INTERNATIONAL RELATIONS ARE A SOCIAL NETWORK: A LATENT NETWORK APPROACH FOR GLOBAL POLITICS

SHAHRYAR MINHAS, CASSY L. DORFF, MARGARET FOSTER, MAX GALLOP, HOWARD LIU, JUAN TELLEZ,

AND MICHAEL D. WARD

ABSTRACT. International relations scholarship is filled with dyads. This includes dyadic hypotheses and propositions, but especially data. These relational data contain information about the interdependencies of various phenomena, including countries, dyads, and even triads. However, most empirical studies of such data do not take into account these dependencies. As a result, the studies are often contradictory and produce results that are not compelling. One reason is that the independence required by the typical methods employed does not exist in the data being studied. We present a different, regression based method, which constructs a latent network which incorporates 1^{st}, 2^{nd}, and 3^{rd}-order dependencies. We replicate five prominent studies in recent IR scholarship and compare the standard approach to the latent factor approach. The additive and multiplicative latent factor approach is shown to produce more precise estimates of covariate effects, and it also dominates standard approaches in terms of out-of-sample cross-validations.

Shahryar Minhas and Michael D. Ward acknowledge support from National Science Foundation (NSF) Award 1259266. Robert Franzese provided welcome and helpful comments on the broader project from which this effort emerges at the 2017 Summer Political Methodology Meetings. We also thank Douglas Gibler, of the University of Alabama, for kindly providing us with a copy of the data and programs he used in Gibler (2017).

1. Introduction

Aronow *et al.* (2015) estimate that over the period from 2010 to 2015, over sixty articles utilizing dyadic data were published in the *American Political Science Review*, *American Journal of Political Science*, and *International Organization*. Most of these studies use a generalized linear model (GLM) to estimate regression coefficients. However, extant approaches to studying dyadic data increase the chance of faulty inferences by treating data as independent and identically distributed (iid) when observations may be highly dependent. Most standard approaches assume that the problems raised by having non-iid relational data can be addressed by recalculating the standard errors of estimated parameters in the link function, so as to reflect the potential clustering of cases. This may work in limited situations, but is not generally effective because these palliatives do not address the fundamental data generating process that remains a threat to inference because of the interdependence of observations or measurements. Namely, it is not just the diagonals of the variance-covariance matrix this affects.

In this article, we discuss a Bayesian approach, the Additive and Multiplicative Effects (AME) model, for directly modeling relational data to reflect the data generating process that yields interdependencies in these types of data structures (Hoff, 2008; Minhas *et al.* , 2016). We focus on three types of interdependencies that can complicate dyadic analyses. First, dependencies may arise within a set of dyads as there may be a particular actor that is more likely to send or receive events such as conflict. Additionally, if the event of interest has a clear sender and receiver, we are likely to observe dependencies within a dyad, specifically, if a rebel group initiates a conflict with a government, the government will reciprocate that conflictual behavior. We capture these two dependencies, often referred to as first- and second-order dependencies, respectively, within the additive effects portion of the model. The multiplicative effects portion of the model captures dependencies that results from groups of actors clustering together or organizing into communities due to meso-scopic features of networks, such as homophily and stochastic equivalence.

We begin conducting a simulation study to show how AME can recover unbiased and well calibrated regression coefficients in the presence of network dependencies. Then we apply the AME model to five prominent studies in the international relations (IR) literature, and compare results from the current state-of-the-art approach (a GLM with robust standard errors) to those obtained with the AME framework. This latent factor approach is able to better capture first, second, and third-order interdependencies than the standard approaches. It also produces results that are more precise, and at times different, than those found in these literatures. As such this approach offers substantive insights, which are occluded by ignoring the interdependent nature of the relational data that characterizes many studies in the field of international relations. Finally, we show that for each replication our network based approach provides substantively more accurate out of sample predictions than the models used in the original studies. Thus the approach that we offer here is one that can be used by scholars in the field to not only generate substantive insights, but also enables us to better model the data generating process behind events of interest in international relations.

2. Dependencies in Dyadic Data

In thinking about dyadic data, scholars in the field begin by structuring it as a set of dyadic observations stacked on top of one another. Each observation is assumed to be independent of the others. Thus, for example, a conflict sent from actor the United States to Japan, is assumed to be independent of any action that Japan may send to the United States. Additionally, every action sent by Japan to others in the system is considered independent even though each of those interactions involves a common sender, i.e, Japan. As a result, the assumption that most begin with is that each dyadic interaction is taking place in isolation of the others.

Sender	Receiver	Event
i	j	y_{ij}
$\vdots$	k	y_{ik}
	l	y_{il}
j	i	y_{ji}
$\vdots$	k	y_{jk}
	l	y_{jl}
k	i	y_{ki}
$\vdots$	j	y_{kj}
	l	y_{kl}
l	i	y_{li}
$\vdots$	j	y_{lj}
	k	y_{lk}

Table 1. Structure of datasets used in canonical design.

$\longrightarrow$

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Table 2. Adjacency matrix representation of data in Table 1. Senders are represented by the rows and receivers by the columns.

To start to move away from this assumption and better understand the dependencies that emerge in relational data it is helpful to shift towards structuring dyadic data in the form of an adjacency matrix as shown in the top right of Table 2. Here rows designate the senders of an event and columns the receivers. The cross-sections in this matrix represent the action that was sent by an actor on the row to one designated in the column. Thus y_{ij} designates an action, such as a conflictual event or trade flows, that is sent from actor i to actor j.

Using the structure of an adjacency matrix we can visualize the types of first and second order dependencies that complicate the analysis of relational data in traditional GLMs. Figure 1 clarifies the types of dependencies that can manifest in these types of data structures. The adjacency matrix on the top left highlights a particular row of an adjacency matrix, to illustrate that values across a particular row of an adjacency matrix may be more similar to each other than other values in the adjacency matrix because each of these values has a common sender. Homogeneity in interactions involving a common sender also manifest heterogeneity in how active actors are across the network when compared to each other. Thus in most relational datasets (e.g., trade flows, conflict, participation in international organizations, even networks derived from Twitter or Facebook) we often find that there are some actors who are much more active than others (Barabási & Réka, 1999). Unless one is able to develop a model that can account for the variety of

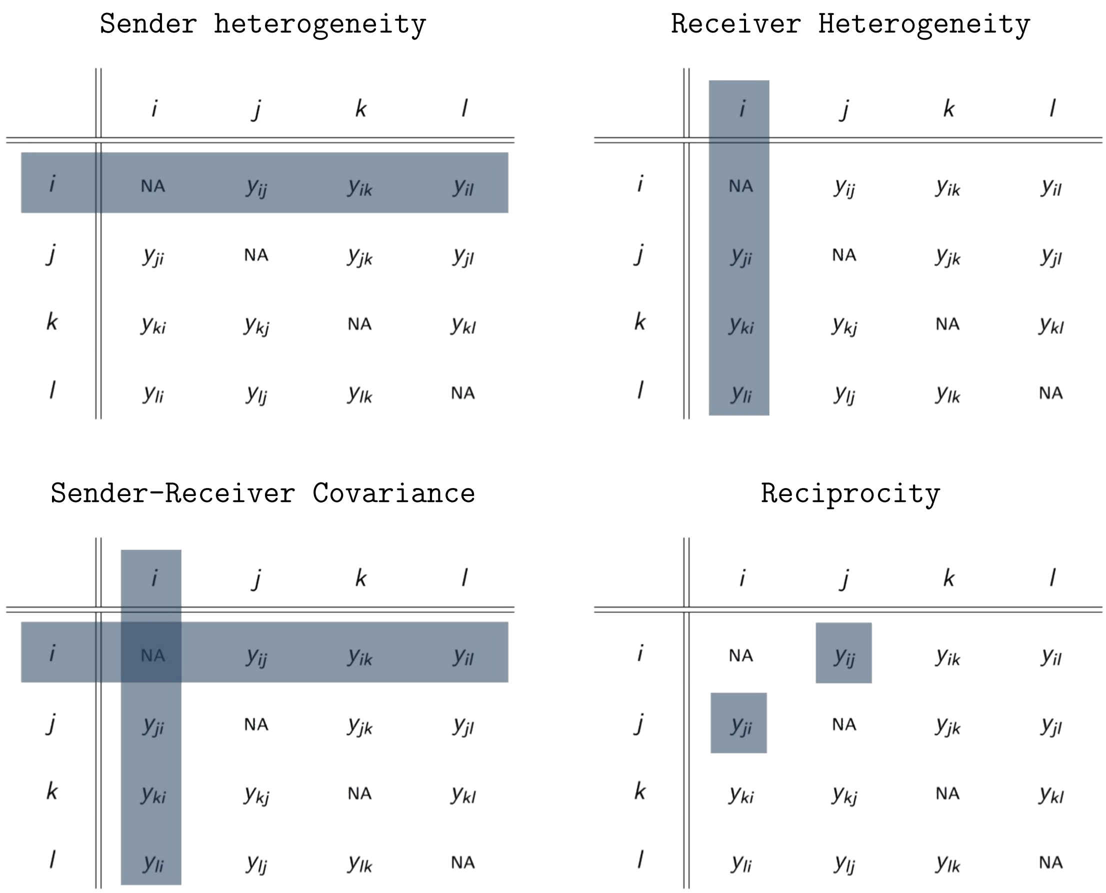


Figure 1. Nodal and dyadic dependencies in relational data.

explanations that may play a role in determining why a particular actor may be more active than others, parameter estimates from standard statistical models will be biased.

