in Figure S3; for simulation, we will assume henceforth that \( N = 20 \).

In a second step, the two incumbents endogenously decide upon whether to offer country 20 \((N)\) membership or not and, at the same time, country 20 endogenously decides whether to choose to join the PTA or not. Country 20 will become a PTA member only if neither one of the incumbents nor the potential entrant faces a welfare loss from the PTA’s enlargement. In the next step, 1, 2, and 20 decide upon offering membership to country 3. Again, the PTA will enlarge only if every one of the incumbents’ and the potential entrant’s utilities increase, and so on.

While Figure S3 is illustrative, it does not reveal the relative welfare gains of the sequencing of country 1’s PTA. Figure S4 illustrates the relative (marginal and cumulative) welfare gains associated with the initial PTA formation and its enlargement. Figure S4 provides several pieces of information, which we will discuss in turn. Figure S4 displays four lines (including the vertical one labeled “Equilibrium PTA size”). For now, we need only describe the top and bottom lines. The top line shows the (net) welfare effect for a nonmember country of joining an existing PTA or – for country 2 – forming a PTA with country 1; this influences the “demand” for membership in the PTA. The bottom line shows the welfare effect of an enlargement for the “worst-off” member of the existing agreement; this influences the “supply” of membership in the PTA. A loss of welfare for this member vetoes any expansion under the assumption, as most (if not all) agreements reveal, that every member of the existing agreement must accept the potential entrant (even though members may not share a common external tariff).

---

25 In the special case of symmetrically sized economies, country 1 is actually indifferent between considering country 2 or country \(N\). We assume arbitrarily the choice of 2 over \(N\) because we have no other economic characteristic to influence relative welfare gains; hence, we assume an infinitesimal epsilon lower cost for neighbors in a clockwise direction.

26 Due to size symmetry across countries, incumbents 1 and 2 together are indifferent between offering membership to country 20 versus country 3. However, this will change when we consider asymmetrically-endowed countries. Of course, country 1 prefers membership of country 20 over that of country 3 and for country 2 the opposite holds true due to our parametrization of trade costs.
Consider first the effect of trade costs on the formation of a new agreement involving country 1 (recall, we choose country 1 randomly). Clearly, the first point on the top line shows that the model generates endogenously that country 2 – country 1’s nearest neighbor – is the *equilibrium* outcome for a partner in the first round. Country 20 would be the next most likely nonmember to form an agreement with country 1 (and consequently country 2). Country 3 is next, and so on. This implies that the closer physically are two countries the more likely they are to form an agreement *sooner*. This suggests PTA events should occur sooner (and the hazard rate should increase) for pairs of countries that have lower trade costs (other things constant).

For empirical purposes, we use bilateral distances and a dummy variable for a common land border, as in gravity-equation analyses of trade. Hence, PTA events should occur sooner the lower two countries’ bilateral distance and/or if they share a common land border.

### A.3.2 Bilateral Economic Size

In Figure S4, all countries had identical labor stocks and consequently identical economic sizes. We now alter the absolute factor endowments of the 20 countries in our world by shifting labor among countries. In this case, we assume that economically large countries are concentrated in the “North” (with country 1 the largest) and the small countries are concentrated in the “South” (with country 11 the smallest). We attain this by reallocating labor ($L_i$) such that the largest country (country 1) holds 140 percent of the initial endowment of 100 units and the smallest country (country 11) holds 60 percent of the initial endowment. The gradual increase or decrease in absolute factor endowments as we move around the circle is identical, irrespective of with which country we start.

– Supplement Figures S5a and S5b here –

Like Figure S4, Figure S5a considers the effects of country 1 pursuing a PTA – but when country 1 and its nearby countries are large. To understand why larger economies form PTAs first, compare Figures S4 and Figure S5a. Both figures illustrate that country 1 benefits the most from – and is likely to form earliest – a PTA with country 2. However, for the PTA between 1 and 2 compare the relative economic gains (vertical axis) in the two figures. When countries 1 and 2 are large relative to other countries (Figure S5a), the welfare effect of a PTA is larger than that when the countries are all identically sized (Figure S4). Moreover, consider an enlargement to include country 20 in the PTA. The marginal welfare effect from adding country 20 when these countries are large (Figure S5a) is much larger than that when the three countries are equally sized (Figure S4).

