Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis*

Peter Egger[†] and Mario Larch[‡]

3rd May 2007

Abstract

Previous empirical work on the determinants of preferential trade agreement (PTA) membership assumes a country's PTA participation to leave other countries' willingness to participate unaffected. More precisely, the presumption is that new PTAs do neither influence the formation of other new PTAs in the future nor do they affect the subsequent enlargement of existing ones. This view is at odds with hypotheses put forward by both political scientists and economists. This paper lays out an empirical analysis to study the role of interdependence in PTA membership in two large data-sets: panel data covering $10,430$ unique country-pairs in eleven five-year intervals between 1950 and 2005, and an even larger set of $15,753$ country-pairs in a cross-section for the year 2005. Applying modern econometric techniques, a PTA membership is found to create an incentive for other countries to form new PTAs or, even more so, to participate in existing ones. This interdependence is stronger among adjacent countries and, more generally, ones with a higher level of 'natural' bilateral trade.

Key words: Preferential trade agreements; Limited dependent variable models; Spatial econometrics

JEL classification: F14; F15; C11; C15; C25

Acknowledgements: We would like to thank Jeff Bergstrand, Daniel Bernhofen, Rick Bond, Rob Feenstra, Eric Fisher, Russell Hillberry, Keith Maskus, Ray Riezman, Andrés Rodríguez-Clare, Jim Tybout, John Whalley, Rainer Winkelmann, and Josef Zweimüller for numerous helpful comments on this and an earlier draft of the manuscript. Furthermore, we have benefitted from discussions with the participants at seminars at the Universities of Munich and Zurich, Pennsylvania State University, the Stockholm School of Economics, the CESifo Venice Summer Institute 2006, the annual meeting of the European Trade Study Group 2006, the Göttingen Workshop in International Economics 2006, the fall Midwest International Economics Meeting at Purdue University 2006, the CESifo Global Economy Area Conference 2007, the Empirical Investigations in International Trade conference in Tokyo 2007, and the annual meeting of the Royal Economic Society 2007. Finally, we would like to acknowledge numerous very helpful comments and suggestions by three anonymous referees and the editor in charge (Daniel Trefler).

[†] *Affiliation:* Ifo Institute for Economic Research, Ludwig-Maximilians University of Munich, CESifo, and Centre for Globalization and Economic Policy, University of Nottingham. *Address:* Ifo Institute for Economic Research, Poschingerstr. 5, 81679 Munich, Germany.

[‡] *Affiliation:* Ifo-Institute and CESifo. *Address:* Poschingerstr. 5, 81679 Munich, Germany.

1 Introduction

If everything in the universe depends on everything in a fundamental way, it might be impossible to get close to a full solution by investigating parts of the problem in isolation.

Hawking, S. and L. Mlodinov (2005), A Briefer History of Time, Bantam Dell, New York, p. 15

The continued integration of the European Union (EU), the formation of the North American Free Trade Agreement (NAFTA), as well as the political discussion about the formation of a preferential trade agreement (PTA) between the Americas have been major sources for the renewed interest in PTAs in the last two decades. With the increasing globalization of the world economy, it seems that there is a raising concern about the global consequences of regionalism (see Krugman, 1991a, Bond and Syropoulos, 1996, Bagwell and Staiger, 1997a,b, 1999, 2005, Bond, Riezman, and Syropoulos, 2004, Riezman, 1999, or Baldwin, 2005, 2006). Another related line of interest focuses on the spread of regionalism and the associated welfare effects of PTA formation and, hence, countries' willingness to be part of a PTA (see Mayer, 1981, Baldwin, 1995, 1997, Frankel, Stein and, Wei, 1995, 1998, Grossman and Helpman, 1995, Yi, 1996, 2000, Maggi and Rodríguez-Clare, 1998, 2007, or Baier and Bergstrand, 2004). The spread of regionalism involves interdependence in the participation decisions among country-pairs which is at the heart of the empirical analysis in this paper.

That the formation of PTAs changes an outsider country's willingness to participate therein is referred to as the *domino theory of regionalism* introduced by and defined in Baldwin (1995, 1997). In Baldwin's work, the source of interdependence in the willingness to participate in a PTA are *political-economy forces.* As Baldwin (1997, p. 877) puts it, idiosyncratic incidents of regionalism trigger a multiplier effect that knocks down "*... bilateral import barriers like a row of dominos.*" Countries desire to participate in an existing PTA since the threat of a loss in the export sector associated with non-participation nourishes lobbying activities to promote membership. The establishment of both NAFTA and the European Single Market created tremendous asymmetries among firms with and without access to these huge markets. Market access is particularly important in a world where firms are mobile across borders and multinationals control goods trade to a large extent. Then, market integration through PTA formation creates an incentive for multinational plant location within the PTA and stimulates a capital influx from abroad (see Baldwin, Forslid, and Haaland, 1996, for simulation-based evidence). In turn, the threat of capital flight into PTAs exerts a pressure on outsiders to join existing PTAs. Baldwin (1995, 1997, 2005, 2006) provides a rich source of examples of PTA memberships that are consistent with the domino theory of regionalism.[1] According to Baldwin's model, bloc size will be finite in equilibrium since it will only pay off to be a PTA member for those countries where the (political) economic gains from participation exceed their resistance to participate. An initial shock associated with the formation of some PTA "*... is amplified by the way in which enlargement makes nonmembership even more costly*" (see Baldwin, 1995, p. 45). If an accession in an existint PTA is not feasible for political reasons, countries might prefer engaging in a new PTA with other outsiders for similar reasons. That some PTAs have even been founded in response to the birth of other PTAs is supported, for instance, by Abbott (1999) who argues in Chapter III-C of his monograph on the North American integration that "*the NAFTA was in part negotiated to counterbalance the growing economic and political influence of the EU. The EU has since pursued negotiations with Mercosur and with Mexico on closer economic relations.*"

This paper lays out an empirical analysis of PTA memberships by explicitly accounting for their interdependence. We use the explanatory variables suggested by Baier and Bergstrand (2004, henceforth referred to as BB) as control variables for PTA membership. Yet, in contrast to previous work we allow the probability of a country-pair's PTA membership to depend on other country-pairs' actions. This demands for recent econometric techniques suitable for the analysis of interdependent limited-dependent variable problems. We provide two pieces of evidence. One is based on panel data for $10,430$ unique country-pairs in eleven five-year intervals between 1950 and 2005. With this data-set we explore the *short-to-medium-run* response in PTA membership probabilities to increased regionalism in the past. The second piece of evidence relies on a cross-sectional data-set for the year 2005. In the latter, we may cover a much larger set of $15,753$ country-pairs for reasons of data availability. However, the questions we may ask in a cross-section differ from the ones with panel data at hand. There, we are interested in interdependent welfare effects of PTA memberships and the associated pattern of membership probabilities in the *long run*. Hence, the associated results based on panel versus cross-sectional data should be seen as complements rather than substitutes.

[1]Whalley (1996) puts forward a different reason for PTA membership, namely the threat of economies of standing alone as PTA outsiders during trade wars. Also Hillberry (2006, p.21) mentions that PTAs may be formed as "*part of a broader foreign policy strategy*".

The empirical findings regarding the economic fundamentals largely support the ones put forward in BB in both data-sets. This is noticeable since BB's results were based on a fairly small cross-sectional data-set of only $1,431$ country-pairs. Beyond that, there is a strong and robust support for interdependence of PTA memberships. Interdependence declines in distance and increases in 'natural' trade among country-pairs.[2] This is consistent with the view that countries wish to participate in PTAs to avoid the welfare loss from trade diversion associated with regionalism among these countries' natural trading partners.

In particular, regionalism in the past creates an incentive to join other countries in an existing PTA as hypothesized by Baldwin (1995, 1997) in his domino theory of regionalism. We also identify a significantly positive incentive to found new PTAs in response to previous regionalism as indicated by Abbot (1999), but this incentive is smaller than the one to join. Also, there is robust evidence that interdependence matters in the very long run. Country-pairs will respond to regionalism in the long run with an even greater probability to participate in an existing or a new PTA than in the short-to-medium run.

The remainder of the paper is organized as follows. In the next section we reconcile hypotheses regarding the PTA-related interdependence of country-pairs. Section 3 lays out the empirical model for interdependent observations with limited dependent variables in cross-sectional as well as panel data-sets. Section 4 summarizes the empirical results, and the last section concludes with a short summary of the most important findings.

2 Hypotheses

Hypothesis 1: The existence of PTAs increases the incentive for a country to join an existing PTA (interdependence and PTA enlargement).

The first hypothesis captures Baldwin's (1995, 1997) domino theory of regionalism. The underlying theory suggests that PTA formation starts a dynamic process of PTA enlargement. Hence, the (random) foundation of PTAs creates an incentive for those countries to participate where consumers in the integrated markets are relatively important. The formation of a PTA diverts trade from outside not only, because insiders trade more with each other, but also because

[2] We define natural trade as the predicted value of bilateral trade without any political trade barriers.

outsider firms relocate their plants into the PTA to serve consumers at lower costs from within the PTA. The enlargement of a PTA increases the incentive for other outsiders to join until all countries participate for whom the gains from joining are at least as large as the costs of becoming a member of the PTA. The willingness to participate in PTAs and the enlargement of PTAs is intermediated by lobbyists.

Hypothesis 2: The existence of PTAs increases the incentive to found new PTAs (interdependence and new PTA foundation).

The second hypothesis roots in the political science literature and talks about interdependence in PTA foundation. According to Abbott (1999) the foundation of PTAs and their enlargement creates an incentive for outsider countries to found new PTAs in response.[3]

Theoretical work in economics has dealt with this and related issues in the context of coalition games of endogenous PTA formation. Yi (1996, 2000) illustrates that the endogenous number of PTAs in equilibrium depends on the structure of the coalition game. For instance, in an unanimous regionalism game, the equilibrium number of customs unions will be two, and they will be asymmetric regarding the number of member countries. In an open regionalism game (where countries can participate in any PTA), the *grand customs union* (i.e., global free trade) is the unique pure-strategy Nash equilibrium outcome. In contrast to customs unions, the *grand free trade area* is not necessarily an equilibrium outcome in the open regionalism game. The formation of a free trade area renders non-member countries better off, while the formation of a customs union makes them worse off.

