Supplemental Appendix For

The Stabilizing Effects of International Politics on Bilateral Trade Flows

In this appendix, we first present summary statistics for our main dependent variable, *Export Volatility*, and our independent variables: *Alliance* and $Diplomacy_{exp}$. We then describe the Bayesian Model Averaging (BMA) approach used in our analysis.

Summary Statistics

In this section we report summary statistics for the key variables used in our analysis. Among those dyads that held alliances or diplomatic ties within our 1950-1999 sample, Table A.1 reports the average duration of alliances and diplomatic missions. As one can see, the average duration of alliances and diplomatic missions in our sample is 17-21 years, suggesting that these institutions are indeed relatively long lasting, while these measures' respective standard deviations of approximately 14 and 16 years suggest that we nevertheless have substantial variation in institutional duration among the allied and diplomatically-tied countries in our 1950-1999 sample. Table A.2 then reports the overall means, standard deviations, maximum and minimums, and skewness for our independent and dependent variables. All variables have moderate to high levels of variation and relatively low levels of skewness. Furthermore, these statistics indicate that a relatively larger share of the dyad-year observations in our 1950-1999 sample (41%) held diplomatic ties, than alliance ties (12%).

Table A.1: Durations of Dyadic Alliances and Diplomatic Missions within 1950-1999 Sample

	Mean	Standard Deviation	Minimum	Maximum	Percentage Right Censored at 1999
Alliance	16.83	14.33	1	49	78%
Diplomacy	20.71	15.93	1	49	63%

Note: summary statistics are in years.

Table A.2: Summary Statistics for 25-year-Period Sample

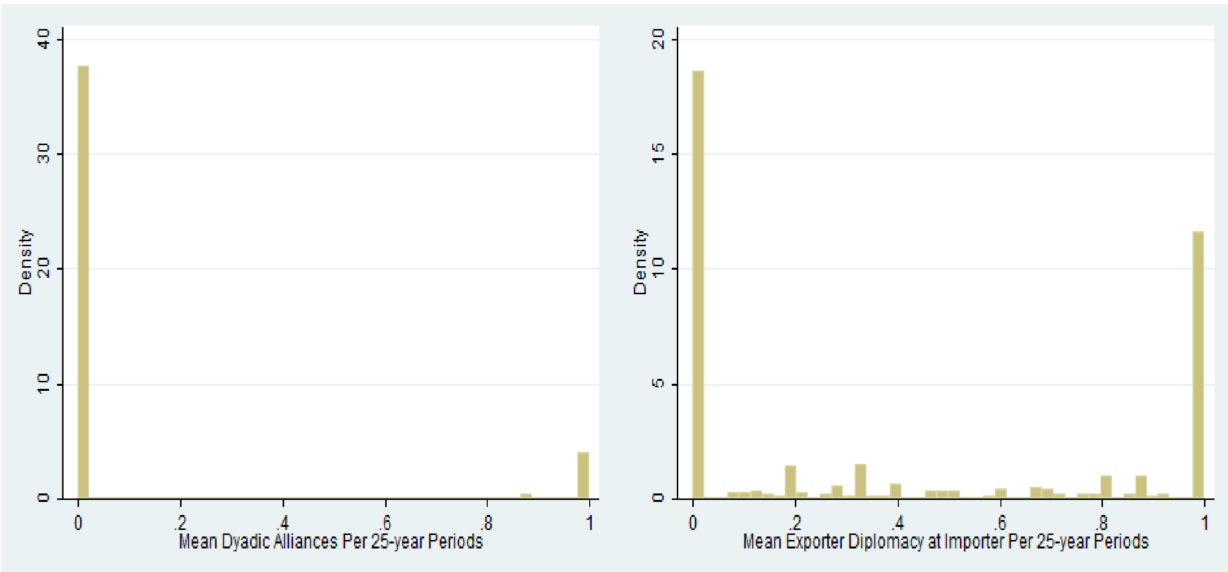
	Mean	Standard Deviation	Minimum	Maximum	Skewness
Export Volatility	0.18	0.35	0.00	9.94	12.73
Alliance	0.12	0.32	0.00	1.00	2.26
$Diplomacy_{exp}$	0.41	0.43	0.00	1.00	0.36