We also considered another scenario where country 1 (11) is the smallest (largest) country. The results of country 1 forming a PTA when 1 and its neighbors are small are shown in Figure S5b. Note that the welfare gain for countries 1 and 2 from forming a PTA is even smaller than that in either Figures S5a or S4. Moreover, the marginal gains from enlargement to country 20 are also smaller in Figure S5b than those in either Figures S5a or S4. A PTA between two large partners increases the volume of trade (at the intensive margin) in more varieties than a PTA between two small partners, and reduces trade (at the intensive margin) in fewer varieties from nonmembers than two small partners, improving utility more in large countries relative to small countries, for any values of transport costs. Also, the consequent larger increase in trade among two large economies causes a larger net expansion of demand and hence a larger rise in real income. Small countries face considerable trade diversion when large countries have a PTA; the excess relative supply of factors in the small countries causes an erosion in their terms of trade. This suggests PTA events should occur sooner (and the hazard rate should increase) for pairs of countries with larger GDPs.

27 Once again, we assume arbitrarily that country 2 is “epsilon” closer to country 1 than country 20 is, and so on for 3 vs. 19, etc.
It is important to note that – even for economically small countries – there are net positive benefits from close partners forming a PTA. Also, Figure S5b illustrates that at some point the trade diversion exceeds the trade creation from further enlargement for the “worst-off” (existing) member, ending expansion. This halt of expansion over time in our model is due to net trade diversion, akin to the net trade diversion from simultaneous continental PTAs at low intra- and inter-continental transport costs in the static models of Frankel (1997) and B-B. We address the importance of this limiting-size effect later.

A.3.3 Bilateral Similarity of Economic Sizes

Figure 3 (in the text) shows empirically that PTAs have occurred sooner among pairs of countries that are more similar in economic size. In our 20-country world, the potential effect of increasing two countries’ economic-size similarity – for given absolute economic size of the two countries – is qualitatively the same as increasing their absolute economic sizes. Intuitively, we know from the Krugman (1980) and the Baier and Bergstrand (2004) models that bilateral economic size and similarity have qualitatively similar impacts. In Figures S5a and S5b, the welfare gains from PTA formations and enlargements are enhanced the larger the GDPs of the countries involved because of greater trade creation and less trade diversion. As size disparity increases, the loss of trade in varieties vis-à-vis the rest-of-the-world (ROW) for the larger country rises relative to its increased trade with a smaller PTA partner. Since one of the countries’ welfare declines with size disparity, the time to a PTA formation is delayed. The figure associated with increased bilateral size similarity is qualitatively identical to previous figures, and omitted for brevity. This suggests PTA events should occur sooner (and the hazard rate should increase) for pairs of countries with more similar economic sizes.

A.4 Endogenous Regionalism

In this section, we investigate “endogenous regionalism.” Endogenous regionalism refers here to the role that existing PTAs play in the timing of countries forming new or joining existing agreements. For tractability, following Baldwin (1995) we categorize the influences of existing PTAs on subsequent PTA events into factors influencing the “demand” for membership and the “supply” of membership.

A.4.1 Distance to the Nearest PTA

Consider now the effect of trade costs on the potential enlargement of an existing agreement between countries 1 and 2. Returning for simplicity to the case of symmetric economies and Figure S4, the second point on the top and bottom lines show that the model generates endogenously that country 20 – the next nearest neighbor – is the equilibrium outcome for a partner in the second round (under our assumption of one PTA event at a time). This suggests that a nonmember (country 20) that is bilaterally close to countries that are already members of a PTA is more likely to “demand” membership in that nearby PTA; this is consistent with Baldwin’s domino theory. The bottom line suggests that the welfare effect of the enlargement for the worst-off member of

28This is in contrast to the inferences in Frankel (1997, Figure 8.4) and B-B (2004, Figure 3) where – at low intra- and inter-continental transport costs – the net welfare gains from PTAs were negative. However, continental PTAs were formed simultaneously in both those studies; they never considered only one pair forming an agreement.

29Even though this is beyond the focus of the present study, notice that Figures S4, S5a, and S5b also suggest that PTAs composed of economically large members (such as NAFTA) will have fewer members and, given the adopted assumption about the timing, will expand over a shorter time span than PTAs composed of economically small members (such as the EU).
the existing agreement (here, country 2) is positive, so that membership is “supplied” by countries 1 and 2 to country 20. Thus, a nonmember is more likely to demand membership in and join an existing PTA sooner the closer the nonmember is to that PTA. In the context of our approach, there is a higher utility gain (to both the potential member and the worst-off existing member) from a nonmember joining a close PTA and hence a nonmember will likely join that PTA sooner. This suggests the hazard rate for a country to join another country in an existing PTA decreases as the distance to the nearest PTA increases (distinct from the bilateral distance to its partner).