Hypothesis 3: The interdependence in PTA membership decisions declines in trade costs (bilateral distance).

The incentive to participate in a very distant PTA is small, since the volume of trade that can be generated with the corresponding PTA members is small. Assume that there are some fixed costs associated with (negotiating) PTA membership which is symmetric across PTAs and countries. Then, there is a high likelihood that the fixed costs associated with PTA membership exceed the associated potential gains from trade with a high distance among trading partners. Hence, the willingness of PTA participation declines in (non-political) trade costs to other po-

[3]Note that the foundation of PTA was assumed to be *idiosyncratic* (i.e., random) in Baldwin (1995, 1997). In fact, the model of Baldwin only considers a single, already existing PTA. In contrast, Abbott thinks of the formation of PTAs as *endogenous* events.

tential PTA partners. Specifically, BB show that (i) countries with lower bilateral trade costs and (ii) ones with higher trade costs from the rest of the world, are expected to face high welfare gains from entering a PTA. This hypothesis cannot be inferred separately from Hypotheses 1 and 2. However, it is explicit about trade costs (and trade flows that can be generated by PTA membership) to be an important channel of interdependence. Hypothesis 3 indicates that interdependence should decline with trade costs (or increase with predicted trade) among PTA members.

Hypotheses about economic fundamentals: Countries of similar size, dissimilar relative factor endowments, and low trade costs should expect high welfare gains from entering a PTA. Similarly, countries that face high trade costs with and/or small differences in relative factor endowments to the rest of the world should be inclined towards entering a PTA.

There is an incentive to increase the size of PTAs (up to a certain level) even in the absence of pressure groups or lobbies. For instance, new trade theory models as in Krugman (1991a,b), Frankel, Stein, and Wei (1995, 1998), and BB are based on economic and geographical fundamentals (such as country size, relative factor endowments, and trade costs) and yet they suggest hypotheses about the desirability of PTAs and even their optimal size. Bond and Syropoulos (1996) determine optimal PTA size (and external tariffs) for symmetric and asymmetric trading blocs in such a model without non-tariff trade frictions. For non-prohibitive, positive external tariffs, they show that trading bloc welfare is maximized, if PTAs cover a finite number of member countries. Welfare of non-member countries decreases monotonically with bloc enlargement similar to Baldwin's model (1995, 1997).

BB use a variant of Krugman's (1991b) framework to motivate an empirical model of endogenous selection into PTAs depending on intra- and intercontinental trade costs, country size, and relative factor endowment differences. They confirm Bhagwati's (1993) and Krishna's (2003) view that positive welfare effects of PTAs are more likely for countries that already trade disproportionately with each other. In particular, BB's hypotheses are that (i) countries with a greater similarity in country size and relative factor endowments, (ii) and ones that are relatively dissimilar in these regards from the rest of the world are expected to face high welfare gains from entering a PTA.

Summary: The literature on non-cooperative PTA formation suggests economic fundamentals such as country size, relative factor endowments, and trade costs to be the key determinants of PTA membership (see BB). From an empirical point of view, such models of PTA membership work very well and exhibit a high explanatory power. Yet, these models do not explicitly talk about interdependence in PTA membership, and they assume that both number and size of trading blocs as well as external tariffs are exogenous. Hypotheses about interdependence in PTA membership have mainly surfaced in the literature on cooperative PTA formation.[4] Models of cooperative regionalism derive analytical results for equilibrium size and numbers of trading blocs as well as endogenous external tariffs. Yet, analytical tractability requires the adoption of quite restrictive assumptions: the models build on strong symmetry assumptions across countries (see Yi, 1996); most of them rely on partial equilibrium analysis and the absence of non-tariff trade frictions (see Maggi and Rodríguez-Clare, 1998; Riezman, 1999); some of them assume a fixed number of PTAs (e.g., Baldwin, 1995, 1997); the results crucially hinge upon assumptions about whether and how gains from PTA formation can be redistributed among the members and whether countries can participate in any PTA (open regionalism) or not (unanimous regionalism; see Yi, 2000); finally, – apart from motivating interdependence – they cannot motivate empirical specifications relying on a number of variables that we know are important for PTA formation – e.g., size or factor endowment (dis-)similarities or non-tariff trade frictions.

Empirically, country asymmetries are the rule rather than the exception and non-tariff trade impediments are ubiquitous. Therefore, we use the empirical model of BB as the workhorse specification and extend it to account for interdependent regionalism. Their specification allows for size and relative factor endowment asymmetries, for distance-related trade costs, and it proved to successfully explain the probability of a country-pair's PTA membership in a small data-set of mostly industrialized countries. In terms of explanatory power, BB's empirical model of a country-pair's probability of PTA membership as a function of these fundamentals works very well. However, they did not focus on interdependence even though a country's PTA membership affects other economies' welfare – and, hence, establishes interdependence in PTA memberships – in their model for similar reasons as in Bond and Syropoulos (1996).[5]

We propose an empirical model which allows for interdependence in country-pairs' decisions

[4] Baldwin's (1995) model may be interpreted as a special case of the open regionalism game where only a single trading bloc may arise in equilibrium (see Yi, 1996).

[5] We illustrate how interdependence matters in the model of BB in Supplement A to this paper.

about founding new PTAs (as in Yi, 1996, or Abbott, 1999) and/or joining existing ones (as in Baldwin, 1995, 1997). Our primary goal is to identify the role played by interdependence, *given the economic fundamentals*. In line with many proponents in the literature on PTA formation, we argue that countries will form/join PTAs if the expected gains from trade are big enough. This is likely the case if countries expect the trade volume generated through PTA formation to be large enough.

3 Methodology

3.1 The problem: interdependence of PTA memberships

Empirical applications treat PTA membership as a binary variable with entry one if two countries are members of the same PTA and zero else (see Magee, 2003, and BB). The binary outcome of PTA participation may be viewed as a reflection of the difference in unobservable utility between membership and non-membership scenarios (see McFadden, 1974, and Domencich and McFadden, 1975, for a random utility interpretation of binary choice models that is applicable here). We follow Baldwin (1995, 1997), Bond and Syropoulos (1996), and BB to assume that a country chooses PTA membership only if it gains in welfare and, accordingly, a PTA will be formed only if all members gain. Similarly, accession of a country to an existing PTA will only take place if both the incumbent(s) and the entrant(s) expect to be better off with a PTA enlargement.

Formally, we can introduce $\text{PTA}^\star_{ij} = \min(\Delta U_i, \Delta U_j)$ with ΔU denoting the membership-to-non-membership utility differential of two (potential) members of a PTA. Notice that $\text{PTA}^\star_{ij}$ – and, hence, welfare – is unobserved. What we can observe instead is the binary indicator variable PTA_{ij} which takes the value 1 if two countries are members of the same PTA (indicating $\text{PTA}^\star_{ij} > 0$), and 0 otherwise (indicating $\text{PTA}^\star_{ij} \leq 0$). In vector form (vectors and matrices are in bold), the unobservable utility differential is determined by the following process

$$\mathbf{PTA}^\star \;=\; \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbf{PTA} = \mathbf{1}[\mathbf{PTA}^\star > \mathbf{0}],$$

where $\mathbf{PTA}$ and $\mathbf{PTA}^\star$ are $n \times 1$ vectors, $\mathbf{1}$ is an $n \times 1$ indicator variable, $\mathbf{0}$ is an $n \times 1$ vector of zeros, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of stochastic residuals with n denoting the number of country-pairs. $\mathbf{X}$ is an $n \times k$ matrix of explanatory variables including the constant and $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters.

In principle, one could estimate the model in (1) by a linear probability model, where the binary variable $\mathbf{PTA}$ is regressed on the explanatory variables determining $\mathbf{PTA}^\star$. However, there are well-known problems associated with this approach. Among those, the most important ones are (i) that the error term is then necessarily heteroskedastic which leads to inefficient test statistics and (ii) that the predicted probabilities of PTA membership can be smaller than zero or larger than unity (see Greene, 2003). Existing research on the determinants of PTA membership avoids these problems by deploying non-linear probability models based on the assumption of normally distributed disturbances.

Magee (2003) and BB estimate probit models, where ε_{ij} is identically and independently distributed following the normal distribution $N(0, \sigma_\varepsilon^2)$. However, these models assume that PTA memberships are independent of each other. But the latter is at odds with the theoretical literature on PTA formation discussed earlier. If PTA memberships are interdependent, we cannot obtain consistent estimates of $\boldsymbol{\beta}$ from estimating (1). Accounting empirically for interdependence, the model to be estimated in vector form reads

$$\mathbf{PTA}^\star \;=\; \rho \mathbf{W} \cdot \mathbf{PTA}^\star + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbf{PTA} = \mathbf{1}[\mathbf{PTA}^\star > \mathbf{0}],$$

where ρ is an unknown parameter and $\mathbf{W}$ is an $n \times n$ matrix of known entries that determines the form of the interdependence across country-pairs. Hence, interdependence is captured by a separate explanatory variable. The latter reflects a weighted average of the dependent variable. The corresponding weights either inversely depend on trade costs (distance) as suggested by Hypothesis 2 or they depend positively on natural bilateral trade flows according to economic theory. The weighted average $\mathbf{W} \cdot \mathbf{PTA}^\star$ is referred to as a spatial lag in the literature.

Unfortunately, there are two serious problems in limited dependent variable models with a spatial lag. First, such a data generating process leads to multiple integrals in the likelihood function, rendering simple maximum likelihood estimation infeasible. Second, the error term is likely heteroskedastic leading to inconsistent parameter estimates if this is not accounted for (see

Yatchew and Griliches, 1985; McMillen, 1992; and Davidson and MacKinnon, 1993). Hence, the spatial binary choice model for interdependent PTA memberships cannot be estimated simply by maximum likelihood as binary choice models usually are.