We next present several figures that reinforce the conclusions drawn above. Figure A.1 presents density histograms for our two independent variables: *Alliance* (Figure A.1a) and $Diplomacy_{exp}$ (Figure A.1b) for our 25-year aggregations. Again reinforcing the notion of high variation in our independent variables, here we can note that while a large number of observations in our sample either held (= 1), or did not hold (= 0), alliances or diplomatic ties for our given 25-year periods, a significant number of observations held these political ties for only a portion of a given 25-year time period (particularly for diplomatic ties).

Figure A.1: Distributions of Independent Variables, 25-year Period Sample

(a) Histogram: *Alliances: 1950-1999*

(b) Histogram: *Diplomacy: 1950-1999*

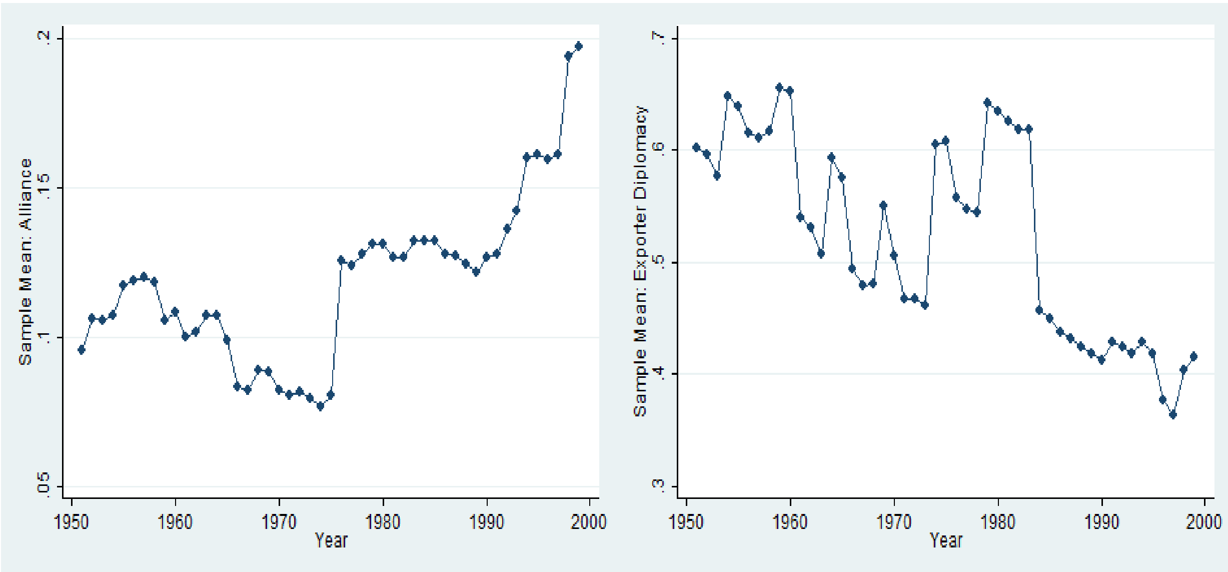


The time series plots in Figure A.2 report the yearly changes in global levels of dyadic alliances and dyadic diplomatic missions for our sample. To create these figures, we calculated the average level of dyadic *Alliances* and $Diplomatic\ Missions_{exp}$ for each year in our sample.

Accordingly, a value of one on either figure would denote a year in which all dyads in the world held alliances or diplomatic ties and a value of zero would denote a year in which no dyads held such ties. Both series exhibit a substantial amount of variation. Beginning first with Figure A.2a we can note that we find significant time varying cross-sectional variation for *Alliance* in our sample, with a strong upward trend that begins with *Alliances* existing within roughly 10% of our dyads in 1950 and ends with an average *Alliance* level of 20% of all global dyads in 1999. $Diplomacy_{exp}$, on the other hand, exhibits a downward trending series for the 1950-1999 time period, declining from diplomatic ties being held within roughly 60% of all dyads in the 1950's to roughly 40% in the 1990's.

