A Bayesian spatial probit estimation of Free Trade Agreement contagion

Dany Jaimovich*
Graduate Institute, Geneva

June 15, 2010

Abstract

This paper analyzes the spatial interdependence of Free Trade Agreements (FTAs) in a cross-section framework using the *Contagion Index* proposed by Baldwin and Jaimovich (2010). A Bayesian heteroskedastic probit model is estimated, where a spatial lag is built based on the *Contagion Index*, finding evidence of interdependence related with a domino-like effect. I compare the results with simple probit estimations and other spatial specifications.

Keywords: Contagion Index, Free Trade Agreements, Spatial probit
JEL Codes: F13

*Graduate Institute of International and Development Studies, Economics Unit. Avenue de la Paix 11A , 1202 Genève, Switzerland. e-mail: dany.jaimovich@graduateinstitute.ch. I thank comments by Patricio Aroca, Richard Baldwin, Peter Egger and Pierre-Louis Vezina.

1 Introduction

The study of interdependence in policy actions is a growing field both in Economics and Political Science, even as the two disciplines do not necessarily share their respective findings[1]. In the arena of international politics, researchers have been motivated by the process of globalization in the last 30 years, that has accelerated the diffusion of ideas and dramatically increased the dependency of local policies on the action of third countries. This interest has been seconded by the development of empirical tools grouped in the discipline of Spatial Econometrics.

In this paper I complement the analysis presented in Baldwin and Jaimovich (2010) related to the interdependence of trade policies, particularly in *contagiousness* of Free Trade Agreements (FTAs). I focus on the cross-sectional case, where the long-term determinants of FTAs are explained, as opposed to the dynamic factors that constitute the core of the empirical study for the model presented in Baldwin and Jaimovich (2010). I devote a special note to this case due to the complexity of the methodology. Indeed, the limited dependent variable and the endogenous spatial lag require the use of non-standard econometric techniques.

The remainder of the paper is organized as follows: In section 2 I describe the empirical strategy and the *Contagion Index* proposed by Baldwin and Jaimovich (2010). I also describe how our weight matrix relates to those of previous studies. In section 3 I present the main results of the Bayesian spatial probit regressions. A last section briefly concludes.

2 Empirical Stretegy

2.1 The Spatial Weights

In the empirical specification, we will follow the spatial econometrics literature and represent the structure of the spatial interdependence as an $N \times N$ zero diagonal weight matrix (W), where each element w_{pq} is a measure of the *distance* between the entities p and q. A natural candidate to measure the space dimension in cross-country studies is the geographical distance between nations. This is the strategy followed by Manger (2006) and Egger

[1]For a recent review from the Political Science side, see Neumayer and Pluemper (2009). A historical review of spatial models in different fields of Economics is provided by Anselin (2009).

and Larch (2008) in previous studies of spatial dependence of FTA formation. Because of the dyadic nature of the data, in which FTAs are signed by pairs of countries, the distance (D) must be defined in terms of the dyad p formed by country-pair i, j and dyad q by country-pair k, l. In Egger and Larch (2008) the element $w_{pq} = w_{k,l}^{i,j} = e^{-\frac{D_{pq}}{500}}$ where $D_{pq} = \frac{(D_{ik}+D_{il}+D_{jk}+D_{jl})}{4}$ and the weight matrix looks like:

$$W_D = \begin{bmatrix} 0 & w_{1,3}^{1,2} & w_{1,4}^{1,2} & \dots & w_{2,3}^{1,2} & w_{2,4}^{1,2} & \dots & w_{n,n-1}^{1,2} \\ w_{1,2}^{1,3} & 0 & w_{1,4}^{1,3} & \dots & w_{2,3}^{1,3} & w_{2,4}^{1,3} & \dots & w_{n,n-1}^{1,3} \\ \vdots & \vdots & 0 & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & 0 & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots & 0 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ w_{1,2}^{n,n-1} & w_{1,3}^{n,n-1} & w_{1,4}^{n,n-1} & \dots & w_{2,3}^{n,n-1} & w_{2,4}^{n,n-1} & \dots & 0 \end{bmatrix}$$