3.2 Cures for PTA membership models with a spatial lag

With interdependence, it is particularly important to distinguish between econometric methods for panel data and ones for cross-sectional data. Cross-sectional data and parameter estimates based on suitable methods for interdependent data are informative about long-run responses to PTA memberships. Panel data and the corresponding methods provide answers about the short-to-medium-run consequences of new PTA memberships in the past on the probability of new subsequent memberships.

Under a set of reasonable assumptions, it turns out that the problem of interdependent country-pairs is easier to tackle with panel data than with cross-sectional data. However, we should not think of panel versus cross-section analysis as substitutes but as complements, here. In the sequel, we discuss upfront solutions for interdependence with cross-sectional data and then turn to the case of panel data.

3.2.1 Cross-sectional data

As indicated before, a model of cross-sectional dependence with an endogenous spatial lag is suited for our problem, since interdependence in the PTA membership-induced welfare effects monotonically declines in trade costs/geographical distance (increases in natural trade) within the empirically relevant range.[6] A spatially lagged dependent variable is the geographical equivalent to a time-lagged dependent variable. There is a large body of research on the estimation of models with a spatial lag of a continuous dependent variable using either maximum likelihood (Anselin, 1988) or generalized method of moments (Kelejian and Prucha, 1999). However, much less research has been undertaken to estimate models with binary dependent variables.

McMillen (1992) is credited with being one of the first to provide an easily tractable solution to the problem. He proposes an EM algorithm which replaces the binary dependent variable

[6] Trade costs are known to increase in geographical distance. See Anderson and van Wincoop (2003) or Baier and Bergstrand (2005), for recent applications of gravity models, where trade costs are associated with distance and common borders.

with the expectation of the underlying continuous latent variable. This variable is then treated as a standard continuous one in the maximum likelihood estimation. The procedure is repeated until convergence. However, several problems arise with McMillen's model (LeSage, 1997, 2000). First, the method prohibits the use of the information matrix approach to determine the precision of the parameter estimates. In particular, the framework rules out estimates of dispersion for the parameter of the spatial lag, which is central to our analysis. Also, the confidence bounds around the other parameters are typically too small. Second, it is not suited for large-scale problems such as ours, covering more than $15,000$ cross-sectional observations. Third, it requires knowledge about the functional form or variables involved in the non-constant variance relationship. Case (1992) derives an alternative estimator to McMillen's. But hers is only applicable to data-sets where the observations can be grouped into regions whose errors are strictly independent of each other (LeSage, 2000).

These problems can be overcome by relying either on geralized method of moments estimates as in Pinkse and Slade (1998) or on the Markov chain Monte Carlo method as proposed by LeSage (1997, 2000). The principal advantages of the Markov chain Monte Carlo approach are its suitability for large-scale problems of spatial dependence such as ours and its flexibility regarding the possible underlying heteroskedasticity of the error term. Formally, the empirical model is a Bayesian heteroskedastic spatial autoregressive probit model as outlined in (1). To allow for heteroskedasticity of the residuals, we assume $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{V})$ with $\mathbf{V} = diag(v_1, v_2, ..., v_n)$. In what follows, we denote the variance of observation i by $\sigma^2 v_i$.

In a Bayesian approach, one applies Bayes' rule to learn about the unknown parameters based on the data. In such a framework, the *posterior density* of the parameters (and hence the parameters that fit the data best) is determined by the product of the likelihood function and the *prior density*. The latter two hinge upon assumptions. In our application, the likelihood function reads

$$L(\rho, \boldsymbol{\beta}, \sigma^2, \mathbf{V}, \mathbf{y}, \mathbf{W}) = \sigma^{-n} \prod_{i=1}^{n} (1 - \rho\mu_i) \prod_{i=1}^{n} v_i^{-1/2} exp \left[- \sum_{i=1}^{n} \frac{\varepsilon_i^2}{2\sigma^2 v_i} \right], \tag{1}$$

where ε_i is the ith element of $(\mathbf{I}_n - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. The determinant $|\mathbf{I}_n - \rho\mathbf{W}|$ is written as $\prod_{i=1}^{n}(1-\rho\mu_i)$, with μ_i denoting the eigenvalues of the matrix $\mathbf{W}$. Priors have to be formed about the set of parameters to be estimated: ρ, $\boldsymbol{\beta}$, σ^2, and $(v_1, v_2, ..., v_n)$. The latter relative variance terms are assumed to be fixed but unknown parameters. However, the Bayesian approach relies on informative priors about the parameters v_i. In particular, an independent $\chi^2(r)/r$ distribution is assumed about the priors on $(v_1, v_2, ..., v_n)$. The χ^2 distribution relies on a single parameter, r. Hence, the n parameters v_i in the model can be estimated by relying on a single parameter r in the estimation.[7] The priors on $\boldsymbol{\beta}$ are assumed to be normally distributed with mean zero and variance 10^{12} (hence, these priors are relatively uninformative), the prior on σ^2 is proportional to $1/\sigma$, and the priors on ρ and r are assumed to be constant. It is assumed that all priors are independent of each other.

Unfortunately, the joint distribution of the parameters is analytically intractable. However, the conditional distributions for the parameters of interest can be set forth (see Albert and Chib, 1993, and Geweke, 1993, for the foundations). LeSage (1997, 2000) derives the conditional distributions for discrete choice models with spatial dependence as ours (see Appendix B for details). Sampling from these conditional distributions then obtains a large set (a chain) of parameter draws. The corresponding estimates of the posterior moments thereof can be shown to converge in the limit to the joint posterior distribution of the parameters (Gelfand and Smith, 1990, LeSage, 2000).

We rely on a chain based on $10,500$ draws. The first and second moments of the chain are computed after skipping 500 burn-ins. Hence, 500 draws are dropped to ensure that there is no systematic information left in the random numbers generation process for the remaining $10,000$ draws. If there is a high autocorrelation in the Monte Carlo chain for each parameter, proper inference on the standard deviation may require dropping further draws from the chain (see Raftery and Lewis, 1992a,b, 1995). The estimates can also be used to compute the first and second moments of the marginal effects to compare the outcome of the spatial probit model of PTA formation to its simple probit counterpart.

[7] Lindley (1971) used this type of prior for cell variances in an analysis of variance problem, and Geweke (1993) in modeling heteroskedasticity and outliers in the context of linear regression. Our runs for the heteroskedastic models rely on $r = 4$.

3.2.2 Panel data

With panel data, we may use a time index t with the process in (1). One may pool the data and the parameter estimates across the available periods and generally use lags of the explanatory variables on the right-hand-side to avoid any bias of the parameter estimates through feedback effects of new PTAs in the future on ones in the past. Furthermore, one may use differences of the variables instead of levels (denoted by a capital $\mathbf{D}$ in front of the respective variable). The change in $\mathbf{PTA}_t$ (i.e., $\mathbf{DPTA}_t$) then indicates *switching into PTAs* rather than just being a member of a PTA at time t. A particular advantage of doing so is that we can implicitly account for a compulsory set of possibly relevant time-invariant variables for PTA membership. In the panel data analysis, we will use five-year differences of $\mathbf{PTA}_t$ between 1950 to 2005.[8] Then, we may adopt the reasonable assumption that new PTA memberships do not affect the probability of new memberships in the past. This avoids the simultaneous determination of $\mathbf{DPTA}_t^\star$ and $\mathbf{W}_t \cdot \mathbf{DPTA}_t^\star$ by using the lagged observable indicator – e.g., $\mathbf{W}_{t-5} \cdot \mathbf{DPTA}_{t-5}$ with a five-year lag – instead of the unobservable contemporaneous variable $\mathbf{W}_t \cdot \mathbf{DPTA}_t^\star$ on the right-hand-side of the model.[9] Formally, the corresponding model reads

$$\mathbf{DPTA}_t^\star = \rho \mathbf{W}_{t-5} \cdot \mathbf{DPTA}_{t-5} + \mathbf{X}_{t-5}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t. \qquad \mathbf{DPTA}_t = \mathbf{1}_t[\mathbf{DPTA}_t^\star > \mathbf{0}_t], \tag{2}$$

Apart from the possible reduction of econometric complexity and immunity against the bias from omitted time-invariant variables, with panel data we may ask about the role of interdependence in *founding new* PTAs versus *joining existing* ones.[10] We then may even ask about the dynamic pattern of interdependence. Also accounting for heteroskedastic disturbances is not a

[8]The advantage of using five-year instead of one-year differences is that we are left with more variation in the left-hand-side variable than in case of taking annual differences.

[9]Without any feedback of new PTA memberships in period t on other country-pairs' memberships in year $t-5$ or even earlier, both $\mathbf{DPTA}_{t-5}^\star$ and $\mathbf{DPTA}_{t-5}$ are strictly exogenous. In contrast to the cross-sectional analysis, where all PTA memberships are simultaneous from a long-run perspective, there are no restrictions on the corresponding interdependence parameters ρ, capturing responses of concurrent new memberships to ones in the past.

[10]For this, we define the latent variable $\mathbf{DPTA}_{found,t}^\star$ and $\mathbf{DPTA}_{enlarge,t}^\star$ to capture the welfare effects of founding new versus joining/enlarging existing PTAs for all country-pairs in year t. Either of them may be determined by the right-hand-side variables in (2). Recall from the introductory section that Abbott argued NAFTA has been created 'in response' to the formation of the European Union. This would be an argument in favor of a positive impact of interdependence on the formation of new PTAs. Instead, Baldwin's domino theory hypothesizes that interdependence matters in particular for the enlargement (or joining) of existing PTAs.

problem under these assumptions anymore.

4 Empirical analysis

4.1 Specification

In the empirical analysis, we rely on a specification that is similar to the one in BB. We use the following variables (the expected signs are in parentheses):

- NATURAL (+) measures the log of the inverse of the great circle distance between two trade partners' capitals.