Figure A.2: Time Series Variation in Independent Variables, Yearly-Averages

(a) Histogram: *Alliances: 1950-1999* (b) Histogram: *Diplomacy: 1950-1999*



Bayesian Model Averaging

In the following section, we present an overview of the version of Bayesian model averaging (BMA) approach that we employed in our empirical analysis. In short summary, our approach closely follows that of Bartels (1997).

Traditional tests of model "robustness" involve estimating a variety of alternative model specifications contingent on the guidance of expert opinion, existing literature, or prevailing statistical wisdom. Generally, if one's results are retained in the face of alternative tests and controls, a researcher concludes that the model is "robust" to alternative specifications (Bartels, 1997; Montgomery and Nyhan, 2010, 246). Although this approach may constitute the norm of political science research and address some of the uncertainty between model specifications, the practice itself vastly understates the degree of uncertainty about the effects of the variables of interest (Montgomery and Nyhan, 2010, 246). By simply discriminating among a set possible models, estimating a few alternatives, but not allowing these different specifications to meaningfully contribute to the final set of reported results, we present an incomplete picture of trade volatility relationship.

BMA provides a solution to this statistical dilemma by better addressing the uncertainty inherent in statistical modeling. In this manner, BMA allows the researcher to employ a stronger metric of robustness of one's results by conducting a battery of alternative specifications, and weighting these alternative models according to their posterior distributions (or probabilities) of how well the models predict the quantity of interest (Montgomery and Nyhan, 2010, 247). In this sense, BMA "produces both parameter estimates and standard errors that more honestly reflect the observed variation of results across a range of possible models" (Bartels, 1997, 643) by allowing each model to contribute to the final set of reported results. In order to carry out this technique, a researcher must first decide how many models are necessary to estimate the quantity of interest. After identifying the relevant models and estimating their effects on the quantity of interest, scholars must then decide on how to average these models by choosing either to exclude model specifications that are considered theoretically and/or inappropriate, or use the entirety of the model space.[1] Second, a Bayes factor which is a criterion used to evaluate the prior veracity of their model, must be selected. For our purposes, we use the Bayesian Information Criteria (BIC), identified by (Raftery, 1995) as an approximate Bayes factor for model selection because it favors parsimonious models and penalizes for overfitting when more parameters are added to the estimation, which is concern given the subset of models we employ (see Schwarz, 1978; Kass and Wasserman, 1995). Thus, in accordance with Bartels

[1]While the former approach relies on a smaller subset of (all possible) models, extant trade volatility research has already reached a large degree of consensus on this subset. Using all 2^p possible model combinations, as Montgomery and Nyhan (2010, 247) suggest, when we believe that the theoretically appropriate model space has already been identified could lead to an overselection on a set on models that (i) are nearly identical to one another and (ii) ignore established theory. This motivates our decision to emulate Bartels (1997) as the model for our BMA approach.

(1997), we use his formulas below for the calculating BIC (Bartels, 1997, see specifically 643-649), which consequently serves as an approximation of the Bayes factor for evaluating model performance under the unit uniform prior assumption (Montgomery and Nyhan, 2010; Raftery, 1995, 129-133):

$$-2\ln(\theta|y)+2k\ln(n)=BIC(M_j) \tag{1}$$

$$BIC(M_j)\approx -2\ln(L)+2k \tag{2}$$

where M_j is the set of models, where j denotes the set of models, n refers to the the number of observations in each model, L is the likelihood of one's model, and k is equivalent to the number of parameters, and finally, θ is the likelihood we would have observed the model given y the data (Bartels, 1997, 648). Overall the degree of uncertainty that is incorporated into the averaging stage depends on the prior beliefs of the researcher and theoretical expectations regarding the quantities of interest. After calculating the BIC, our approximate Bayes factor, for each model, researchers must decide on a Bayesian prior. The following equation shows the notation of a prior.