For similar reasons one also needs to take into account that there is a shared dependence between dyadic observations that share a common receiver. The bottom-left panel, illustrates that sender and receiver type dependencies can also blend together. Specifically, as actors who are more likely to send ties in a network tend to also be more likely to receive them. As a result, the rows and columns in an adjacency matrix are often correlated. For example, consider that trade flows both from and to many wealthy, developed countries. The bottom-right panel, highlights a second order dependence, specifically, reciprocity. This is a dependency occurring within dyads involving the same actors whereby values of y_{ij} and y_{ji} are correlated. The concept of reciprocity has deep roots in the study of relations between states (Richardson, 1960; Keohane, 1989). The purpose of highlighting each of these dependencies through the set of panels in Figure 1 is to more easily visualize how each of these dependencies manifest in relational data. Alternatively, when we simply simply stack dyads in some arbitrary order, these dependencies are easy to ignore.

The dependencies discussed so far involve at most dependence between two actors. For most relational data, however, dependencies do not simply manifest at this level. More often we find significant evidence of higher order structures that result from dependencies between multiple groups of actors. These dependencies arise because there may be a or some set of latent attributes between actors that affects their probability of interacting with one another (Wasserman & Faust, 1994; Zinnes, 1967). In Figure 2 we provide a visualization of a hypothetical relational dataset. Here the nodes designate actors and edges between the nodes indicate that an interaction between the two took place. Each node is colored by the latent group to which they belong.

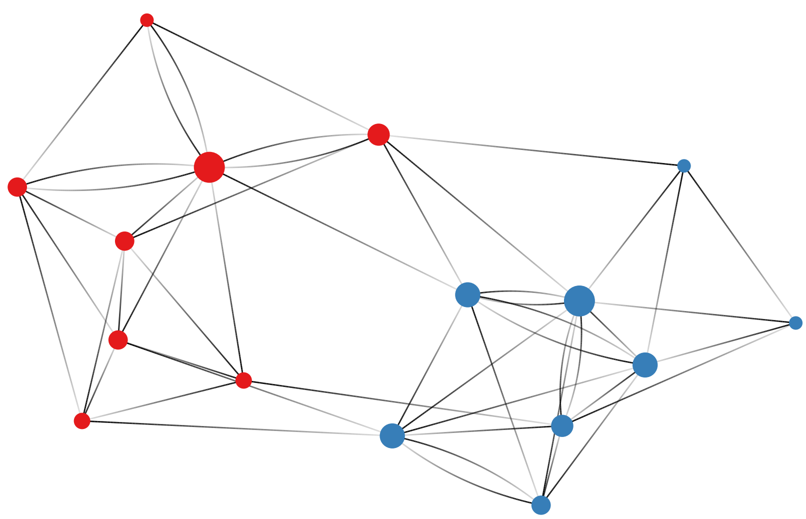


Figure 2. Visualization of network with meso-scopic features.

Clear from the visualization is that the actors belonging to the same group have a higher likelihood of having an interaction with each other than those from the other group. A prominent example of a network with this type of structure was found by Adamic & Glance (2005), who visualized the ways in which right and left leaning political blogs linked to one another in the 2004 United States Election (Adamic & Glance, 2005). They found that the degree of interaction between right and left leaning blogs was minimal, and that most of these blogs simply linked to those of their own ilk. This showcases the types of higher order dependencies that can emerge in relational data. First, the fact that interactions was determined by a shared attribute, in this case political ideology, is an example of homophily. Homophily can be used to explain the emergence of patterns such as transitivity ("a friend of a friend is a friend") and balance ("an enemy of a friend is an enemy"), which also have a long history in international relations. The other major type of mesoscopic feature that emerges in relational data is community structure, which is often formalized through the concept of stochastic equivalence (Anderson *et al.* , 1992). This concept simply refers to the idea that groups of nodes that act similarly in the network are stochastically equivalent. In the example we have laid out above each of the left leaning blogs would be considered stochastically equivalent to one another.

The major implication of the presence of homophily, stochastic equivalence, and the other dependencies we discussed above are that they complicate the practical assumption of observational independence. Inferences drawn from models that ignore potential interdependencies between dyadic observations face a number of well-known challenges: a) biased estimates of the effect of independent variables, b) uncalibrated confidence intervals, and c) poor predictive performance. By ignoring these potential interdependencies, we often ignore important features of the problem under study. The study of international relations is founded on the relations among actors. Why ignore the interdependencies that led to the study of IR in the first place?

3. Additive and Multiplicative Effect Models for Networks

The AME approach can be used to conduct inference on cross-sectional and longitudinal networks with binary, ordinal, or continuous linkages. It is flexible and easy to use for analyzing the kind of relational data often found in social science. It accounts for nodal and dyadic dependence patterns, as well as higher-order dependencies such as homophily and stochastic equivalence. We do not know whether interdependence dominates international relations, but as noted by Thoreau in reference to rumors that dairymen on strike were watering down milk: "…some circumstantial evidence is very strong, as when you find a trout in the milk." At a very minimum, it is necessary to examine whether there is interdependence since it challenges substantive arguments as well as statistical modeling (Snijders, 2011).

Further parameter interpretation in the AME framework is straightforward because it has a simple regression basis as shown in Equation 1 which portrays directed matrix

$$
\begin{aligned}
y_{ij} &= g(\theta_{ij}) \\
\theta_{ij} &= \boldsymbol{\beta}^\top \mathbf{X}_{ij} + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} + \alpha(\mathbf{u}_i, \mathbf{v}_j) \text{ , where} \\
\alpha(\mathbf{u}_i, \mathbf{v}_j) &= \mathbf{u}_i^\top \mathbf{D}\mathbf{v}_j = \sum_{k \in K} d_k u_{ik} v_{jk}.
\end{aligned} \tag{1}
$$

With this framework, it is straightforward to model dyadic observations as conditionally independent given $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ depends on the the unobserved random effects modeled to account for the potential 1^{st}, 2^{nd}, and 3^{rd}-order dependencies. In Equation **??**, $a_i + b_j + \epsilon_{ij}$ represent the additive random effects in this framework and account for sender, receiver, and within-dyad dependence. The multiplicative effects, $\mathbf{u}_i^\top \mathbf{D}\mathbf{v}_j$, capture higher-order dependence patterns in $\boldsymbol{\theta}$ after accounting for any known covariate information.[1]

4. Simulation study

Here we consider inference for a regression parameter β of a linear or generalized linear model for a network in the case where there is an omitted variable. The true data-generating models we consider are of the form

$y_{i,j} \sim \mu + \beta x_i x_j + \gamma w_i w_j + \epsilon_{i,j}$

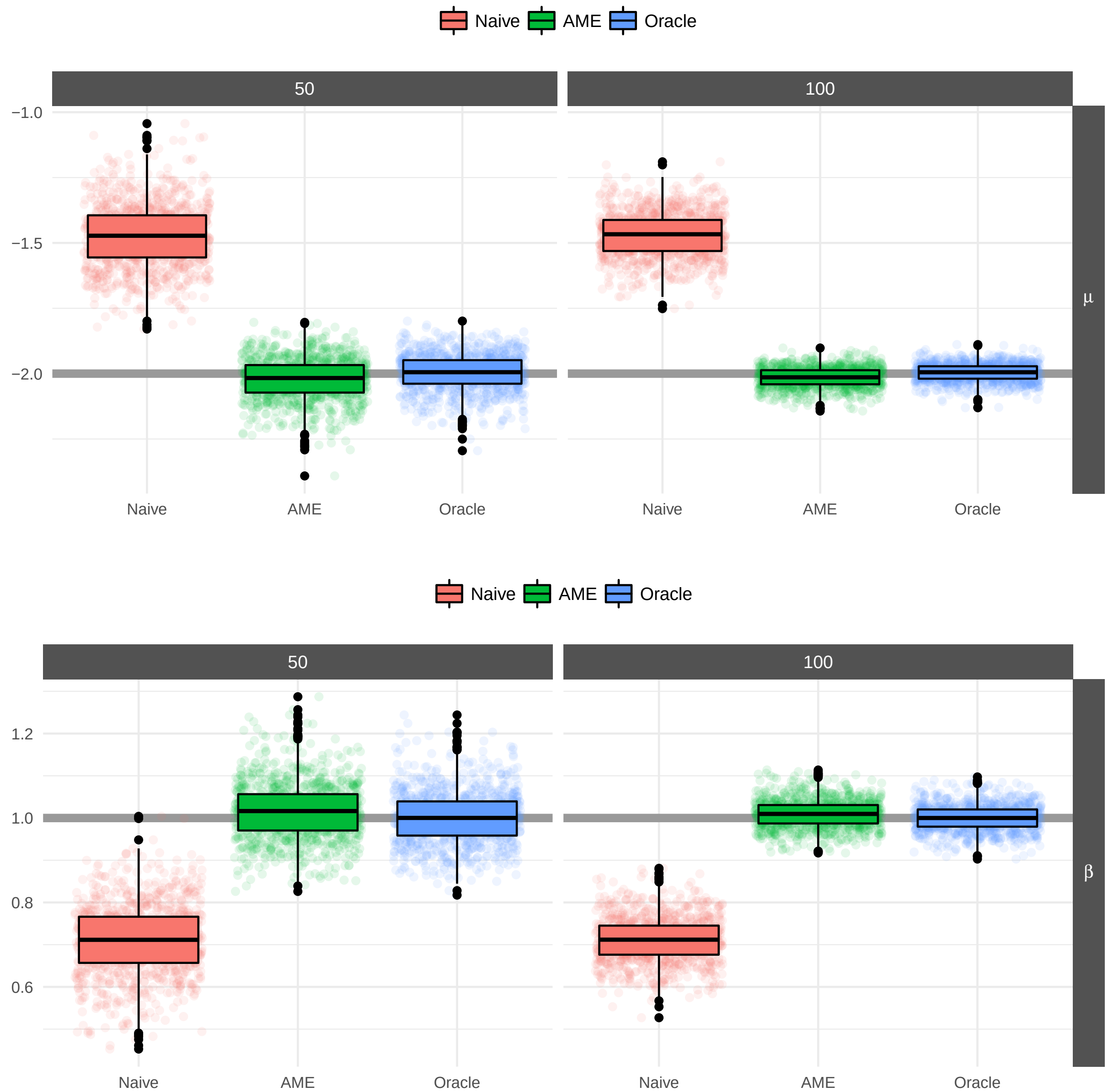
where $Y = \{y_{i,j}\} \in \mathbb{R}^{n \times n}$ is an observed sociomatix, $x = \{x_i\} \in \mathbb{R}^n$ is a vector of observed node-specific characteristics, and $w = \{w_i\} \in \mathbb{R}^n$ is a vector of unobserved node-specific characteristics. We compare inference for β using three models:

- A naive regression model assuming independent errors;
- A latent factor model;
- An "oracle" regression model that includes both $x_{i,j} = x_i x_j$ and $w_{i,j} = w_i w_j$.

We make these comparisons in the context of a binary network outcome with a probit regression model.

[1] This procedure is described in detail in Hoff (2008) and Minhas *et al.* (2016). A Bayesian procedure in which parameters are iteratively updated using a Gibbs sampler is described in the appendix.

Figure 3. Bias in parameter when homophily is ignored.

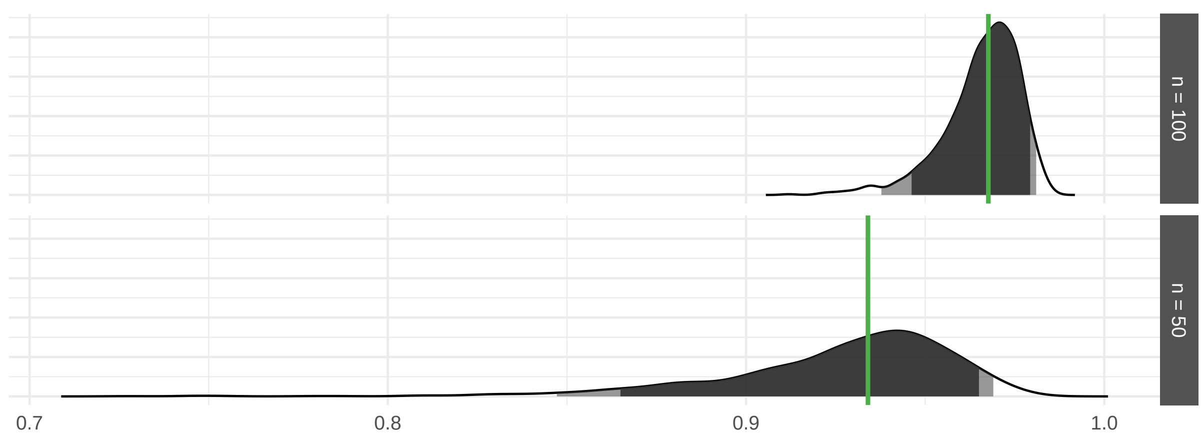


One way to compare inferences across models is with bias and variance of the parameter estimates. Both of these quantities are combined to get the MSE.

Alternatively, bias and precision can be summarized by confidence interval coverage and width. Coverage should ideally be at the nominal level. If two methods have the same actual coverage rate, the one with the narrower intervals is preferred.

Figure 4. Coverage in parameter when homophily is ignored.

Figure 5. Correlation between missing variable and multiplicative random effect in AME.



5. Replications

5.1. **Design.** For the purpose of this study, we choose five prominent studies from the broad field of international relations and international political economy that utilize relation data (McDonald, 2004; Reiter & Stam, 2003; Rose, 2004; Weeks, 2012; Gibler, 2017). We chose to replicate studies that were written fairly recently, since 2000, and that have been cited over 100 times. Each of these pieces was published in a prominent journal and is well-known in the literature. They all used the standard approach in political science, which is to employ some form of general linearized regression that ignores dyadic interdependencies. Post estimation, standard errors are often adjusted in an attempt to account for clustering of observations.

We obtained the data for each of these studies from their replication archives and replicated the main results of each of the articles.[2] We examine each of the models using the AME framework described above. Our goal is to ascertain whether the ignored interdependencies—the non-iid structure of the relational data—would result in different model estimates when they were addressed in an AME framework, and more importantly to see if there were substantive opportunities that were presented with the dynamic factor approach.

Finally, we assess whether there is any substantive finding that emerges or indeed if any disappear once the interdependencies in the data are modeled.

The broader goal, beyond introducing the use of the AME framework in an applied setting, is to examine the extent to which interdependencies within typical dyadic data make much difference in what we have learned about international relations from empirical studies over the past decade or so. We believe that it does, and that the dynamic latent factor model provides a step forward.

[2]Without exception this was straightforward to accomplish, thanks to an increasing norm in the social sciences of open data sharing.

Table 3. Features of the Studies Replicated.

	Model	# of Actors	Years	# of Dyads	Type of Dyads	Clustered $\sigma_{\hat{\beta}}$
Weeks (2012)	Logit			$901,540$	Directed	Robust
Reiter & Stam (2003)	Logit			$753,456$	Directed	Robust
McDonald (2004)	Logit			$92,354$	Undirected	Robust
Gibler (2017)	Logit			$650,557$	UnDirected	no
Rose (2004)	OLS			$234,597$	Directed	Robust

	Central Finding	Does it Replicate in a Network Model?
Weeks (2012)	Bosses, Juntas and Strongmen are more Aggressive, Machines are Not	**Fails to Replicate**
Reiter & Stam (2003)	Personalist Regimes Attack Democracies, Not Vice Versa	Replicates
McDonald (2004)	Lower Trade Barriers and Higher Trade Lead to Peace	**Fails to Replicate**
Gibler (2017)	Power Parity at time of Entry to International System Inceases Conflict	**Fails to Replicate**
Rose (2004)	WTO Membership Does not Effect Trade	Replicates

Table 4. Here we provide a brief summary of the key variable in each of the five replications and a note about whether or not the finding is replicated when using our network based approach. Cases in which the finding is not replicated are highlighted in bold.

We also examine the accuracy of the predictions made with each approach. Out of sample cross validation strategy.

By accounting for exogenous and network dependent patterns that give rise to conflict systems we are able to better account for the data generating process underlying relational data structures. To show that this is the case, we examine whether our approach achieves better predictive performance in an out of sample context than traditional dyadic models. To evaluate our model, we randomly divide the $\binom{n}{2} \times T$ data values into $k = 30$ sets, letting $s_{ij,t}$ be the set to which pair ij, t is assigned. Then for each $s \in \{1, \ldots, k\}$, we:

(1) estimate model parameters with $\{y_{ij,t} : s_{ij,t} \neq s\}$, the data not in set s,

(2) and predict $\{\hat{y}_{ij,t} : s_{ij,t} = s\}$ from these estimated parameters.

The result of this procedure is a set of sociomatrices $\hat{Y}$, in which each entry $\hat{y}_{ij,t}$ is a predicted value obtained from using a subset of the data that does not include $y_{ij,t}$.

We set a number of benchmarks for comparison. First we compare the AME model to a GLM model using the same covariates to show the effect of accounting for network dependencies on predicting conflict. We supplement this with an alternative GLM that includes not just these covariates, but also a lagged dependent variable and a lagged reciprocity term. The lagged dependent variable is the equivalent of saying that conflict and peace are relatively likely to persist between dyads, while the inclusion of a lagged reciprocity term in a GLM framework is a simple way to account for retaliatory strikes.

We utilize three performance criterions to compare the models: Receiver Operator Characteristic (ROC) curves, Precision Recall (PR) curves, and separation plots. ROC curves look at the trade-off between true positive rates and false positive rates at different thresholds of classification. An issue with an ROC Curve when looking at conflict, is that it is relatively rare at the dyadic level: in most years only 3% of possible dyads are in conflict with one another. If peace is common, even a poor model will have a very low False Positive Rate.

To better assess which models predict the presence of conflict, not just its absence, we look at PR Curves. These examine the trade-offs between the percentage of conflicts a model predicts, and the percentage of predicted conflicts which occur. Lastly, we examine separation plots (Greenhill *et al.* , 2011). These provide an intuitive visualization of the accuracy of our predictions by juxtaposing a line showing the predicted probability of conflict with whether conflict actually occurs for all cases (where the cases are sorted by the predicted probability and then colored to indicate the outcome). Here a perfect model would have all cases where conflict actually exists on the right with a predicted probability of 1, and would predict 0 in all other cases. All of the models' performance out of sample by these metrics are displayed in figure **??**. The AME model with covariates is the best performing model out of sample in all cases. This model outperforms each of the GLM variants by a notable margin.