- DCONT(+) is a dummy variable that takes the value one if two countries are located at the same continent and zero else.[11]

- REMOTE $=$ DCONT $\cdot\, 0.5\{log[\sum_{k\neq j} \text{Distance}_{ik}/(N-1)] + log[\sum_{k\neq i} \text{Distance}_{kj}/(N-1)]\}$ (+) is remoteness of a pair of continental trading partners from the rest of the world.

- total bilateral market size RGDPsum $= log(\text{RGDP}_{it}+\text{RGDP}_{jt})$ (+) with RGDP_{it}, RGDP_{jt} denoting the real GDP of countries i, j in year t.

- RGDPsim $= log\{1 - [\text{RGDP}_{it}/(\text{RGDP}_{it} + \text{RGDP}_{jt})]^2 - [\text{RGDP}_{jt}/(\text{RGDP}_{it} + \text{RGDP}_{jt})]^2\}$ (+) measures the similarity of two countries in terms of their real GDP.[12]

- DKL $= |log(\text{RGDP}_{it}/\text{POP}_{it}) - log(\text{RGDP}_{jt}/\text{POP}_{jt})|$ (+) is the absolute difference in real GDP per capita.[13]

- SQDKL $=$ DKL^2 ($-$) is the square of DKL.

[11]BB use only NATURAL instead of NATURAL and DCONT together. However, our results indicate that both of them should be included.

[12]BB use the absolute value of the difference in log real GDP of two countries instead. Consequently, the expected sign for their parameter is negative.

[13]Already Kaldor (1963) pointed to the high correlation of capital-labor ratios and real GDP per capita. Capital stock data for a large country sample as ours are not available. Even perpetual inventory method based estimates thereof as in BB can not be derived due to missing data on gross fixed capital formation and investment deflators (see Leamer, 1984). If interdependence matters, the enormous loss of observations due to the use of capital stock values can not be justified. With a serious decline in observations, the problem of interdependence could not be consistently accounted for anymore, leading to eventually biased probit estimates.

- DROWKL $= 0.5\{|log(\sum_{kt\neq it} \text{RGDP}_{kt} / \sum_{kt\neq it} \text{POP}_{kt}) - log(\text{RGDP}_{it}/\text{POP}_{it})| + |log(\sum_{kt\neq jt} \text{RGDP}_{kt}/ \sum_{kt\neq it} \text{POP}_{kt}) - log(\text{RGDP}_{jt}/\text{POP}_{jt})|\}$ $(-)$ is the relative factor endowment difference between the rest of the world and a given country-pair.

We set up the database such that every country-pair arises only once in the cross-sectional data-set. With a cross-section of N countries in the sample, there are then $N(N-1)/2$ unique pairs in the sample.[14] Similarly, every country-pair appears only once in an arbitrary year of the panel data-set. Hence, with N_t countries in year t we have $N_t(N_t-1)/2$ pairs in that year. Moreover, with a focus on changes in PTA membership in the panel data analysis, we need to take care of the fact that two countries can not eliminate their tariffs bilaterally if they are already members of a PTA. Hence, we have to exclude the subsequent observations after two countries entered a new membership. This also ensures that zeros in the data (i.e., non-switchers) reflect only country-pairs that do not participate in the same PTA. The cross-sectional data-set covers 178 economies and the panel data-set covers 146 economies (for reasons of availability of coherent GDP data). There are 127 preferential trade agreements. The Supplement to the manuscript provides details on both country and PTA coverage. Sources of the data and descriptive statistics for the dependent and the explanatory variables in use are provided in Table 8 in Appendix A for both the cross-sectional and the panel data-set.

4.2 Spatial weighting

However, our primary interest is on interdependence. Hence, we include the variable $\mathbf{W} \cdot \mathbf{PTA}^{\star}$ in our cross-sectional model and $\mathbf{W}_{t-5} \cdot \mathbf{DPTA}_{t-5}$ in the panel model, respectively, as outlined before. For this, we need to specify the weighting matrix $\mathbf{W}$ (and $\mathbf{W}_t$ for all t). As suggested by new trade theory models, we hypothesize that interdependence should decline in trade costs and, more generally, increase in expected bilateral trade flows. Accordingly, we presume that the elements of the weighting matrix are inversely related to the distance (trade costs) between country-pairs ℓ and m.[15] Suppose that country-pair ℓ consists of economies i and j and country-

[14] Hence, US-Canada and Canada-US are treated as being the same pair.

[15] In one of the sensitivity checks, we employ an alternative weighting scheme which relies on elements that are proportional to average 'natural' bilateral trade flows between pairs ℓ and m. We define 'natural' trade flows as the prediction from a bilateral gravity model as developed by Anderson and van Wincoop (2003). This model captures log-nonlinear effects of geographical trade frictions and country size through CES utility-based price

pair m of countries h and k. We define the distance between pairs ℓ and m as $\text{Distance}_{\ell m} = \left(\sum_\iota \sum_\kappa \text{Distance}_{\iota\kappa}\right)/4$ with $\iota = i, j$ and $\kappa = h, k$.

The inverse-distance-based weighting scheme exhibits elements $\omega_{\ell m}$ that are based on $w_{\ell m} = e^{-\text{Distance}_{\ell m}/500}$ if $\text{Distance}_{\ell m} < 2000$. We use a cut-off distance of 2000 kilometers to avoid problems associated with an excessive memory requirement for matrix elements that are close to zero anyway.[16] We divide the exponent in $w_{\ell m}$ to ensure that the decay of the interdependence is slow enough (i.e., that the coverage of third countries is large enough). We use alternative distance-based weights in the sensitivity analysis. In general, $\mathbf{W}$ is row-normalized for econometric reasons such that $\omega_{\ell m} = w_{\ell m}/\sum_m w_{\ell m}$. Similarly, $\mathbf{W}_t$ is row-normalized for each year t. In the cross-sectional analysis, all PTA memberships occur simultaneously. Accordingly, a prerequisite for proper inference is that the parameter measuring the strength of interdependence meets the restriction $0 \leq |\rho| \leq 1$. This is not the case with the panel data-set, since there is no feedback of new memberships in the future on ones in the past by assumption.

In general, we expect $\rho > 0$ irrespective of whether we consider cross-sectional or panel data. Moreover, with panel data we expect $\rho > 0$ for the joining of existing PTAs, according to Baldwin's (1995, 1997) domino effect of regionalism. The latter would indicate that new PTA memberships in the past create an incentive to join existing PTAs, in particular, if they are not too far away. Similarly, a positive interdependence parameter for new PTA foundations would indicate that new PTA memberships in the past create an incentive to found new PTAs in response, as hypothesized by Abbot (1999). We again hypothesize that the effect should be particularly important if distance is small. However, the latter has not been spelled out by Abbott. In that regard, we are interested in the relative importance of the two mechanisms in the post World War II period.

We have put great effort into ensuring efficiency of the implementation of spatial binary choice models following LeSage (1999a,b). Just to portray the size of the problem: the sheer construction of the matrix $\mathbf{W}$ for the cross-sectional data-set by using a standard loop (running

index terms. We compute the corresponding predictions for a world without political trade impediments such as PTAs. However, we do not base the inference here on the trade-weights-based results, since the corresponding models have somewhat less explanatory power than the inverse-distance-based ones. The Supplement to the manuscript provides more details on this case.

[16]Note that it is impossible to handle (invert, transpose, and even store) a full $15,753 \times 15,753$ as required in our cross-sectional analysis for any modern personal computer. With the chosen cut-off value, 2 percent of the cells of $\mathbf{W}$ are non-zero.

over $15,753 \times 15,753$ country-pairs) in MATLAB takes about 48 hours.[17] The estimation of the spatial probit model with heteroskedasticity-robust standard errors based on the cross-section and $10,500$ Monte Carlo draws takes more than 60 hours.

4.3 New membership events between 1950 and 2005 in the data

Before turning to the model estimation, it seems useful to provide some information about the frequency of new PTA formation and the joining of existing PTAs after World War II. Table 1 reports on the percentages of these events for all 5-year intervals between 1950 and 2005 in the data. This 'time series' information is based on $10,430$ country-pairs whereas we use $15,753$ in the cross-sectional analysis.

−− Table 1 −−

The figures in the table indicate that particularly many country-pairs became members of (new or existing) PTAs between 1985 and 2000.[18] However, most of the new memberships were enlargements of existing PTAs rather than foundations of new ones. Moreover, only part of the new memberships were customs unions or free trade areas (FTAs). Overall, about 14 percent of the country-pairs in the data became members of PTAs between 1950 and 2005. Of those, slightly less than a fifth (about 3 percentage points) were memberships in customs unions or free trade areas. A PTA membership for a randomly drawn country-pair in the panel data-set between 1950 and 2005 is about as 'likely' as one in the larger cross-sectional data-set used above (according to Table 8, about 14 percent of the country-pairs were PTA members in the cross-section). The number of unique pairs changes slightly over time due to the political 'birth' and 'death' of countries and also due to the treatment of European Community (or European Union) members as a single country.[19] Furthermore, we have excluded all pairs at time t with a PTA in place as of $t-5$ or earlier. This is to acknowledge that two countries with a PTA in place in period $t-5$ will not choose to have another PTA among them in period t.[20]

[17] The hardware in use is a Fujitsu Siemens PC with 2 gigabyte RAM and a 3.2 gigahertz processor.

[18] The observation of a positive trend in PTA memberships since the 1950s is consistent with one of the hypotheses in Freund (2000) that falling tariffs render bilateral agreements easier to enforce.

[19] This is to acknowledge that new members can either liberalize their tariffs with all existing member countries or with none of them. This is not a general feature for customs unions or free trade areas, but it needs to be taken into account with the European Union members.