$$Pr(M_i)\equiv \pi_i^0 \tag{3}$$

where π_i^0 is representative of a uniform prior for a specific model, which is the equivalence of the prior probability of observing model, notated as $pr(M_i)$ with i indexing a particular model in the set j. Typical critiques of this prior assignment technique are that it is heavily biased and subjective to the preferences of the researcher and the strong assumptions regarding the choice of prior and its overall flexibility (Montgomery and Nyhan, 2010, 249-253). We understand the uneasiness of this practice, and recognize there are other possible prior choices, but we argue that our approach is not to prefer one model to another. Rather, we seek to give each model a chance to influence this process according to their balance between performance and fit, which we believe is captured best by the use of a uniform prior and its assumption of normality—an assumption that we believe is correct based on the established consensus regarding the trade volatility model subset [2] Because we believe that we have identified the correct set of models—

[2]Note that each model does meaningfully contribute to our core set of results, and that—in line with assumptions regarding the BIC—all BMA coefficient estimates roughly correspond to their previously established patterns given the assumptions regarding our choice of prior and Bayes factor.

but remain agnostic as to which model is best model—we assign a uniform prior allowing each model an equal prior probability of contributing information to our analyses, thus minimizing some of the subjectivity that is often criticized in the selection of priors. Once these priors are selected, we multiply them by our approximate Bayes factor (i.e. the BIC). After multiplication, these products are normalized by the sum of all Bayes factors corresponding to each model in order to return the posterior model probabilities displayed by:

$$\pi_i = \frac{B_{i_0}\pi_i^0}{\Sigma_j B_{j_0}\pi_j^0} \tag{4}$$

where π_i is the posterior probability of observing a particular model M_i, that is equal to B_{i_0}, which is Bayes factor for one model subscripted by i among the set of all our models M_j, multiplied by π_i, which is the uniform prior for each model, divided by the sum B_{j_0}, which is the set of Bayes factors for all models, and π_j^0 is the set of uniform priors for all the models j (Bartels, 1997, 648). The returned posterior model probabilities are conditional on the performance of each model, which was determined by the product of the BIC and the prior in the previous equations. In the final stage, the averaging stage, the expected values of the coefficients approximate the sum of the product of the posterior model probabilities and their respective coefficients. In a similar approach, the variances are calculated by sum of the product of the posterior model probabilities and their respective coefficients plus the sum of the product of the squared deviations between the conditional and unconditional posterior means (Bartels, 1997, 645). More formally,

$$E(\beta|X,y) \equiv \bar{b} = \Sigma_j \pi_j b_j \tag{5}$$

where $E(\beta|X,y)$ is the expected value of the coefficient estimate for variable X in model y, π_j is the set of posterior probabilities, and $\bar{b}$ is the vector of BMA coefficients.

$$V(\beta|X,y) = \Sigma_j \pi_j V(b_j) + \Sigma_j \pi_j (b_j - \bar{b})^2 \tag{6}$$

where $V(\beta|X,y)$ is the variance around the estimated model coefficients defined as β, π_j is the set of posterior probability for the models, b_j is a vector of coefficient values for set of models

M_j, and $\bar{b}$ is the vector of BMA coefficient values defined above (Bartels, 1997, 645).

References

Bartels, Larry M. 1997. "Specification Uncertainity and Model Averaging." American Journal of Political Science 41:641–674.

Kass, Robert E. and Larry Wasserman. 1995. "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion." Journal of American Statistical Association (431):928–934.

Montgomery, Jacob and Brendan Nyhan. 2010. "Bayesian Model Averaging: Theoretical Developments and Practical Applications." Political Analysis 18(2):245–270.

Raftery, Adrian. 1995. Bayesian Model Selection in Social Research. In Sociological Methodology, ed. Peter V. Marsden. Cambridge, MA: Blackwells.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." Annals of Statistics 6(2):461–464.