In an alternative definition, Egger and Larch (2008) use the *natural* bilateral exports as defined in Anderson and van Wincoop (2003), with a weight matrix $W_{\widehat{X}}$ where the element $w_{pq} = w_{k,l}^{i,j} = \frac{(\widehat{X_{ik}}+\widehat{X_{il}}+\widehat{X_{jk}}+\widehat{X_{jl}})}{4}$ and $\widehat{X_{ij}} = \widehat{X_{ji}}$ are the estimated exports between countries i and j.

Both specifications use a symmetric matrix where it is sufficient to exploit the information contained in the lower triangular sub-matrix, belonging to the *undirected dyad contagion* type of spatial weights matrices, as classified by Neumayer and Pluemper (2009).

The weight matrix used in Baldwin and Jaimovich (2010) differs to those. They use a theoretically motivated definition of *distance* that relates to the domino theory of regionalism proposed by Baldwin (1993), where the driving force of the contagiousness of FTAs is the trade diversion created by previous agreements. This suggests that two elements must be considered: the number of FTAs signed among a nation's trade partners, and the importance of those markets to the nation's exporters. Considering this, a *Contagion Index* is proposed by Baldwin and Jaimovich (2010), defined as:

$$Contagion_{ij,t} \equiv \sum_{k \in \Omega_{j,t}} \frac{BilateralExports_{ij}}{TotalExports_i} \times \frac{BilateralExports_{kj}}{TotalImports_j} \quad (1)$$

where $\Omega_{j,t}$ is the set of nations with which nation-j has FTAs in year t. From this *Contagion Index*, a weight matrix $W_{Contagion}$ is derived, with the following characteristics:

- The element $w_{pq} = w_{k,l}^{i,j} = \frac{BilateralExports_{ij}}{TotalExports_i} \times \frac{BilateralExports_{kj}}{TotalImports_j} = ExportShare_{ij} \times ImportShare_{jk}$.
- Since $ExportShare_{ij} \neq ExportShare_{ji}$ and $ImportShare_{jk} \neq ImportShare_{kj}$ (except when both are zero), the matrix $W_{Contagion}$ is not symmetric.
- The relevant unit of contagion is a triad: a country-pair's probability to sign an FTA will not be affected by FTAs signed by countries outside the pair, then $w_{k,l}^{i,j} = 0, \forall l \neq j$.

The last point implies that $W_{Contagion}$ will be very sparse and looks like:

$$W_{Contagion} = \begin{bmatrix} 0 & 0 & \dots & w_{3,2}^{1,2} & w_{4,2}^{1,2} & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & w_{2,3}^{1,3} & w_{4,3}^{1,3} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & 0 & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & 0 & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ w_{2,1}^{4,1} & w_{3,1}^{4,1} & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & 0 & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}$$

Following Neumayer and Pluemper (2009) classification, $W_{Contagion}$ belongs to the category of *specific target contagion*.

2.2 Estimation methodology

In a very influential paper, Baier and Bergstrand (2004) estimate the long run economic determinants of FTA formation. Their empirical specification is as follows:

$$FTA^* = P(y^* > 0) = G(\beta_0 + \beta X) \tag{2}$$

where y^* is an unobserved latent variable related to the utility of the FTA that will be expressed as a binary variable $FTA = 1$ if $y^* > 0$ and zero otherwise, $G(.)$ is the standard normal cumulative distribution function and X is a vector of exogenous explanatory variables with associated vector of parameters β that can be estimated with a probit.