		GLM	AME
Weeks (2012)	Area Under ROC Curve (AUC-ROC)	0.64	**0.97**
	Area Under PR Curve (AUC-PR)	0.00	**0.15**
Reiter & Stam (2003)	AUC (ROC)	0.92	**0.96**
	AUC (PR)	0.08	**0.15**
McDonald (2004)	AUC (ROC)	0.92	**0.99**
	AUC (PR)	0.13	**0.28**
Gibler (2017)	AUC (ROC)	0.52	**0.91**
	AUC (PR)	0.00	**0.08**
Rose (2004)	Root Mean Squared Error (RMSE)	3.23	**1.99**
	Root Median Squared Error (RMDSE)	2.01	**1.06**

Table 5. Here we provide a summary of the out-of-sample performance based on our cross-validation strategy for each of the five replications when using the standard dyadic approach and our network based approach. Four of the five studies involved a binary dependent variable, so for those measures we provide area under the curve (AUC) statistics. The fifth involved a gaussian dependent variable and for that we use the root mean squared error (RMSE) and root median squared error (RMDSE). Cases in which our network based approach outperformed the standard approach are highlighted in bold.

5.2. **Replication of Weeks (2012).** Weeks (2012) examines the influence of domestic institutions on the initiation of military conflicts by autocratic leaders. She argues that in some circumstances autocrats are held accountable for their foreign policy decisions. She adds the nuance that autocratic audiences are not homogeneous. When the autocratic regime is nonmilitary, the domestic audience do not favor military actions, but in military autocracies this is not the case. Further she argues that in personalistic regimes without a military or civilian domestic audience, the leaders tend to be more likely to employ military force in their foreign policy. To study this question, she uses a dyadic design in which the dependent variable is "whether country A in a directed dyad initiated military conflict against country B during year t" (page 337). These data come from the Militarized Interstate Disputes database (Maoz, n.d.). One major innovation in her study resides in the nuanced way in which she conceptualized and coded regime type into four types: a) Machine, b) Junta, c) Boss, and d) Strongmen. She also includes a variety of putative control variables focusing on capabilities for both sides of the dyad, alliances, geography, trade dependence, regime instability, and the regime type of "side B." She uses a logistic regression, but follows Beck *et al.* (1998) and includes splines to capture temporal covariation in the dependent variable along with fixed, unit effects. The analysis is done for dyads, but is considered to be from the perspective of the actor that initiated the dispute. Her time frame is from 1946-1999. The basic finding is that a) juntas, bosses, and strongmen are more likely to initiate conflict than machines (and maybe democracies) and that b) machines are no more belligerent than democracies. These insights are mainly determined by looking at coefficients and their putative significance in Tables 1 and 2 (pages 339–340). She makes the argument that ignoring important nuances between different types of autocracies hinders our understanding of the initiation of military conflict by autocracies.

The replication of Weeks (2012) likely has the sharpest divergence between the General Linear Model results and those of the AME Model. In Weeks's initial models, as detailed in figure 6 she finds that machines are less prone to initiate conflict than the reference category, whereas Juntas, Bosses and Strong-men are more conflict-prone, in line with her theoretical expectations. When we look at the posterior distribution of these coefficients in the AME results, we find that none of these values are distinguishable from zero. We similarly find less pronounced effects for military capabilities. One explanation for this divergence is the AME model's ability to account for third order effects. Inspection of the multiplicative effects in figure 8 reveals a number of clusters of states which exhibit structural equivalence – in the top right corner we see states like the US, the UK, and Israel who often send conflict to similar targets. In particular, we observe a cluster of "rogue states," who are receivers of conflict, in the top right (Iran, Iraq, Syria, Libya, North Korea). In the bottom corner we see a cluster of authoritarian senders including Iraq, Russia, Syria, North Korea and China. In general, these clusters have similar governmental types (Iraq, Syria, Libya, and North Korea all fell under the "boss" category), and similar conflict behavior. So, in the GLM, which ignores these third order dependencies, much of these results might have been attributed to regime type. The AME model, on the other hand, finds that it is more efficient to attribute this behavior to the multiplicative effects. In terms of out of sample performance, shown in figure 7(a), the AME model performs markedly better than the GLM out of sample, lending credence to the possibility that these third order dependencies cause spurious effects for regime type.

5.3. **Replication of Reiter & Stam (2003).** Reiter & Stam (2003) examine the relationship between democracy, dictatorship and the initiation of militarized disputes. They use directed dyads and find that dyads involving a democratic leader on the one hand and a personalist dictator on the other tend to be violent. They also discover that dictators are likely to challenge democracies, but that

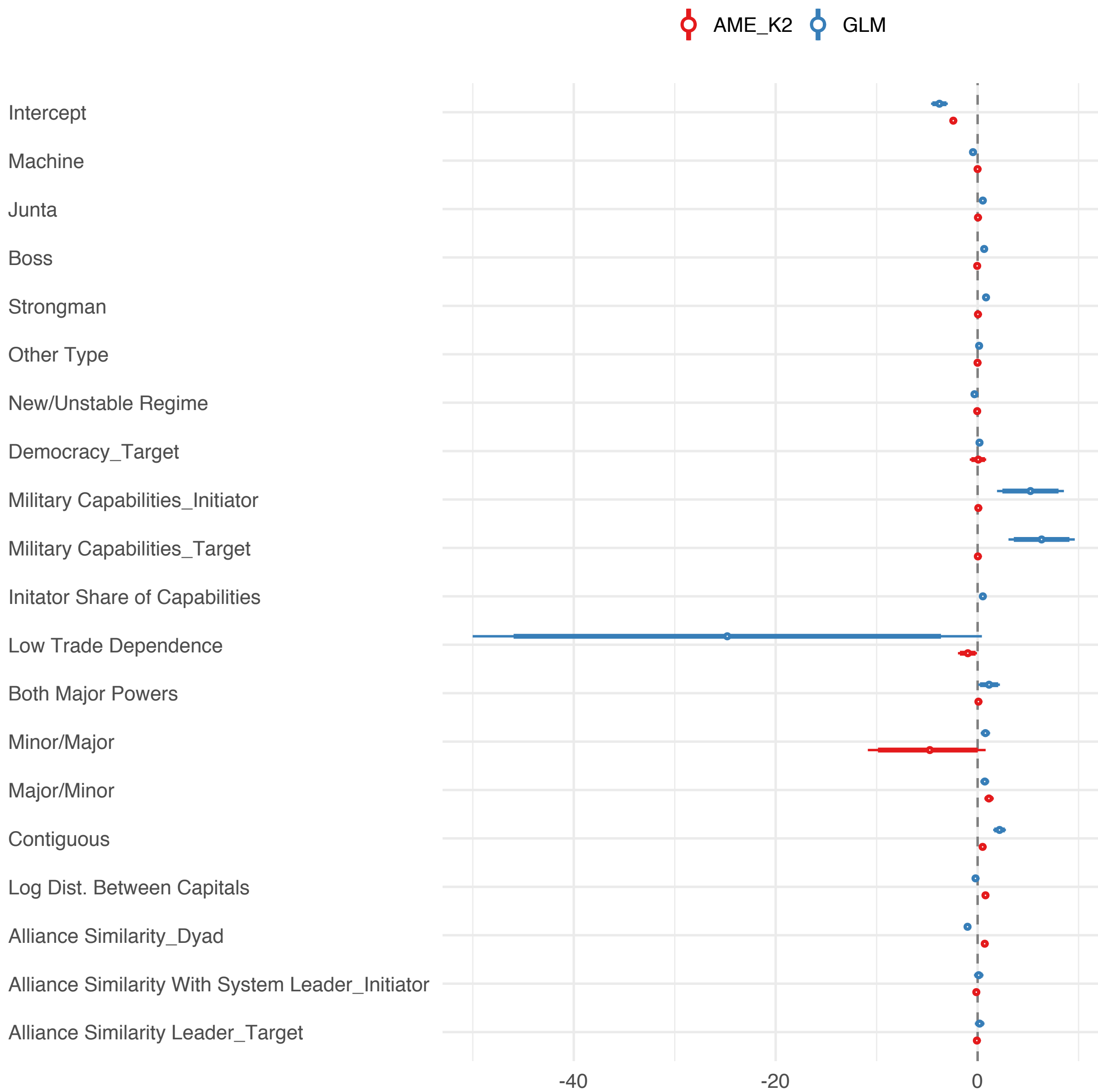
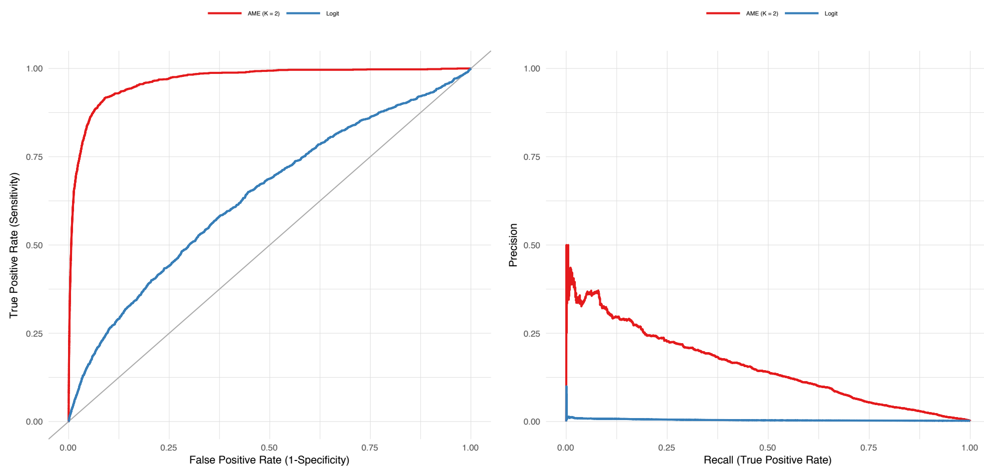


Figure 6. Coefficient plot of Weeks' (2012) original model (blue) compared to AME model (red).

this is not reciprocal. In addition, military regimes and single-party regimes are more prone to initiate disputes with democracies, than the other way around. They use the MID data, but note that "We code a state as having initiated a dispute if it is on 'side A' of a MID, the conventional approach to coding initiation. This means that the state was on the side that took the first action in the dispute, whether that action was the threat, display, or use of force. We code joiners as



(a) AUC (b) Precision and Recall

Figure 7. Assessments of out-of-sample predictive performance for Weeks (2012) using ROC curves and PR curves.

initiators or targets, though the results do not change if we do not code joiners as initiators or targets. ...Though coding initiation will always be difficult, the 'side A' variable has been widely used in past conflict scholarship (page 334)." Independent variables are largely taken from an earlier study and focus on various encodings of regime types, contiguity, alliance, and capability measures. As is prevalent in these kinds of studies, Reiter & Stam employ a logistic regression that includes an indicator of the time since the last dispute as well as three cubic splines. The database for this study is constructed using EUGene (Bennett & Stam, 2000) and comprises approximately three-quarters of a million stacked dyads. Based on their statistical analysis, they conclude that institutional constraints affect the propensity of democratic and non-democratic leaders to engage in military conflict.