[20] They might integrate others which, of course, we account for. However, we want to exclude the simple relabeling of an existing PTA or its replacement with another one. The latter is not associated with a change

In the sequel, we estimate standard probit models similar to BB and spatial probit models based on the weighting matrix $\mathbf{W}$ ($\mathbf{W}_t$ for the case with panel data). In the spatial models, we account for heteroskedastic disturbances.[21]

4.4 Estimation results

We first present the panel data models for short-to-medium term inference about the determinants of new PTA memberships between 1950 and 2005 in Table 2 (i.e., the foundation or joining thereof) over time. Here, we use the indicator variable $\mathbf{DPTA}_t$ stacked for all years on the left-hand-side of the probit model (we refer to this vector as $\mathbf{DPTA}$). The first three columns in the table refer to probit models for all new PTA memberships, PTA foundations, and PTA enlargements, respectively. The remaining three columns include $\mathbf{W}_{t-5} \cdot \mathbf{DPTA}_{t-5}$ as a determinant of any of these events as indicated in Section 3.2.2.

The simple probit model obtains results that are similar to the ones in BB. This is remarkable since BB focused on PTA membership in a cross-sectional analysis rather than in a panel of events and their data-set was much smaller (covering only $1,431$ country-pairs rather than $10,430$ pairs/$79,649$ observations.) Countries that are closer to each other in geographical terms and that are located at the same continent exhibit a higher probability of a new PTA membership ($\hat{\beta}_{\text{NATURAL}} > 0$, $\hat{\beta}_{\text{DCONT}} > 0$). Country-pairs that are relatively remote from the rest of the world will more likely enter a PTA ($\hat{\beta}_{\text{REMOTE}} > 0$). Also larger and more similarly sized economies tend to become a new PTA member more likely than others ($\hat{\beta}_{\text{RGDPsum}} > 0$, $\hat{\beta}_{\text{RGDPsim}} > 0$). Regarding relative factor endowments, we find that $\hat{\beta}_{\text{DKL}} > 0$ and $\hat{\beta}_{\text{SQDKL}} < 0$. These point estimates are qualitatively in line with those of BB. The marginal effect peaks at a value of DKL where $\hat{\beta}_{\text{DKL}} - 2\hat{\beta}_{\text{SQDKL}} DKL = 0$ at the mean of the data. Hence, larger relative factor endowment differences exert a positive impact on PTA membership only at fairly small values of DKL. Furthermore, we do not find a significantly negative effect of

from the viewpoint of regional tariff liberalization. Therefore, we do not permit country-pairs in the empirical analysis to have another PTA among them, if there is already one in place. Country-pairs tend to engage in a PTA only once in our panel data-set. The design of the analysis there is the following: a country-pair that became a PTA member in some period since 1950 exhibits a unitary entry only once across all years. Hence, the cumulative sum of zero entries across all periods is larger in the panel data-set than in the cross-section while the cumulative sum of unitary entries is comparable (it would be the same if the cross-sections were identical).

[21] See the Supplement to this manuscript for the results of a homoskedastic cross-sectional model.

the difference in relative factor endowments from the rest of the world. Rather, the corresponding point estimate is positive ($\hat{\beta}_{\text{DROWKL}} > 0$) but not significantly different from zero.[22] Also the models for new PTA foundations and new PTA enlargements obtain qualitatively similar results.

Let us now turn to the spatial models that account for interdependence in PTA membership. There, new memberships in year $t - 5$ exert an impact on the probability of other new memberships to take place in year t. Again, we run models using $\mathbf{DPTA}$, $\mathbf{DPTA}_{found}$, and $\mathbf{DPTA}_{enlarge}$ as the dependent variable, respectively. Interestingly, we find that there is only a minor change in the parameters of the economic fundamentals used in the simple probits. But we identify significant, positive interdependence parameters: $\hat{\rho} = 0.922$ (with $\mathbf{DPTA}$), $\hat{\rho} = 1.457$ (with $\mathbf{DPTA}_{found}$), and $\hat{\rho} = 0.992$ (with $\mathbf{DPTA}_{enlarge}$), respectively. This finding supports Hypotheses 1-3 at the same time.[23]

The significance of the spatial interdependence terms leads to higher log-likelihood statistics (and pseudo-R^2s) in the spatial models than in their simple counterparts. The pseudo-R^2s of the simple probit models in the panel data case are smaller than the ones estimated in the much narrower country sample of BB. However, part of the reason for this is the relatively larger number of zeros (i.e., the design-related larger number of PTA non-members) in the panel dataset as compared to the cross-section, according to Table 1. The explanatory power for the panel model using $\mathbf{DPTA}_{found}$ as the dependent variable is higher than that for the one based on $\mathbf{DPTA}$ or $\mathbf{DPTA}_{enlarge}$.

−− Tables 2 and 3 −−

Table 3 summarizes our findings for PTA membership as of 2005 for the larger cross-section of $15,753$ country-pairs. Of course, in a cross-section there is no difference between membership in new versus existing PTAs since we take a long-run perspective where PTA membership of all country-pairs is simultaneously determined out of a situation where no PTAs exist. Therefore, we focus on PTA membership as such, there. Interestingly, the cross-sectional parameter estimates are qualitatively identical to the ones based on the time-series variation.[24] Many of the parameters are even quantitatively similar to the ones for all new PTAs in Table 2 (e.g.,

[22] Notice that BB did not include SQDKL and DROWKL simultaneously.

[23] As indicated before, we may not infer Hypothesis 3 (that interdependence declines in trade costs/increases in natural trade) independent of Hypotheses 1 and 2 (that interdependence matters for joining/founding a PTA).

[24] This is a first indication of the robustness of our findings with respect to omitted time-invariant variables, the consideration of short-run versus long-run effects, and also the composition of the sample. We provide further evidence on the robustness in an extensive sensitivity analysis documented in the Supplement to the manuscript.

$\hat{\beta}_{\text{RGDPsum}}$, $\hat{\beta}_{\text{RGDPsim}}$, and $\hat{\beta}_{\text{SQDKL}}$). The pseudo-$R^2$ values are now higher than those of the comparable models in Table 2, since the control group of PTA non-members is naturally smaller in the cross-section. Again, the parameters of the control variables of the spatial cross-sectional model are quite similar to its simple counterpart.

The significance of the spatial interdependence term also leads to a higher value of the corresponding pseudo-log-likelihood statistics (see LeSage, 1997, 2000; a usual log-likelihood value is not available for the spatial models). The simultaneous and interdependent choice of PTA membership across country-pairs in the cross-sectional model (i.e., the long-run perspective) relies on the Bayesian approach described in Section 3.2.1. The inference is based on a sequence containing every second of the $10,000$ draws (after dropping the burn-ins) according to the Raftery and Lewis (1992a,b, 1995) diagnostic statistics to avoid an excessive autocorrelation in the sequence.[25] Accordingly, the first and second moments of the posterior parameter distributions reported in Table 3 are based on $5,000$ draws only. The diagnostics indicate that there are enough draws and burn-ins for proper inference. The ratio between the total number of draws needed to achieve an accuracy for testing at 5 percent and the ones required under identically and independently distributed draws exhibits a value that is much lower than 5, as required for proper convergence. Also, a set of further convergence diagnostics suggested by Geweke (1992) supports this conclusion but is suppressed in Table 3.

The parameter estimates for both the panel data-set and the cross-section are qualitatively robust to changes in the assumptions about the decay and cut-off values for the inverse-distance-based weighting scheme. They are qualitatively insensitive to choosing 'natural' trade instead of inverse-distance-based weights. Furthermore, the results are not driven by the exclusion of potentially important variables, they also hold in smaller samples than the considered ones (in particular, in the one considered by BB), and they are qualitatively insensitive to the distributional assumptions about the residuals in the Bayesian models. Details about the sensitivity analysis are provided in the Supplement to the manuscript.

[25]This would unnecessarily inflate the standard deviation of the parameters.

4.5 Quantifying the impact of interdependence

Tables 2 and 3 suggest that the *parameter estimates* of the control variables are affected only to a minor extent once we account for interdependence in PTA memberships. However, this does not mean that the *marginal effect* or the *total effect* of interdependence and, hence, its role for a country-pair's *predicted probability* is negligible.

Of course, the effect of interdependence is not identical to the parameter estimate of ρ. However, we may quantify the role of spatial interdependence by means of a comparison of the predicted response probabilities in the spatial versus the simple probit models. We do so for both the panel data models (columns one and four in Table 2) for short-run inference and the cross-sectional models (Table 3) for long-run inference. A first insight in the relevance of modeling interdependence can be gained from looking at the average and the extreme (minimum and maximum) values of the predicted response probabilities. Table 4 does so for the panel data models at the top and for the cross-sectional models at the bottom.

−− Table 4 −−

Obviously, the predictions are on average quite similar between the simple and the spatial models. However, this is not surprising and only means that the models are appropriately centered. However, with binary choice data it matters how well they predict the binary outcomes relative to each other. Obviously, the difference between the spatial versus the simple models is not big for non-memberships (i.e., for the minimum response probability), neither with panel data nor in the cross-section. However, irrespective of using panel versus cross-sectional data, the spatial models do a better job in predicting actual PTA membership. There are several pieces of evidence confirming the latter.

First, the predicted new membership probabilities for those country-pairs that actually became new PTA members are always higher for the spatial model than for the simple one, irrespective of whether we consider all new PTAs, newly formed PTAs, or enlarging ones (the figures are suppressed here but available upon request). The same result is obtained for the cross-sectional models for PTA membership in general.

Second, the predicted maximum response probabilities are always higher for the spatial models than for the non-spatial ones (see the results in the last column of Table 4). The maximum spatial-to-simple model prediction difference for all new PTAs (0.243) is about as big as the maximum predicted probability of a new PTA membership with the simple probit model (0.300) using panel data. The corresponding difference is 0.407 with the cross-sectional models, where the maximum predicted probability in the simple probit model amounts to 0.966. Hence, similar conclusions apply for the ignorance of interdependence of PTA memberships in the short run and the long run: the probability of entering a PTA is downward biased.[26]

Two further comparisons are of interest: predicted new PTA foundations versus PTA enlargements, and predicted membership probabilities in the short run versus the long run. In general, our models predict enlargements of existing PTAs at a higher probability than foundations of new PTAs. One reason for this is that new PTA foundations occur at a fairly low frequency which entails a difficulty for econometric models to predict new events.[27] For a similar reason, it is harder to predict new PTA membership (either foundations or enlargements) at a specific point in time in the short run (with panel data) than PTA membership as such in the long run (with a cross-sectional model). The latter can be seen from the much higher maximum response probabilities in the spatial cross-sectional probit model at the bottom of Table 4 as compared to the spatial probit for all new PTAs at the top of that table.