The set of regressors in the Baier and Bergstrand (2004) study are:

- Geographical proximity between countries: $NATURAL_{ij} = \dfrac{1}{D_{ij}}$
- Remoteness from the rest of the world: $REMOTE_{ij} = DCONT_{ij} \times 0.5 \times \left[\log(\sum_{k=1,k\neq j}^{N} \dfrac{D_{ik}}{N-1}) + \log(\sum_{k=1,k\neq i}^{N} \dfrac{D_{jk}}{N-1}) \right]$, where $DCONT_{ij}$ is a dummy equal 1 if both countries are in the same continent.
- Market size: $GDPsum_{ij} = \log(GDP_i) + \log(GDP_j)$
- Similarity: $GDPdif_{ij} = |\log(GDP_i) - \log(GDP_j)|$
- Factor endowment: $KL_{ij} = |\log(\frac{K}{L})_i - \log(\frac{K}{L})_j|$, where $\frac{K}{L}$ is the capital-labor ratio in each country.

Following Spatial Econometrics, a *spatial lag* must be added to the basic specification to test the spatial interdependence hypothesis:

$$FTA^* = P(y^* > 0) = G(\beta_0 + \beta X + \rho(W \times FTA^*)) \tag{3}$$

where the spatial lag is given by the expression $(W \times FTA^*)$ in which W is one of the matrices described in section 2.1 (W_D, $W_{\widehat{X}}$ or $W_{Contagion}$) that is used to weight the dependent variable FTA^* according to the different definitions of *distance*. The sign and magnitude of the spatial correlation is given by the parameter ρ.

Unfortunately, (3) can not be estimated using a simple probit, because the spatial lag is endogenous implying the well known associated econometric problems. This is usually the case in Spatial Econometrics, but here the

problem is even more complicated because the dependent variable is categorical and the solutions developed for the continuous case, both using Maximum Likelihood and Generalized Methods of Moments[2], do not apply. Maximum likelihood estimations are not suitable for spatial autoregressive models with limited dependent variable, because the heteroskedasticity in the error term yields inconsistent estimators and the spatial likelihood function involves the evaluation of an n-dimensional integral, which imposes a computational burden difficult to handled in practice. Some proposed solutions for the estimation of spatial models with limited dependent variable involve variable simulation with recursive importance sampling (RIS) and the EM algorithm or GMM estimation.

In this paper I focus on a methodology that is increasingly used by practitioners, based on a Bayesian simulation using a Markov chain Monte Carlo (MCMC), also known as Gibbs sampler, as proposed by LeSage (2000). The advantage of this method is that it simulates the continuous latent variable -basically taking each realization of the latent variable as a parameter to be estimated- and then treats the data as in a regular linear regression. Similarly, it can deal with the problem of heteroskedasticity simulating the elements in the diagonal of the variance-covariance matrix.

The estimation is implemented in a heteroskedastic spatial autoregressive model:

$$\begin{aligned} &FTA^* = P(y^* > 0) = \beta X + \rho(W \times FTA^*) + \epsilon \\ &\epsilon \sim N(0, \sigma^2 V), \\ &V = diag(v_1, v_2, ..., v_n) \end{aligned} \tag{4}$$

In order to implement the MCMC estimation it is necessary to specify the distribution of each parameter conditional on the others. To derive the conditional *posterior distributions*, the Bayesian theorem is used combining the prior distribution times the likelihood function. Following LeSage (2000), the likelihood function is estimated as:

$$L(y, W, \rho, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{(n/2)}} |I_n - \rho W| exp\left\{-\frac{1}{2\sigma^2}(\epsilon'\epsilon)\right\}, \tag{5}$$

and the priors are:

[2]For a recent review of the ML methods can be found in LeSage and Pace (2009). The GMM methods for Spatial Econometrics were introduced in Kelejian and Prucha (1999)

$$\beta \sim N(c, T) \tag{6}$$

$$\sigma \sim \frac{1}{\sigma} \tag{7}$$

$$v_i \sim \frac{\chi^2(r)}{r} \tag{8}$$

The posterior conditional distributions of the parameters β, σ, v_i and ρ, and the sequence of the process are taken from Albert and Chib (1993). In the case of the simulated latent dependent variable the conditional distribution is:

$$f(y_i^*|\rho, \beta, v_i) \sim \begin{cases} N(\tilde{y}_i, \sum_j \omega_{ij}^2), \text{left truncated by 0 if } FTA_i = 1 \\ N(\tilde{y}_i, \sum_j \omega_{ij}^2), \text{right truncated by 0 if } FTA_i = 0 \end{cases} \tag{9}$$

where $\tilde{y}_i$ is the predicted value for the latent variable in the ith row of $\tilde{y} = \beta X$ and ω_{ij} denotes the ijth element of $(I_n - \rho W)^{-1}\epsilon$.

3 Empirical Results

3.1 Data and priors

The data for FTAs come from the recently available Preferential Trade Agreements Database described in Hufbauer and Schott (2009) which is constructed using the date the agreements entered into force, keeping track of all participants. This source is more comprehensive than the WTO database which lists only agreements that have been officially notified. It has 570 agreements recorded from 1948 to 2007; of these, 329 agreements were still force in 2007. In terms of dyads, a cumulative total of 1319 country-pair trade deals are registered up to 2005, but just 1134 (nearly 17% of the total pairs) are still in force or signed for later implementation.

To build the export and import shares for the $W_{Contagion}$ weight matrix I use information from the UN Comtrade database [3]. Only countries with at least 70 partners are included in the sample in order to avoid exaggeration of trade shares related to lack of data. I end up with a sample of 75 countries and the year 2005 is used to maximize data availability, with a coverage

[3]Using the World Integrated Trade Solution of the World Bank. I use information aggregated at the 1-digit Standard International Trade Classification (SITC) level. Whenever a country did not report data on exports, the partner's import image is used.

of 787 country-pair agreements. Appendix A shows the list of countries in both the full and cleaned sample and the number of country-pair agreements signed by each one.

As for the other variables used in the estimation, the GDP data are taken from the World Bank's World Development Indicators (WDI), and the capital-labor ratio are extracted from the same source as Baier and Bergstrand (2004). To build the respective control variables, I use the oldest datum available for each country. The geographical distances are extracted from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) database. In Appendix B the descriptive statistics for the dependent and control variables are shown. After cleaning, the percentage of dyads covered by an FTA rises to 22.4%, reflecting the fact that the countries dropped from the sample have less agreements than the average. Table B.1 shows that in the cleaned sample the means of $NATURAL_{ij}$, $REMOTE_{ij}$ and $GDPsum_{ij}$ do not change considerably with respect to the full sample, but $GDPdif_{ij}$ and KL_{ij} decrease by around 29%, as countries not included in the final sample are smaller (in GDP terms) and less capital-intensive than the average. The last section of Table B.1 presents the descrptive statistics for the cleaned sample when missing values are filled (as zeros), a step required for the implementation of the Gibbs sampler procedure, in order to have a complete weight matrix. Control variables do not change significantly.

In terms of the assumptions of the Gibbs sampler procedure, I stick to the parameters proposed by LeSage (1999): $\beta \sim N(0, 10^{12})$ and $r = 4$ for the distribution of v_i. Since the priors are very uninformative, it is very likely that the first draws are going to be far from the final value of the parameters and must be eliminated. I will implement a total of 10,000 draws, from where the first 500 will taken as burn-in.

3.2 Results

I first analyze the results using simple probit estimations. In the first column of Table 1 the specification of Equation (2) is estimated with the full sample of dyads in the data for which controls are available. The results of Baier and Bergstrand (2004) are mostly replicated here. All the variables are statistically significant at the 1% level with the probability of an FTA increasing with the proximity of the countries ($NATURAL_{ij}$), their relative distance from the rest of the world ($REMOTE_{ij}$) and the size of their markets ($GDPsum_{ij}$) and decreasing with the differences in GDP ($GDPdif_{ij}$). In the case of the difference in factor endowments, I find a result ooposite

to Baier and Bergstrand (2004), because KL_{ij} is negative. When the data is cleaned, the results hold for the first three regressors, as can be seen in column 2, but $GDPdif_{ij}$ and KL_{ij} are not significant anymore. This can be explained by the change in sample. With the filled data I obtain similar results.