In the original model, the variable "Pers/Democ Directed Dyad" (which represents a Personalist $\rightarrow$ Democractic directed dyad) is clearly positive while the variable "Democ/Personalist Directed Dyad" is zero and the difference between the two coefficients is clearly distinct from zero. In our replication using the AME framework, we also find that Pers-Democ directed dyad has a positive effect with zero excluded from the 95% confidence interval while Democ-Pers directed dyad is indistinguishable from zero. Using this model, however, we can no longer conclusively say that the Pers/Democratic coefficient is larger than the Democ/Personalist one. Our replication using the

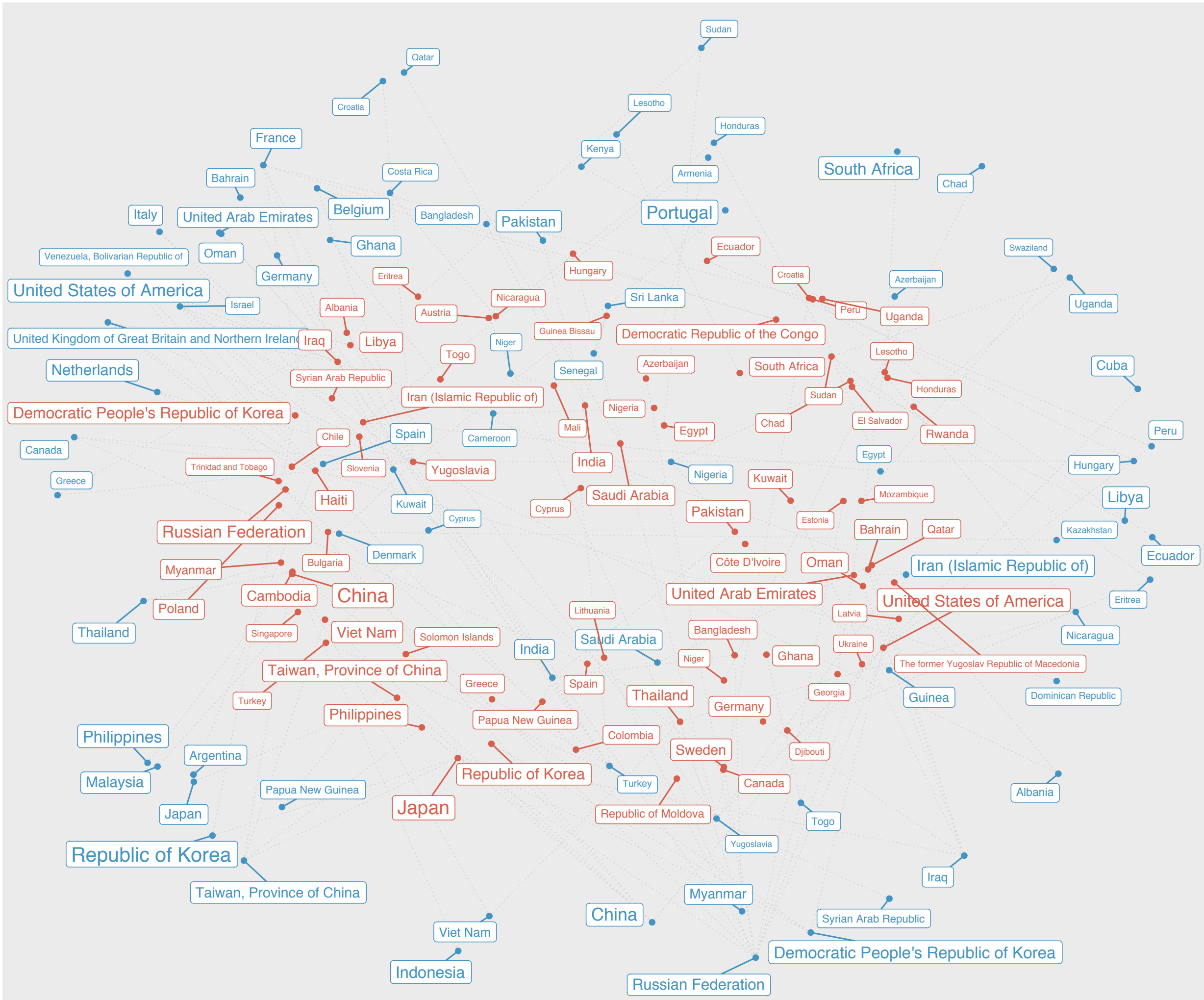


Figure 8. Visualization of multiplicative effects for Weeks (2012). Blue represents groups with common sending patterns and red represents groups with common receiving patterns.

AME approach therefore cast doubt on Reiter & Stam's key claim that MIDs initiated by personalist dictatorships against democracies are more likely than MIDS initiated by democracies. Further, the effect of most of the covariates in the literature thought to predict interstate MIDs are much closer to zero when using the AME framework, as seen in Figure 9. Finally, our modeling approach outperforms the original model by better, and more accurately, predicting MIDs out-of-sample (Figure 10(a) and Figure 10(b)).

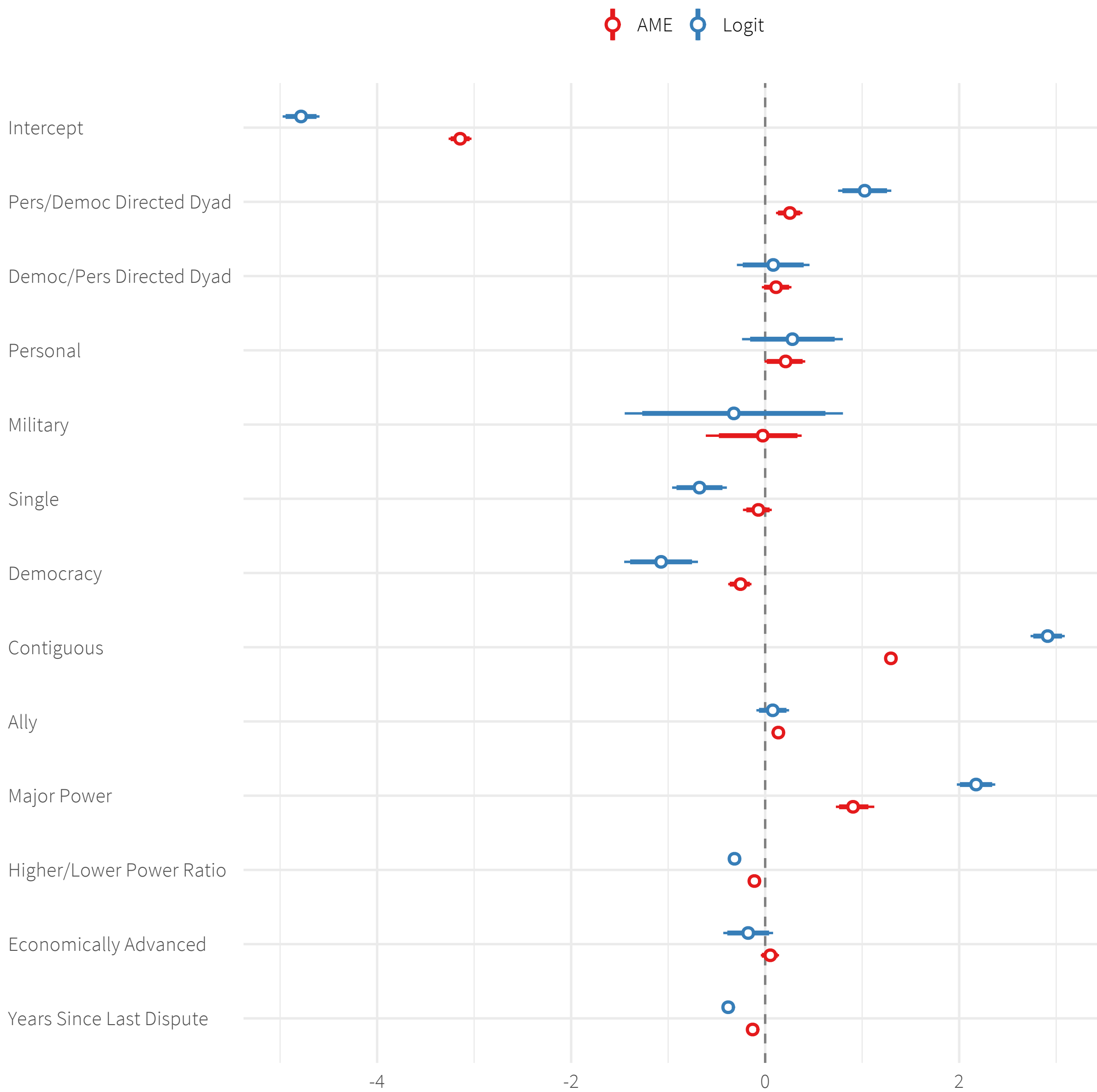
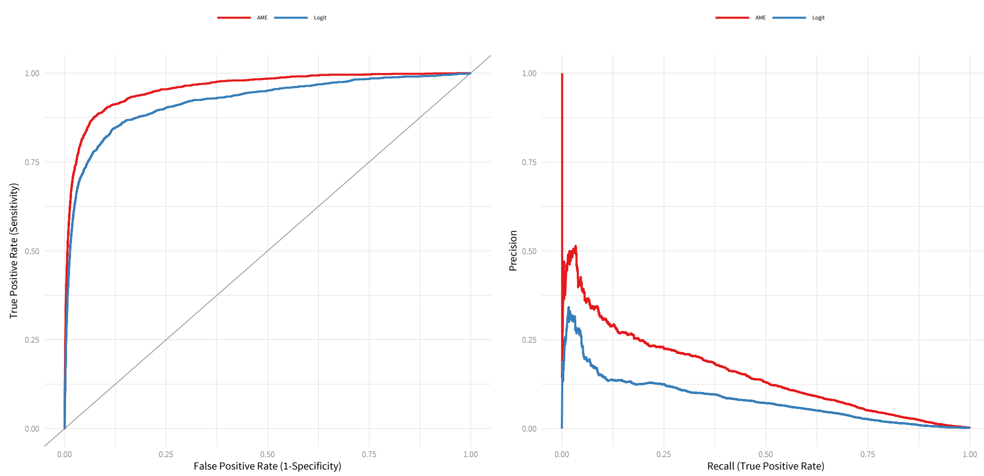