For empirical purposes, we construct a variable $DIST_{PTA}^{i,j,t-5}$, which measures the distance of a country pair to its nearest PTA five years prior. Consequently, a limiting factor in the enlargement of an existing PTA will be the farther a potential entrant is from the nearest PTA, limiting the demand for membership.

A.4.2 The “Degree” of Regionalism

As just established, countries outside of PTAs face potential trade diversion by not becoming part of an existing nearby agreement. However, we do not observe the potential trade diversion caused by the enlargement of the European Union to cause every country to apply to the EU! It is likely that some country pairs instead form new agreements, such as NAFTA or MERCOSUR. Hence, the formation of new agreements is likely to be an endogenous response to the intensity, or “degree,” of regionalism in the world that creates potential trade diversion and countries’ governments becoming concerned about being left out of “competitive liberalization.”

Our model can simulate the potential effect of a higher overall degree of regionalism on raising the likelihood of a pair of countries forming an agreement sooner. We simulated the model again, returning for simplicity to the case of symmetrically-sized economies, to consider the timing of PTA membership in two agreements. In this simulation, we first introduce exogenously a PTA in the “North” as before (beginning with countries 1 and 2). However, in this case, we now allow endogenously country 11 in the “South” (the country farthest from country 1) to choose joining the North PTA, forming a new PTA, or doing nothing. (As before, we must assume only one PTA formation or enlargement can occur at a time.) The simulation yields the endogenous outcome that country 11 forms a new agreement with country 12, rather than joining countries 1 and 2 in their PTA or doing nothing.  

The next agreements formed endogenously – in sequence – are country 20 joining the North PTA, country 10 joining the South PTA, country 3 joining the North PTA, country 13 joining the South PTA, and so on. In equilibrium, the North PTA is exactly the same 7 countries as before (1, 2, 20, 3, 19, 4, and 18). However, in equilibrium the South PTA also includes 7 countries (11, 12, 10, 13, 9, 14, and 8), owing to the symmetries in economic size and bilateral trade costs.

We can use this simulation to infer the potential effect of a higher “degree of regionalism” in the world on increasing the likelihood of countries forming a PTA sooner. For the North PTA, the associated figure is qualitatively identical to Figure 6; for brevity, we do not provide any new figures. However, the welfare effects associated with potential North PTA entrants and the North PTA’s potentially “worst-off” existing member are quantitatively different. In this case, the welfare effects for entrants and the worst-off existing member are all higher in the case where another PTA can form in the South than when one cannot. In effect, the welfare effects for entrants and members of the North PTA from forming and enlarging their PTA are increased from the presence of regionalism in the South. In the context of the model, this suggests that they are more likely to form a PTA sooner. This suggests the hazard rate for a nonmember to form a new agreement with another

\[30\] The richness of the model given considerable potential asymmetries in economic size, size similarity, and bilateral trade costs can also explain readily whether a nonmember country would endogenously form a PTA with another nonmember or with a member of an existing PTA. We address this below.

\[31\] Recall, in the case of symmetric economies, neighbors in a clockwise direction have epsilon smaller trade costs.
country (or join another country in a PTA) increases as the “degree of regionalism” the pair faces increases.

Capturing empirically the influence of the degree of regionalism on potential entrants’ demand for membership is no easy task. To measure the effect of the degree of regionalism facing a country pair, we constructed for each pair a variable $WPTA_{ij,t-5}$. This variable is a spatially weighted average of all the PTAs countries $i$ and $j$ face in “third” markets (i.e., ROW), five years earlier. We assume the elements of the weighting matrix to be inversely related to the distance (hence, trade costs) between country-pairs $\ell$ and $m$, as in Egger and Larch (2008). For instance, suppose that country-pair $\ell$ consists of economies $i$ and $j$ and country-pair $m$ consists of countries $h$ and $k$. We define the distance between pairs $\ell$ and $m$ as $\text{Distance}_{\ell m} = (\sum_i \sum_\kappa \text{Distance}_{i\kappa}) / 4$ with $\iota = i, j$ and $\kappa = h, k$. All diagonal elements of the weights matrix $W_t$ are set to zero. Consequently, $WPTA_{ij,t-5}$ measures the spatially weighted number of PTAs that country-pair $ij$ faces (in terms of potential trade diversion), with closer PTAs weighted more heavily.\footnote{The inverse-distance-based weighting scheme exhibits elements $\omega_{\ell m}$ that are based on $\omega_{\ell m} = e^{-\text{Distance}_{\ell m}/500}$ if $\text{Distance}_{\ell m} < 2000$. We use a cut-off distance of 2000 kilometers in $W_t$ to avoid problems associated with an excessive memory requirement for matrix elements that are close to zero anyway. We divide the exponent in $\omega_{\ell m}$ to ensure that the decay of interdependence is slow enough (i.e., that the coverage of third countries is large enough).} Hence, nonmembers’ demand for membership in a PTA increases with a higher degree of regionalism (and potential trade diversion) they face.