The naïve specification of Equation (3) is estimated in the last three columns of Table 1 for each of the different endogenous spatial lags. When W_D is used, ρ is significant and positive, with value 1.718. It is interesting to note that the value of $NATURAL_{ij}$ is reduced by about a half, indicating that in this specification some of its explanatory power is captured by the spatial lag based on geographical distances. When $W_{\widehat{X}}$ is used, the value for ρ increases to 6.113, and the coefficient for $NATURAL_{ij}$ is also reduced, but to a lesser extent. In the last column the endogenous lag is built using $W_{Contagion}$, and a value of 1.878 is obtained while the controls keep close values to the non-spatial specification.

In order to tackle the endogeneity issue, Table 2 shows the results of the Bayesian procedure described above for the three different weight matrices, for which the same sample of 75 countries is used yielding a 5550×5550 matrix ($N(N-1) \times N(N-1)$). Following a standard practice to give some interpretability to ρ in terms of a correlation, these matrices are row normalized to sum up to unity.

The first column shows the results for the estimation of Equation(4) when the space dimension based on geographical distance is used and W_D is the matrix of spatial weights, replicating the estimation of Egger and Larch (2008). In order to alleviate the computation burden of dense weight matrices, the distance is considered zero for $D_{pq} > 2000$ kilometers. Even though I have a smaller sample and a different FTA database from Egger and Larch (2008), the value obtained for the spatial parameter ρ is very close to theirs (0.741 compared with 0.749) but with slightly bigger standard errors. This indicates that ρ in the naïve specification was upward biased. In term of the control variables, the results of the simple probit with the cleaned data hold here and in the rest of the panels of Table 2.

The second column of Table 2 presents the results when the $W_{\widehat{X}}$ matrix, based on estimated bilateral exports, is used. The values for $\widehat{X}_{ij}$ were obtained as the country-pair average of the first observation in the sample of predicted values from a simple gravity equation estimated with fixed effects

and the log GDP of nations in the dyad as regressors[4] A cut-off value of 0.02 was imposed for the averaged export share not to be null. The obtained value for ρ is 0.916 and it is statistically significant.

The last column of Table 2 reports the Bayesian spatial probit estimation of the *Contagion Index* in FTA, where the matrix $W_{Contagion}$ is used to create a spatial lag defined by the theoretical explanation of interdependence proposed in Baldwin and Jaimovich (2010), where the threat of trade diversion generated by agreements signed by third countries is the driving force of the spatial contagion. As shown in section 2.1, the matrix $W_{Contagion}$ is very sparse, and no assumptions about the relevant range of *distance* are necessary. To estimate the export and import shares of $W_{Contagion}$ we use the same procedure described above for $\widehat{X}_{ij}$. A positive and significant value of 0.306 is obtained for ρ. This result is interesting because the parsimonious matrix $W_{Contagion}$, extremely light compared with W_D and $W_{\widehat{X}}$, can give evidence of spatial interdependence with a clear theoretical interpretation.

4 Conclusions

This paper complements the analysis presented in Baldwin and Jaimovich (2010) related to the interdependence of trade policies, particularly the *contagiousness* of FTAs, expanding the analysis to the cross-sectional case, where the long-term determinants of FTAs are examined. Using a Bayesian heteroskedastic probit model that includes a spatial lag built based on their *Contagion Index*, I find support for the hypothesis of spatial interdependence following a domino-like effect.

References

Albert, J., and S. Chib (1993): "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88(422), 669–679.

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

[4]Baier and Bergstrand (2009) use Montecarlo simulations to show that the bias of using fixed-effects instead of the Anderson and van Wincoop (2003) procedure is very small, and that the former can be even more unbiased in the case of asymmetric trade costs.