Figure 9. Coefficient plot of Reiter & Stam (2003)'s original model (blue) compared to AME model (red).

5.4. **Replication of Rose (2004).** In 2004, Andrew Rose published a study in the *American Economic Review* that proved to be quite controversial in terms of macroeconomic trade theory and in terms of trade policy in a variety of nations. It also provoked a number of responses in the international political economy literature Tomz *et al.* (2007); Ward *et al.* (2013). Rose's basic argument is that despite longstanding arguments made by trade theorists and the World Trade Organization



(a) AUC (b) Precision and Recall

Figure 10. Assessments of out-of-sample predictive performance for Reiter & Stam (2003) using ROC curves and PR curves.

that WTO membership fosters greater cooperation and thereby more trade among its members, the empirics do not bear out such claims. He uses a standard gravity model with dyadic data on bilateral merchandise trade (not services) for 175 countries over a period of five decades. Estimating this model using OLS within many differing contexts, his conclusion was that: "An extensive search reveals little evidence that countries joining or belonging to the GATT/WTO have different trade patters from outsiders... (2004, page 98, abstract)." The data for this study have been widely used in replications by many searching for the missing effects of the WTO—as well as preferential trade agreements, bilateral investment theories, and other aspects of modern trade theory.

When we compare the results of Rose's original OLS model to our model that accounts for network dependencies, the results are generally similar. As you can see in Figure 11, the main result of the model – the null effect of membership in the WTO, as represented by the "One-In" and "Both-In" variables – remains when we move from an OLS to a Gaussian AME model. The most striking difference between the models is that, while in the original model there was a clear positive relationship between Real GDP and Trade, most of this effect vanishes in the AME model. The random effects shown in Figure 12 reveal the cause of much of this divergence. Here, the states with the most positive random effects are also states with high GDP, though not necessarily high

GDP/capita.[3] Thus, the effect of GDP in the original model was, in part, an artifact of first-order dependencies. Most of the other results of the model are constant across each model, though some geographic features, such as islands and landlocked states, have a more clear effect on trade once we account for these network dependencies. Excitingly, when we account for network interdependencies, we observe a markedly lower Root Mean Squared Error out of sample – 3.23 for the OLS model and 1.77 for the AME model.

[3]Note: Qatar exhibits strongly negative random effects.

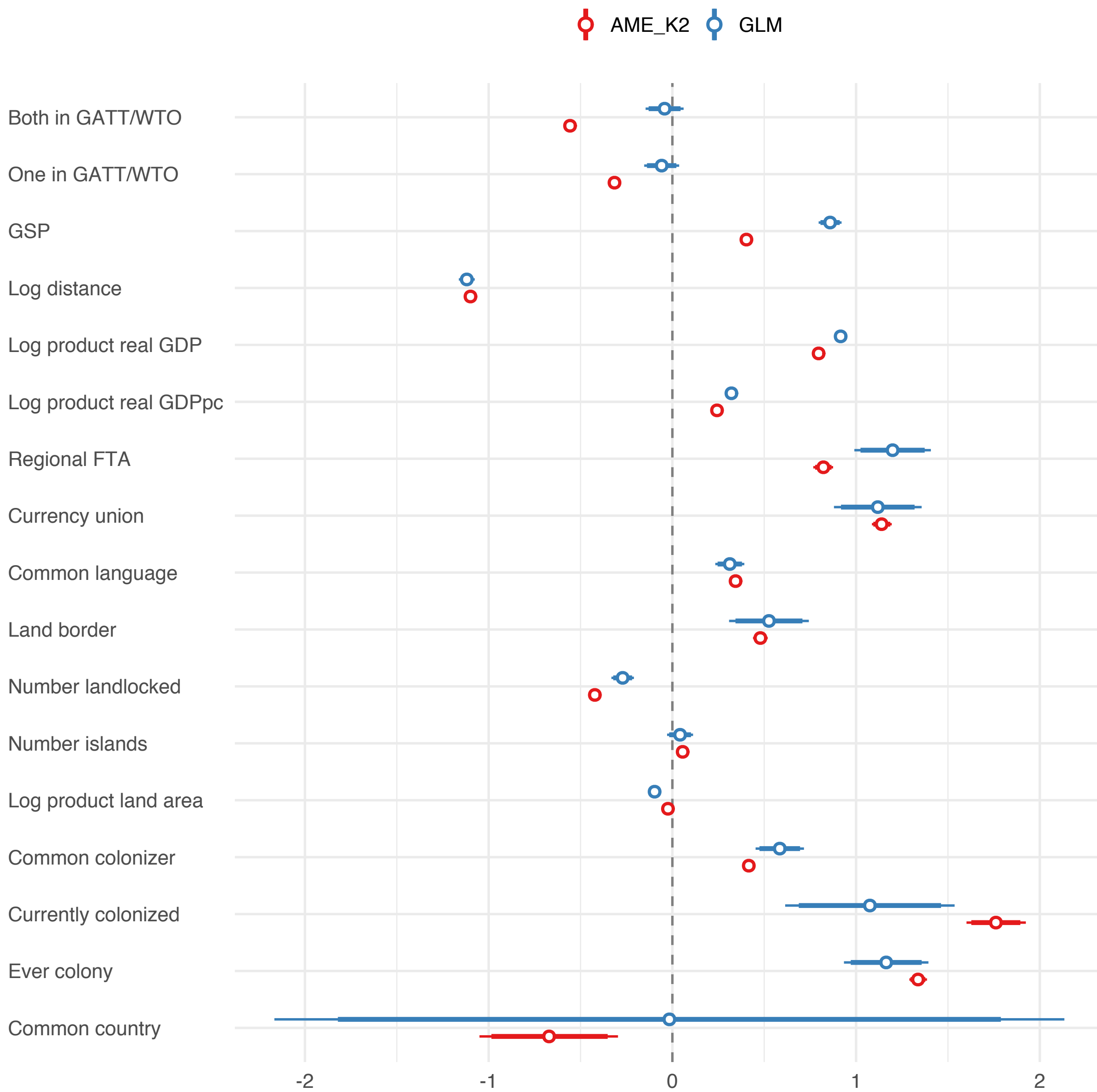


Figure 11. Coefficient plot of Rose (2004)'s original model (blue) compared to AME model (red).