Note that the results at the top of Table 4 refer to the average unit of observation in the panel data-set between 1950 and 2005. To illustrate the merits of accounting for interdependence with panel data, it is useful to consider the predicted simple probit-based new membership probabilities in each of the 11 covered five-year intervals along with the maximum negative and the maximum positive difference of the spatial model from the simple probit. We summarize the corresponding results in Table 5.

−− Table 5 −−

The results in the table suggest the following conclusions. First, there is a trend in the simple probit model's predictions which is consistent with a larger number of actual memberships in the 1980s and 1990s as compared to the 1950s and 1960s. However, there is not enough variation in these predictions over time to capture the variance in new membership activity across periods

[26] The average absolute difference in the predictions between the spatial and the non-spatial models is 0.053 for the cross-section. The corresponding absolute differences for all new PTAs, newly enlarged PTAs, and newly founded PTAs with panel data are 0.015, 0.015, and 0.001, respectively.

[27] The lower average membership probability for newly founded PTAs as compared to newly enlarged PTAs reported in Table 4 is also associated with a smaller marginal effect of PTA formation in the past on new memberships of the former kind (the marginal effect is 0.0002) as compared to the latter kind (the marginal effect is 0.0258).

reasonably well (see Table 1). The spatial model performs much better also in that regard. This can be seen from the volatility in predicted new membership probabilities across the years, which matches quite well with that one of the actual new memberships. Obviously, the maximum positive deviation (this reflects the prediction for actual new members) is in line with the actual new memberships as of Table 1. However, this pattern is much weaker and the difference is smaller for the maximum negative deviations of the spatial model from its simple counterpart.

For now, we know that the simple probit model-based predictions can be dramatically biased for some country-pairs. Yet we did not identify the corresponding pairs and years these large biases of the simple probit model accrued to. We do so in Tables 6 and 7 for the panel data-set and the cross-section, separately. The distinction between the cross-section versus panel analysis is important here for two reasons. First, we can identify extreme deviations for a country-pair and year when focusing on the short run rather than the long run. However, we had to exclude country-pairs in the panel to avoid problems associated with unbalanced spatial panels due to missing data. Therefore, the cross-sectional data-set is much larger and we can identify sources of a systematic bias for country-pairs there which are not covered in the panel.

−− Tables 6 and 7 −−

In line with our findings from Table 5, the differences in predicted new membership propensities between the spatial and the simple probability models with panel data in Table 6 are largest from 1995 onwards. Overall, the spatial model obviously predicts less action in the short to medium run than the simple one in Europe for the mid 1990s.[28] In contrast, it predicts more new memberships than the simple probit model from the year 2000 onwards in Asia. Overall, the problem of downward-biased simple probit estimates seems more serious than that of upward-biased ones. Again, these estimates should be seen as complements to the cross-sectional ones rather than substitutes since they reflect short-to-medium-term responses rather than long-term ones (recall also that the cross-sectional models outperform the panel-based ones in terms of predicting PTA membership in general).[29]

[28]Note that the discrepancy is not as large for the same countries in later periods. Hence, the predicted probability of the Eastern Enlargement of the European Union based on the spatial model in the period 2000-2005 is higher both than before and than in the simple probit model.

[29]We could also investigate the time-pattern of the long-run response probabilities based on the cross-sectional model parameters. Similar to the within estimates, this would lead to an increasing propensity to participate in a PTA. For instance, it would suggest that Poland's probability of an EU membership rose from about 78 percent in 1960 to about 96 percent in 2000 while that of Romania rose from about 72 percent to 90 percent over the

In the cross-sectional analysis of Table 7, we cover a broader set of country-pairs. In the top panel of the table, the probability of PTA membership predicted by the simple probit model is quite high, and it is much lower in the spatial model. Obviously, the largest negative deviations of the spatial model from the simple one arise for Djibouti-Somalia (-37 percentage points), Oman-Saudi Arabia (-35 percentage points), India-Iran (-34 percentage points), Iran-Saudi Arabia (-33 percentage points), and Israel-Saudi Arabia (-33 percentage points). By and large, these countries are located at or close to the Arabian Peninsula. Since there are only a few PTA members in the neighborhood of these countries, the impact of interdependence on predicted membership is small.

In the bottom panel of Table 7 the opposite holds true. There, the predictions of the simple probit model tend to be low (except for Belize-Nicaragua, where the predicted membership probability is higher than 50 percent) whereas those of the spatial model are much higher. The highest positive deviations of the predicted membership probabilities of the spatial probit from its simple counterpart arise for Aruba-Haiti (41 percentage points), Bahamas-Haiti (37 percentage points), Haiti-Netherlands Antilles (35 percentage points), Belize-Nicaragua (34 percentage points), and Haiti-Nicaragua (31 percentage points). Notice that these countries belong to the Caribbean with numerous PTA members in the neighborhood. Ignoring the latter (i.e., omitting interdependence) leads to downward-biased predicted membership probabilities.

Overall, this illustrates that PTA membership decisions are indeed interdependent. This interdependence declines in distance (increases in natural trade) among country-pairs. Our findings support Richard Baldwin's domino theory of regionalism since the impact of new PTA memberships on subsequent enlargements of existing PTAs is particularly strong. Ignoring interdependence has two consequences: the goodness of fit of nonlinear probability models determining PTA membership is reduced and the predicted PTA membership probabilities are biased. The latter bias can be substantial and it exhibits a geographical pattern.

same time-span.

5 Conclusions

This paper puts forward novel empirical insights about the determinants of preferential trade agreement (PTA) memberships. The focus is on the interdependence of PTA memberships in the world economy. We derive the following three testable hypotheses regarding interdependence: (i) the formation of PTAs and their enlargement generates an incentive for a country-pair to join an existing PTA; (ii) there is a similar incentive to found a PTA in response (e.g., if joining an existing one is politically infeasible; (iii) the interdependence among PTA memberships declines in the distance to (or increases in natural trade with) foreign PTAs since the associated trade diversion is then lower.

These hypotheses are investigated in two large samples of data: a panel data-set covering $10,430$ country-pairs between 1950 and 2005 and a cross-sectional data-set for the year 2005 based on $15,753$ country-pairs. We employ spatial models for discrete choice panel data and a Bayesian spatial discrete choice model for interdependent cross-sectional data. There is significant support for any of the hypotheses which seems to be very robust to the chosen sample, the set of explanatory variables, and various model assumptions. We illustrate that interdependence does not only matter as such, but its ignorance seriously affects the predicted membership probabilities. We provide evidence that the estimated probabilities of PTA membership are biased in absolute value by up to 24 percentage points in the short run (using panel data) and by up to 40 percentage points in the long run (using cross-sectional data).

Appendix

A Data sources

We use information on PTAs that are notified to the World Trade Organization. These data are augmented and corrected by using information from the CIA's World Fact Book and PTA secretariat homepages and they are compiled to obtain a binary dummy variable reflecting PTA memberships for each year between 1950 and 2005. In the panel data-set, we take five-year differences of the binary PTA indicator and use the changes of eleven intervals 1950-1955, 1955-1960, ..., 2000-2005. In the cross-sectional data-set we only use PTA membership of 2005.

For construction of the explanatory variables, we use real GDP figures at constant parent country exchange rates and population. In the cross-sectional analysis we take these data from the World Bank's World Development Indicators. With the panel data-set, this is not possible since the World Development Indicators are only available from 1960 onwards and for many of the covered countries our time series start earlier. Therefore, we rely on the large panel data-set covering real GDP and population collected in Maddison (2003). Bilateral distances are based on the great circle distance between two countries' capitals (own calculations, using coordinates as available from the CIA World Fact Book). The following table summarizes the descriptive statistics of the dependent and independent variables employed in the empirical specification.

−− Table 8 −−

Most importantly, about 14 percent of the $15,753$ country-pairs in the cross-sectional data-set for 2005 (and a similar fraction in the year 2005 of the longitudinal data-set of $10,430$ pairs) are members of the same PTA. About 21 percent of the pairs are intracontinental ones.

B Econometric issues

Following Albert and Chib (1993) and Geweke (1993), LeSage (1997, 2000) derives the conditional posterior distributions of the parameters of interest in the discrete choice model with a spatial lag:

$$\begin{aligned}
p(\boldsymbol{\beta}|\rho,\sigma,\mathbf{V}) &= N[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\mathbf{I}_n-\rho\mathbf{W})\mathbf{y},\sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}],\\
p(\sigma|\boldsymbol{\beta},\rho,\mathbf{V}) &\propto \sigma^{-(n+1)}e^{-\sum_{i=1}^{n}\varepsilon_i^2/(2\sigma^2 v_i)},\\
p(\rho|\boldsymbol{\beta},\sigma,\mathbf{V}) &\propto |\mathbf{I}_n-\rho\mathbf{W}|e^{-(1/2\sigma^2)(\boldsymbol{\varepsilon}'\mathbf{V}^{-1}\boldsymbol{\varepsilon})},\\
p(v_i|\boldsymbol{\beta},\rho,\sigma,\mathbf{V}_{-i}) &\propto \left(\varepsilon_i^2/\sigma^2+r\right)/v_i,
\end{aligned}$$

where $\propto$ indicates that the expression on the left-hand side is proportional up to a constant to the one on the right-hand side, and $\mathbf{V}_{-i}$ indicates all elements except v_i.