Anselin, L. (2009): "Thirty Years of Spatial Econometrics," Working Paper 2009-02, GeoDa Center for Geospatial Analysis and Computation School of Geographical Sciences and Urban Planning Arizona State University.

Baier, S., and J. Bergstrand (2004): "Economic determinants of free trade agreements," *Journal of International Economics*, 64(1), 29–63.

——— (2009): "*Bonus vetus* OLS: A simple method for approximating international trade-cost effects using the gravity equation," *Journal of International Economics*, 77(1), 77–85.

Baldwin, R. (1993): "A Domino Theory of Regionalism," NBER Working Paper Series, Vol. w4465. Eventually published in Baldwin, Haaparanta and Kiander (eds.), *Expanding membership of the European Union*, Cambridge University Press, 1995.

Baldwin, R., and D. Jaimovich (2010): "Are Free Trade Agreements Contagious?," NBER Working Paper Series, Vol. w16084.

Egger, P., and M. Larch (2008): "Interdependent preferential trade agreement memberships: An empirical analysis," *Journal of International Economics*, 76(2), 384–399.

Hufbauer, G., and J. Schott (2009): "Fitting Asia-Pacific agreements into the WTO system," in *Multilateralising Regionalism: Challenges for the global trading system*, chap. 12, pp. 554–635. Cambridge University Press.

Kelejian, H., and I. Prucha (1999): "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40(2), 509–533.

LeSage, J. (1999): "The Theory and Practice of Spatial Econometrics," Department of Economics. University of Toledo. Available in www.spatial-econometrics.com.

——— (2000): "Bayesian Estimation of Limited Dependent-Variable Spatial Autoregressive Models," *Geographical Analysis*, 32(1), 19–35.

LeSage, J., and R. K. Pace (2009): *Introduction to spatial econometrics*. Taylor & Francis, Inc.

Manger, M. (2006): "The Political Economy of Discrimination: Modelling the Spread of Preferential Trade Agreements," Montreal, Department of Political Science, McGill University.

Neumayer, E., and T. Pluemper (2009): "Spatial Effects in Dyadic Data," *International Organization*, 64(1), 145–166.

Table 1: Simple probit estimations

	B-B Full Sample	B-B Cleaned	B-B Filled	W_D	$W_{\widehat{X}}$	$W_{Contagion}$
$NATURAL_{ij}$	0.979***	0.946***	0.832***	0.481***	0.633***	0.862***
	(0.0337)	(0.0412)	(0.0358)	(0.0390)	(0.0409)	(0.0290)
$REMOTE_{ij}$	0.0451***	0.0512***	0.0677***	0.0371***	0.0984***	0.0770***
	(0.0063)	(0.0079)	(0.0076)	(0.0078)	(0.0083)	(0.0070)
$GDPsum_{ij}$	0.119***	0.0863***	0.157***	0.111***	0.108***	0.160***
	(0.0074)	(0.0099)	(0.0061)	(0.0067)	(0.0073)	(0.0050)
$GDPdif_{ij}$	-0.074***	0.0268	0.00502	0.00389	-0.071***	0.00166
	(0.0128)	(0.0169)	(0.0164)	(0.0175)	(0.0180)	(0.0169)
KL_{ij}	-0.071***	-0.0362	-0.0147	-0.0611**	-0.00145	-0.0102
	(0.0191)	(0.0294)	(0.0296)	(0.0299)	(0.0307)	(0.0311)
ρ				1.718***	6.113***	1.878***
				(0.1160)	(0.4695)	(0.1052)
Observations	9274	5342	5550	5550	5550	5550
Pseudo R^2	0.397	0.384	0.383	0.421	0.413	0.413
Robust standard errors in parentheses						
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$						