5.5. **Replication of McDonald (2004).** McDonald (2004) studies whether trade promotes peace between nations. He observes that knowledge about the link between conflict and trade is indeterminate in the field of international relations, noting that competing explanations persist. He calls for more precise empirical tests. Importantly, McDonald (2004) includes the interdependence

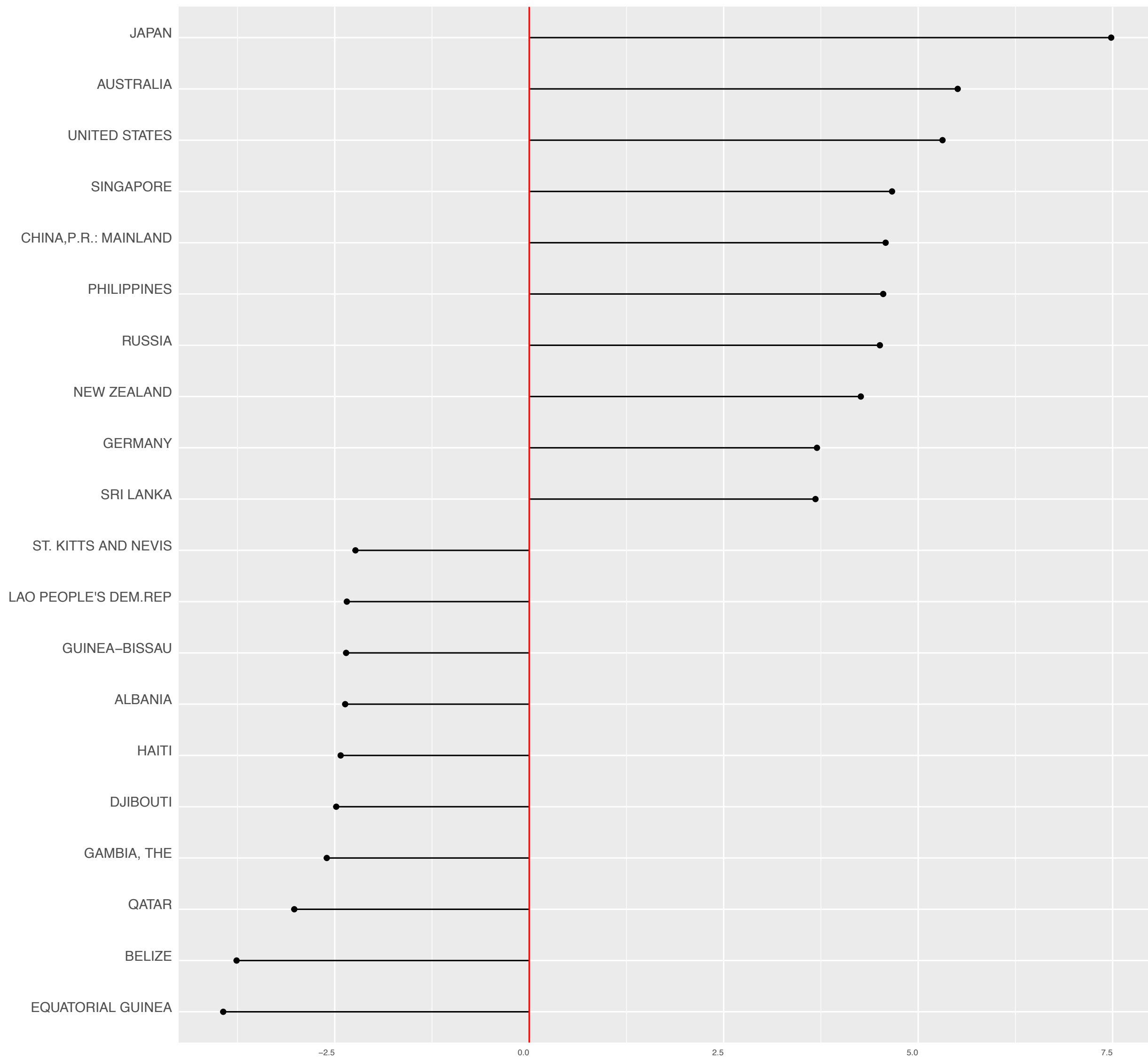


Figure 12. Nodal Random Effects for AME estimation of Rose (2004).

argument—that interdependence between states "makes conflict less likely because of its efficiency over conquest in acquiring resources. . . (547)" in his overview of underdeveloped hypotheses.

Accordingly, the primary contribution of the study is to provide evidence challenging the generalized linkage between peace and trade and to offer a new measurement of the key independent variable, trade. To do so, McDonald (2004) refines the trade variable, arguing that *free* trade, rather than trade alone, reduces the likelihood of conflict between states. His key hypothesis that greater levels of protection increase the probability of interstate conflict, an argument that builds on the work of classic liberalism and connects free trade to the power of domestic audiences. McDonald (2004) measures free trade in two ways. The first captures the idea that larger protected sectors generate greater societal pressures resulting in pockets of support for war. This protection variable measures the proportion of customs revenue divided by total imports in the state that possesses the greater such ratio in each dyad. This measure captures the score of the state in the dyad that possesses higher barriers to trade (560). McDonald (2004) also includes a measure of economic integration calculated as "the lower proportion of total dyadic trade (imports plus exports) divided by state i's GDP or total dyadic trade divided by state j's GDP." (560). Finally, the binary, dependent variable is the onset of a new militarized interstate dispute within a given dyad. McDonald (2004) employs logistic regression to examine the putative statistical significance of these variables. The models include a spline correction for time-series data as well as robust standard errors clustered on each dyad.

Our replication reveals that trade relations are highly interdependent and exhibit important patterns of transitivity. Or, in other words, if countries i and j are highly dependent and countries j and k are also highly dependent, then we are likely to observe high dependency between countries i and k. Our findings support those of Traag and Lupu (2013) which argue that indirect trade relations reduce the probability of conflict. This indicates that conflict is less likely between members of a trade community. Once we control for these dependencies, we can more clearly interpret the positive link between trade and conflict. Figure 13 shows the coefficients for both the original and replicated model and Figure 14 demonstrates the predictive performance of each model. Figure 15 represents the multiplicative effects of the model. This plot enables us to consider actors with similar sending and receiving patterns resulting from stochastic equivalence and homophily. In the upper arch for senders, there is a clear clustering of United States' allies (the United Kingdom, Netherlands, and Japan). We also observe an African cluster in the top right segment for receivers (Egypt, Niger, and Zambia for example). These countries both trade amongst themselves and have similar conflict behavior – stochastic equivalence is present among the first group, homophily among the second – and so once we account for these clusters using the multiplicative effects, the residual effect of trade actually increases the likelihood of conflict.

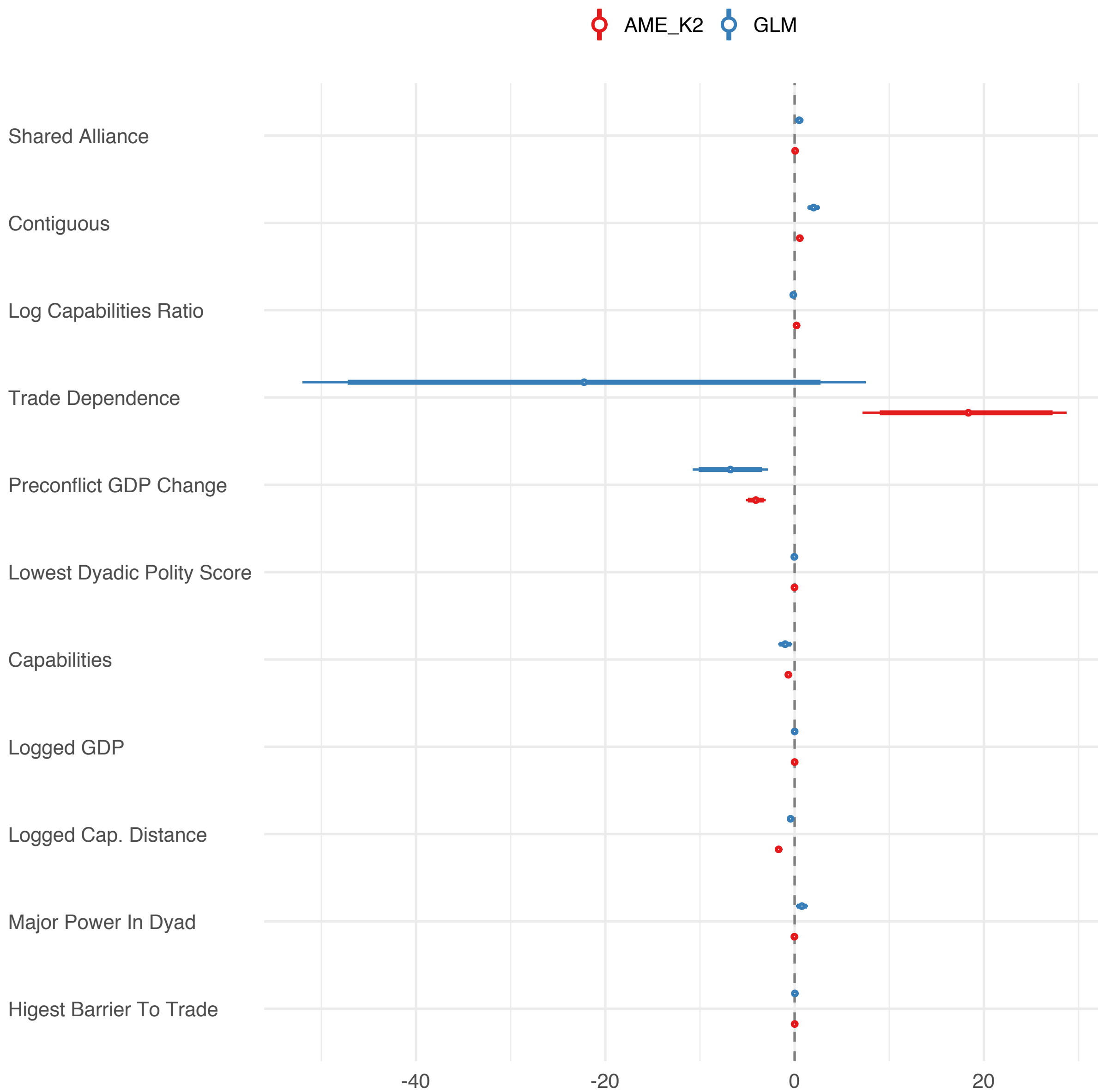
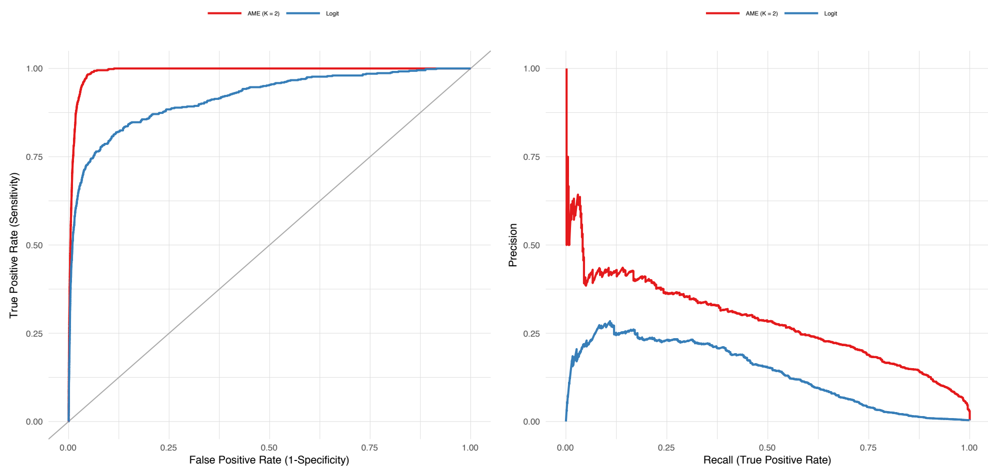