The posterior distribution of $\mathbf{PTA}^\star$ conditional on the model parameters takes the form of a truncated normal distribution. The latter is derived by truncating the function $N[\widehat{\mathrm{PTA}}_i^\star,\sum_j\omega_{ij}^2]$ from the right by zero if $\mathrm{PTA}_i=0$ and from the left by zero if $\mathrm{PTA}_i=1$. There, $\widehat{\mathrm{PTA}}_i^\star$ is the predicted value of the ith row of $\text{PTA}_i^\star$, and $\sum_j \omega_{ij}^2$ denotes the variance of the prediction with ω_{ij} denoting the ijth element of $(\mathbf{I}_n - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$. The probability density function of the latent variable $\mathbf{PTA}^\star$ is:

$$f(\text{PTA}_i^\star|\rho, \boldsymbol{\beta}, v_i) \sim \begin{cases} N(\widehat{\text{PTA}}_i^\star, \sum_j \omega_{ij}^2), & \text{truncated at the left by 0 if } \text{PTA}_i = 1, \\ N(\widehat{\text{PTA}}_i^\star, \sum_j \omega_{ij}^2), & \text{truncated at the right by 0 if } \text{PTA}_i = 0. \end{cases}$$

References

Abbott, Frederick M. (1999), *The North American Integration Regime and its Implications for the World Trading System*, Jean Monnet Center at NYU Law School.

Albert, James H. and Siddhartha Chib (1993), Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association* **88**, 669-679.

Anderson, James E. and Eric van Wincoop (2003), Gravity with gravitas: a solution to the border puzzle, *American Economic Review* **93**, 170-192.

Anselin, Luc (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Boston.

Bagwell, Kyle and Robert W. Staiger (1997a), Multilateral tariff cooperation during the formation of free trade areas, *International Economic Review* **38**, 291-319.

Bagwell, Kyle and Robert W. Staiger (1997b), Multilateral tariff cooperation during the formation of customs unions, *Journal of International Economics* **42**, 91-123.

Bagwell, Kyle and Robert W. Staiger (1999), An economic theory of GATT, *American Economic Review* **89**, 215-248.

Bagwell, Kyle and Robert W. Staiger (2005), Multilateral trade negotiations, bilateral opportunism and the rules of GATT/WTO, *Journal of International Economics* **42**, 91-123.

Baier, Scott L. and Jeffrey H. Bergstrand (2004), Economic determinants of free trade agreements, *Journal of International Economics* **64**, 29-63.

Baier, Scott L. and Jeffrey H. Bergstrand (2005), Bonus vetus OLS: a simple OLS approach for addressing the "border puzzle" and other gravity-equation issues, unpublished manuscript, University of Notre Dame.

Baldwin, Richard E. (1995), A domino theory of regionalism, in: R.E. Baldwin, P. Haaparanta, and J. Kiander (eds.) *Expanding Membership in the European Union*, Cambridge: University Press.

Baldwin, Richard E. (1997), The causes of regionalism, *The World Economy* **20**, 865-888.

Baldwin, Richard E. (2005), Stepping stones or building blocs? Regional and multilateral integration, in: J. McKay, M. Armengol and G. Pineau (eds.), *Regional Integration in a Global Framework*, Frankfurt: European Central Bank.

Baldwin, Richard E. (2006), Multilateralising regionalism: spaghetti bowls as building blocs on the path to global free trade, *The World Economy* **29**, 1451-1518.

Baldwin, Richard E., Rikard Forslid, and Jan Haaland (1996), Investment creation and investment diversion: a simulation study of the EU's Single Market Programme, *The World Economy* **19**, 635-659.

Bhagwati, Jagdish (1993), Regionalism and multilateralism: an overview, in: J. de Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration*, Cambridge University Press, Cambridge, UK, 22-51.

Bond, Eric W. and Constantinos Syropoulos (1996), The size of trading blocs: market power and world welfare effects, *Journal of International Economics* **40**, 411-437.

Bond, Eric W., Raymond Riezman, and Constantinos Syropoulos (2004), A strategic and welfare theoretic analysis of free trade areas, *Journal of International Economics* **64**, 1-27.

Case, Anne (1992), Neighborhood influence and technological change, *Regional Science and Urban Economics* **22**, 491-508.

Davidson, Russell and James G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford, University Press.

Domencich, Tom A. and Daniel McFadden (1975), *Urban Travel Demand: A Behavioral Analysis*, Amsterdam: North Holland.

Frankel, Jeffrey A., Ernesto Stein, and Shang-Jin Wei (1995), Trading blocs and the Americas: the natural, the unnatural, and the super-natural, *Journal of Development Economics* **47**, 61-95.

Frankel, Jeffrey A., Ernesto Stein, and Shang-Jin Wei (1998), Continental trading blocs: are they natural or supernatural, in: J.A. Frankel (ed.) *The Regionalization of the World Economy*, Chicago, IL: University Press.

Freund, Caroline (2000), Multilateralism and the endogenous formation of preferential trade agreements, *Journal of International Economics* **52**, 359–376.

Gelfand, Allan E. and Adrian F.M. Smith (1990), Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**, 398-409.

Geweke, John (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in: Bernardo, J.M., J.O. Berger, A.P. Dawid, and A.F.M Smith (eds.), *Bayesian Statistics*, Vol. 4, Oxford: Clarendon Press.

Geweke, John (1993), Bayesian treatment of the independent student t linear model, *Journal of Applied Econometrics* **8**, 19-40.

Greene, William H. (2003), *Econometric Analysis*, 5^{th} ed., Prentice Hall, New Jersey.

Grossman, Gene M. and Elhanan Helpman (1995), Trade wars and trade talks, *Journal of Political Economy* **103**, 675-708.

Hillberry, Russell (2006), Recent empirical studies of preferential trade agreements, University of Melbourne.

Kaldor, Nicholas (1963), Capital accumulation and economic growth, in: Lutz, F.A. and D.C. Hague (eds.), *Proceedings of a Conference Held by the International Economics Association*, London, Macmillan.

Kelejian, Harry H. and Ingmar R. Prucha (1999), A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* **40**, 509-533.

Krishna, Pravin (2003), Are regional trading partners "natural"?, *Journal of Political Economy* **111**, 202-226.

Krugman, Paul R. (1991a), Is bilateralism bad?, in: E. Helpman and A. Razin (eds.) *International Trade and Trade Policy*, MIT Press, Cambridge, MA.

Krugman, Paul R. (1991b), The move toward free trade zones, in: *Policy Implications of Trade and Currency Zones, A Symposium Sponsored by the Federal Reserve Bank of Kansas City*, Jackson Hole, WY.

Leamer, Edward E. (1984), *Sources of International Comparative Advantage: Theory and Evidence*, MIT Press, Boston, MA.

LeSage, James P. (1997), Bayesian estimation of spatial autoregressive models, *International Regional Science Review* **20**, 113-129.

LeSage, James P. (1999a), *Applied Econometrics Using MATLAB*, unpublished manuscript, University of Toledo.

LeSage, James P. (1999b), *The Theory and Practice of Spatial Econometrics*, unpublished manuscript, University of Toledo.

LeSage, James P. (2000), Bayesian estimation of limited dependent variable spatial autoregressive models, *Geographical Analysis* **32**, 19-35.

Lindley, David V. 1971, The estimation of many parameters, in: Godambe, V.P. and D.A. Sprout (eds.), *Foundations of Statistical Science*, Toronto: Holt, Rinehart, and Winston.

Maddison, Angus (2003), *The World Economy: Historical Statistics*, Paris: OECD.

Magee, Christopher S. (2003), Endogenous preferential trade agreements: an empirical analysis, *Contributions to Economic Analysis & Policy* **2**, Article 15.

Maggi, Giovanni and Andrés Rodríguez-Clare (1998), The value of trade agreements in the presence of political pressures, *Journal of Political Economy* **106**, 574-601.

Maggi, Giovanni and Andrés Rodríguez-Clare (2007), A political-economy theory of trade agreements, *American Economic Review*, forthcoming.

Mayer, Wolfgang (1981), Theoretical considerations on negotiated tariff adjustments, *Oxford Economic Papers* **33**, 135-153.

McFadden, Daniel (1974), Conditional logit analysis of qualitative choice behavior, in: Zarembka, P. (ed.) *Frontiers in Econometrics*, New York: Academic Press.

McMillen, Daniel P. (1992), Probit with spatial autocorrelation, *Journal of Regional Science* **32**, 335-348.

Pinkse, Joris and Margaret Slade (1998), Contracting in space: an application of spatial statistics to the discrete choice model, *Journal of Econometrics* **85**, 125-154.

Raftery, Adrian E. and Steven M. Lewis (1992a), How many iterations in the Gibbs sampler?, in: Bernardo, J.M., J.O. Berger, A.P. Dawid, and A.F.M Smith (eds.), *Bayesian Statistics*, Vol. 4, Oxford: Clarendon Press, 763-773.

Raftery, Adrian E. and Steven M. Lewis (1992b), One long run with diagnostics: implementation strategies for Markov chain Monte Carlo, *Statistical Science* **7**, 493-497.

Raftery, Adrian E. and Steven M. Lewis (1995), The number of iterations, convergence diagnostics and generic Metropolis algorithms, in: Gilks, W.R., D.J. Spiegelhalter, and S. Richardson (eds.), *Practical Markov Chain Monte Carlo*, London: Chapman and Hall.

Riezman, Raymond (1999), Can bilateral trade agreements help to induce free trade?, *Canadian Journal of Economics* **32**, 751-766.

Santos Silva, João M.C. and Silvana Tenreyro (2006), The log of gravity, *Review of Economics and Statistics* **88**, 641-658.

Whalley, John (1996), Why do countries seek regional trade agreements?, in: J.A. Frankel (ed.) *The Regionalization of the World Economy*, Chicago University Press, Chicago, IL.

Yatchew, Adonis and Zvi Griliches (1985), Specification errors in probit models, *Review of Economics and Statistics* **67**, 134-139.

Yi, Sang-Seung (1996), Endogenous formation of customs unions under imperfect competition: open regionalism is good, *Journal of International Economics* **41**, 153-177.