Table 2: Bayesian spatial probit estimations using different weight matrices

	W_D		$W_{\widehat{X}}$		$W_{Contagion}$	
Variable	Coef.	S.Dev.	Coef.	S.Dev.	Coef.	S.Dev.
$NATURAL_{ij}$	0.733	0.034	0.863	0.035	0.863	0.027
$REMOTE_{ij}$	0.053	0.007	0.065	0.008	0.071	0.006
$GDPsum_{ij}$	0.1	0.006	0.128	0.006	0.136	0.004
$GDPdif_{ij}$	0.009	0.017	0.035	0.018	0.009	0.018
KL_{ij}	-0.037	0.031	-0.026	0.031	-0.030	0.032
ρ	0.741	0.06	0.916	0.051	0.306	0.021
Number of Draws	10000					
Burned Draws	500					
Countries	75					
Min. Export Partners	70					

Appendices

A Countries and FTAs

Table A.1: Countries in the cleaned sample

Country	FTAs	Country	FTAs	Country	FTAs
Algeria	26	Ghana	6	Pakistan	13
Argentina	10	Greece	8	Panama	7
Australia	7	Guatemala	8	Peru	10
Austria	34	Hong Kong	1	Philippines	6
Bangladesh	6	Hungary	24	Poland	24
Belarus	11	India	9	Portugal	34
Belgium	36	Indonesia	6	Russia	11
Bolivia	10	Iran	7	Singapore	14
Brazil	11	Ireland	36	Slovak Republic	24
Bulgaria	23	Israel	22	South Africa	17
Canada	7	Italy	36	Spain	34
Chile	30	Japan	2	Sri Lanka	6
China	12	Jordan	24	Sweden	34
Colombia	19	Kazakhstan	13	Switzerland	24
Costa Rica	12	Kenya	6	Tanzania	6
Cyprus	21	Korea,South	5	Thailand	9
Czech Republic	23	Latvia	22	Trinidad and Tobago	6
Denmark	36	Lithuania	22	Tunisia	22
Ecuador	10	Malaysia	6	Turkey	27
Egypt	32	Mexico	35	Ukraine	11
El Salvador	9	Morocco	21	United Kingdom	36
Estonia	22	Nepal	4	United States	7
Finland	32	Netherlands	36	Uruguay	10
France	36	New Zealand	6	Venezuela	17
Germany	36	Norway	24	Vietnam	0

Table A.2: Countries just in the full sample

Country	FTAs	Country	FTAs
Albania	3	Madagascar	2
Armenia	9	Malawi	7
Azerbaijan	13	Mali	4
Burkina Faso	4	Moldva	13
Cambodia	6	Mozambique	4
Central African Rep.	1	Nicaragua	8
Chad	1	Niger	4
Fiji	3	Papua N.Guinea	3
Gabon	0	Paraguay	10
Gambia	2	Sierra Leone	3
Georgia	11	Syria	14
Guinea	3	Tajikistan	11
Guinea-Bissau	0	Togo	6
Guyana	4	Turkmenistan	13
Haiti	2	Uganda	5
Honduras	8	Uzbekistan	13
Jamaica	6	Yemen, Republic Of	6
Kyrqyz Republic	13	Zambia	7

B Descriptive statistics

Table B.1: Descriptive statistics for the different samples

	Full Sample (9274 Obs.)		Cleaned Sample (5374 Obs.)		Filled Sample (5500 Obs.)	
Variable	Mean	S. Dev.	Mean	S. Dev.	Mean	S. Dev.
FTA_{ij}	0.168	0.373	0.224	0.417	0.215	0.411
$NATURAL_{ij}$	-8.156	0.833	-8.171	0.884	-7.865	1.778
$REMOTE_{ij}$	2.129	3.635	2.106	3.631	2.027	3.585
$GDPsum_{ij}$	46.765	2.477	47.831	2.269	46.039	9.354
$GDPdif_{ij}$	2.415	1.742	1.874	1.399	1.804	1.418
KL_{ij}	1.427	1.210	1.108	0.821	1.066	0.832