A.4.3 Number of Members of Nearest PTA

It turns out – for the particular parameterization in Figure S4 (with size-symmetric economies) – that after the PTA enlarges to 7 countries (namely, 1, 2, 20, 3, 19, 4, and 18) – a further enlargement would induce a welfare loss for at least one member, halting expansion of the PTA. Figure S4 (or S5a or S5b) displays four lines. The vertical line demarcates the endogenously-determined equilibrium number of members in the agreement. Recall, the top line shows the (net) welfare effect for a nonmember country of joining an existing PTA (or, for country 2, forming an PTA with country 1); the net trade creation from having a PTA with a larger group of countries increases with the size of the existing PTA, as discussed above. However, while the potential entrant’s welfare gain from joining an existing PTA increases with the PTA’s size as discussed above, the other relevant economic characteristic for a potential entrant’s actually joining the PTA is the welfare gain or loss of the (marginal) “worst-off” existing member – which determines the “supply of membership.” Recall that the bottom line shows the welfare effect of enlargement for the “worst-off” member of the existing agreement. A loss of welfare for this member vetoes any expansion under the assumption, as most (if not all) agreements reveal, that every member of the existing agreement must accept the potential entrant (even though members may not share a common external tariff).

The middle line illustrates the welfare effect of a PTA enlargement for the average member as the number of members expands; this is the average of all existing members and the potential entrant. This is the critical line. In reality, the number of members will be determined by the interaction of demand for and supply of membership. The middle line reveals that – at first – the demand for membership dominates. However, eventually the worst-off member’s utility limits the size of the agreement; note that the welfare effect of the “average” member peaks when the worst-off member’s utility change is negative. Thus, the middle line in Figure S4 suggests the hazard rate for a country to join a country in an existing PTA is a hump-shaped function of the number of members in such PTA, due to the influence of supply of membership by the marginal “worst-off” existing member.

For this, we create a variable that is the actual number of members of the closest existing PTA (actually, the number of members 5 years earlier), $NPTA_{ij,t-5}$, as a determinant of the Time-to-Event. However, since the expected relationship between the hazard rate and the number of members of the closest existing PTA is
quadratic, we also create $SQNPTA_{ij,t-5}$, which is the square of $NPTA_{ij,t-5}$. We expect the hazard rate to be increasing in $NPTA_{ij,t-5}$ and decreasing in $SQNPTA_{ij,t-5}$. This hypothesis reflects the finite elasticity of supply of membership here, in contrast to the infinitely elastic supply of membership in Baldwin (1995).

References for Theoretical Supplement


Circular ordering of countries:
Benchmark case is symmetric endowments with labor (L)

Figure S1
Figure S2

Bilateral trade cost stair-case for Country 1


Bilateral trade cost stair-case for Country 6

Figure S3

Enlargement scenario:

White countries are part of the RTA
(all countries are symmetric)
Figure S4

Equilibrium PTA size:
7 countries

Welfare effect of PTA foundation/enlargement for an entrant

Welfare effect of PTA foundation/enlargement for average member (incumbents and entrant) as compared to no PTA

Welfare effect of PTA foundation/enlargement for worst-off member (incumbent)
Figure S5a

Equilibrium PTA size: 6 countries

Welfare effect of PTA foundation/enlargement for an entrant

Welfare effect of PTA foundation/enlargement for worst-off member (incumbent)

Welfare effect of PTA foundation/enlargement for average member (incumbents and entrant) as compared to no PTA
Figure S5b

Equilibrium PTA size:
7 countries

Welfare effect of PTA foundation/enlargement for
an entrant

Welfare effect of PTA foundation/enlargement for average
member (incumbents and entrant) as compared to no PTA

Welfare effect of PTA foundation/enlargement for worst-off member

Country