Yi, Sang-Seung (2000), Free-trade areas and welfare: an equilibrium analysis, *Review of International Economics* **8**, 336-347.

Table 1: Preferential trade agreement memberships in $10,430$ country-pairs since 1950

5-year period	Country-pairs	Percent of observations with DPTA=1 per period					
		All PTAs			Customs unions and FTAs		
		All	Foundations	Enlargements	All	Foundations	Enlargements
1950-1955	7,748	0.00	0.00	0.00	0.00	0.00	0.00
1955-1960	7,138	0.29	0.29	0.00	0.29	0.29	0.00
1960-1965	7,258	0.39	0.00	0.39	0.39	0.00	0.39
1965-1970	7,258	0.00	0.00	0.00	0.00	0.00	0.00
1970-1975	6,901	1.56	0.00	1.56	0.04	0.00	0.04
1975-1980	6,901	0.13	0.00	0.13	0.00	0.00	0.00
1980-1985	6,784	0.88	0.06	0.83	0.03	0.01	0.01
1985-1990	6,553	3.89	0.00	3.89	0.02	0.00	0.02
1990-1995	8,776	4.19	0.15	4.06	0.76	0.03	0.73
1995-2000	8,776	2.83	0.17	2.65	1.25	0.15	1.11
2000-2005	8,776	0.02	0.00	0.02	0.02	0.00	0.02
1950-2005	82,869	14.18	0.67	13.51	2.80	0.49	2.32

Table 2: Probit results for the probability of new preferential trade agreement memberships (non-spatial and spatial models)

		Non-spatial Probits			Spatial Probits (inverse distance-based W)		
	Theory	All PTAs	Foundations	Enlargements	All PTAs	Foundations	Enlargements
ρ	+				0.922***	1.457***	0.992***
					0.204	0.453	0.202
NATURAL	+	0.160***	0.430***	0.145***	0.150***	0.428***	0.133***
		0.019	0.045	0.019	0.019	0.047	0.019
RGDPsum	+	0.167***	0.201***	0.162***	0.165***	0.198***	0.161***
		0.005	0.024	0.005	0.006	0.025	0.006
RGDPsim	+	0.072***	0.035	0.071***	0.072***	0.038	0.071***
		0.007	0.031	0.007	0.007	0.031	0.007
DKL	+	0.046	0.023	0.051	0.041	0.019	0.046
		0.050	0.296	0.050	0.050	0.301	0.050
SQDKL	-	-0.102***	-0.159	-0.101***	-0.099***	-0.162	-0.099***
		0.019	0.126	0.019	0.019	0.131	0.019
DCONT	+	0.257***	0.272*	0.251***	0.254***	0.277**	0.248***
		0.031	0.140	0.031	0.031	0.141	0.031
REMOTE	+	0.213***	-0.129	0.217***	0.216***	-0.121	0.220***
		0.041	0.120	0.041	0.041	0.121	0.042
DROWKL	-	0.122***	0.079	0.120***	0.118***	0.073	0.116***
		0.028	0.097	0.028	0.028	0.097	0.028
Const		-4.588***	-1.062	-4.719***	-4.688***	-1.149	-4.829***
		0.366	1.055	0.371	0.366	1.063	0.372
Pseudo-R2 (MacFadden)		0.080	0.240	0.074	0.081	0.251	0.076
Log-likelihood		-5552.092	-255.585	-5460.569	-5543.017	-251.922	-5450.085
Log-likelihood for constant only		-6032.092	-336.242	-5896.6036	-6032.092	-336.242	-5896.6036

Notes: There are 79,649 observations and 10,430 country-pairs. The number of observations is smaller here than in Table 1, since we exclude all those pairs from the regression among which a PTA was already effective in period $t-5$. Figures below coefficients are standard errors. All estimated models assume heteroskedastic disturbances. *,**,*** denotes significance at 10%, 5% and 1%, respectively.

Table 3: Probit results for the probability of a preferential trade agreement (non-spatial and spatial models)

Parameters	Theory	Non-spatial	Spatial, $\mathbf{W}^c$
ρ	+		0.805***
			0.035
NATURAL	+	0.517***	0.761***
		0.024	0.030
RGDPsum	+	0.191***	0.128***
		0.009	0.011
RGDPsim	+	0.050***	0.035***
		0.009	0.011
DKL	+	0.203***	0.065
		0.043	0.051
SQDKL	−	-0.111***	-0.062***
		0.011	0.012
DCONT	+	0.516***	0.504***
		0.039	0.050
REMOTE	+	0.518***	0.297***
		0.095	0.108
DROWKL	−	0.001	0.062**
		0.023	0.030
Constant		-6.022***	-0.640
		0.793	0.919
Pseudo-R^2		0.229	
Log-likelihood		-4999.999	
Log-likelihood[a]		-76193.787	-62567.977
Used draws from Markov Chain			5000
Thinning ratio[b]			1
Required number of burn-ins[b]			4
I-statistic[b]			1.507

Notes: There are 15,753 observations (country-pairs). Figures below coefficients are standard errors. [a] LeSage (1999a). - [b] Raftery and Lewis (1992a,b, 1995). - [c] The parameter of the χ^2-distribution of the residuals is set at 4 to account for heteroskedasticity. *,**,*** denotes significance at 10%, 5% and 1%, respectively.

Table 4: Predicted probabilities of a preferential trade agreement (PTA)

Predicted probabilities of a PTA in panel models of Table 2				
Models for all new PTAs	Mean	Std.	Minimum	Maximum
Simple probit	0.015	0.016	0.000	0.300
Spatial probit	0.015	0.016	0.000	0.335
Difference of spatial to simple probit	0.000	0.003	-0.029	0.243
Models for newly founded PTAs				
Simple probit	0.000	0.002	0.000	0.076
Spatial probit	0.000	0.002	0.000	0.124
Difference of spatial to simple probit	0.000	0.001	-0.012	0.119
Models for newly enlarged PTAs				
Simple probit	0.014	0.015	0.000	0.276
Spatial probit	0.014	0.015	0.000	0.349
Difference of spatial to simple probit	0.000	0.004	-0.029	0.262
Predicted probabilities of a PTA in cross-sectional models of Table 3				
Models	Mean	Std.	Minimum	Maximum
Simple probit	0.144	0.166	0.000	0.966
Spatial probit	0.107	0.169	0.000	0.996
Difference of spatial to simple probit	-0.037	0.070	-0.371	0.407

Table 5: Predicted probabilities of new preferential trade agreement membership per period (DPTA)

	Average predicted change in propensity of new PTA membership		
Period	Simple Probit	Maximum positive deviation of spatial from simple probit	Maximum negative deviation of spatial from simple probit
1950-1955	0.0102	0.0003	-0.0126
1955-1960	0.0105	0.0003	-0.0129
1960-1965	0.0118	0.0257	-0.0129
1965-1970	0.0129	0.0395	-0.0134
1970-1975	0.0140	0.0003	-0.0136
1975-1980	0.0151	0.0828	-0.0137
1980-1985	0.0161	0.0194	-0.0142
1985-1990	0.0165	0.0909	-0.0146
1990-1995	0.0176	0.1782	-0.0292
1995-2000	0.0171	0.2435	-0.0076
2000-2005	0.0177	0.1700	-0.0188

Table 6: Extreme differences of spatial to non-spatial model response probabilities based on the panel models of Table 2

Country-pair		Year	Non-spatial probit	Difference spatial-non-spatial
		Largest negative differences		
European Union	Poland	1995	0.116	-0.029
Bulgaria	European Union	1995	0.149	-0.024
European Union	Hungary	1995	0.099	-0.023
Norway	Poland	1995	0.060	-0.023
European Union	Iran	1995	0.041	-0.022
		Largest positive differences		
China	Uzbekistan	2000	0.092	0.243
Nepal	Philippines	1995	0.050	0.178
Mongolia	Thailand	2005	0.045	0.170
Pakistan	Russian Federation	2000	0.040	0.138
Pakistan	Ukraine	2000	0.031	0.136

Table 7: Extreme differences of spatial to non-spatial model response probabilities based on the cross-sectional models of Table 3

Country-pair		Non-spatial probit	Difference spatial-non-spatial
		Largest negative differences	
Djibouti	Somalia	0.541	-0.371
Oman	Saudi Arabia	0.653	-0.350
India	Iran	0.565	-0.340
Iran	Saudi Arabia	0.644	-0.328
Israel	Saudi Arabia	0.646	-0.327
		Largest positive differences	
Aruba	Haiti	0.198	0.407
Bahamas	Haiti	0.260	0.374
Haiti	Netherlands Antilles	0.303	0.352
Belize	Nicaragua	0.535	0.338
Haiti	Nicaragua	0.372	0.306

Table 8: Descriptive statistics for PTAs

Variable	Mean	Std. Dev.	Minimum	Maximum
		For panel data-set		
DPTA	0.014	0.120	0.000	1.000
NATURAL	-8.721	0.740	-9.894	-4.454
RGDPsum	10.886	1.672	4.745	16.561
RGDPsim	-0.587	1.525	-9.878	0.693
DKL	1.168	0.855	0.000	4.657
SQDKL	2.094	2.705	0.000	21.690
DCONT	0.236	0.424	0.000	1.000
REMOTE	8.726	0.326	6.471	9.688
DROWKL	1.004	0.472	0.004	3.010
		For cross-sectional data-set		
PTA	0.144	0.351	0.000	1.000
NATURAL	-8.269	0.778	-9.420	-3.247
RGDPsum	24.681	1.906	18.783	30.182
RGDPsim	-0.733	1.798	-11.000	0.693
DKL	1.847	1.317	0.000	6.100
SQDKL	5.147	6.230	0.000	37.211
DCONT	0.207	0.405	0.000	1.000
REMOTE	8.463	0.165	8.200	9.065
DROWKL	1.605	0.668	0.061	3.884

Notes: There are $15,753$ observations in the cross-sectional data-set and $79,649$ observations ($10,430$ country-pairs) in the panel data-set.