Figure 13. Coefficient plot of McDonald (2004)'s original model (blue) compared to AME model (red).

5.6. **Replication of Gibler (2017).** The replications we have undertaken are all from articles over the past fifteen years. A more recent example is Gibler (2017) which examines the onset of militarized disputes using capabilities, joint democracy, alliances, and power parity in a undirected dyadic study using logistic regression and dyad clustered standard errors. In addition to this, Gibler shows that the long-standing relationship between the relative parity of capabilities and initiation



(a) AUC (b) Precision and Recall

Figure 14. Assessments of out-of-sample predictive performance for McDonald (2004) using ROC curves and PR curves.

of international conflict is almost completely mediated by the initial conditions for the members of the dyad when they joined the international system as sovereign members. This finding calls into question many IR theories about the role of balance in terms of generating international conflict (Organski, 1958).

We re-estimated model 6 from Table 6 (2017,34). The results are presented in Table 6 in order to facilitate explicit comparison. The results obtained with `amen` stand in stark contrast to those found with a logistic regression (with dyad clustered, robust standard errors). Most importantly, the primary variable from the Gibler study, parity of the members of a dyad at the year in which they entered the international system, is shown to be unimportant in the amen results. Not only is the value of the this parameter small, but it has a very large relative standard error, over a magnitude larger than the parameter itself ($z = 0.038$). In addition, joint democracy follows the same pattern of importance in the logistic results, but this disappears once interdependencies are modeled. As might be expected the strong geographic clustering in the original study is about one-quarter as strong in the `amen` estimations. Similarly, rivalry coefficients are about one-third the size in the `amen` formulation, but a great deal more precisely measured ($z = 18.116$).

Beyond more informative fixed effect coefficients, the `amen` approach also provides information about the interdependencies that were modeled. The most pertinent of these may be the dyadic

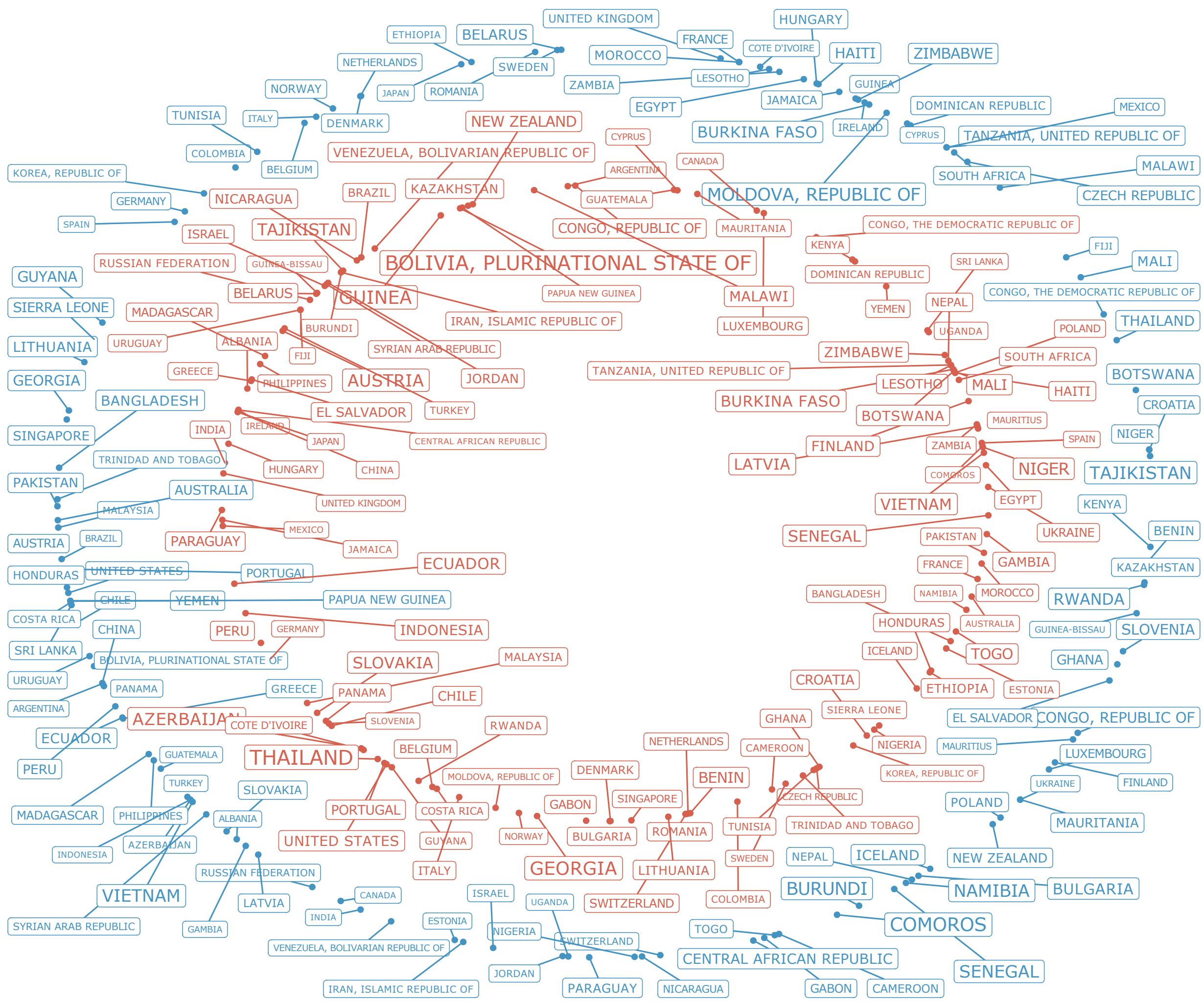


Figure 15. Visualization of multiplicative effects. Blue represents groups with common sending patterns and red represents groups with common receiving patterns.

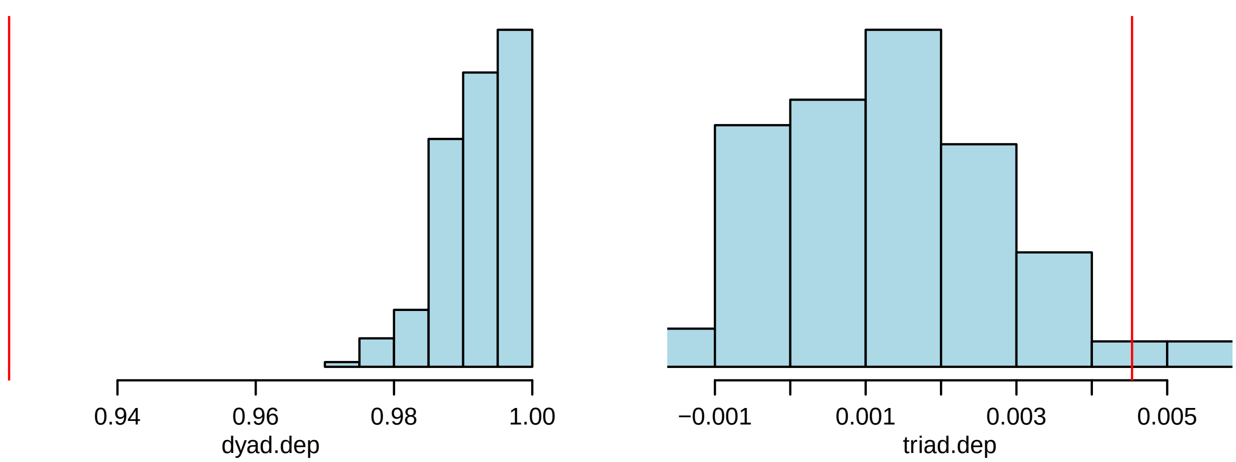
and triadic dependencies which are presented in Figure 16, which further reveal that the assumption of independence among the dyadic data in this study can be strongly rejected. Perhaps most importantly, the main substantive conclusions of the 2017 Gibler study do not seem warranted from the perspective of the results obtained with the `amen` estimation that explicitly models 1st-2nd-, and 3rd-order interdependencies. Not only do the estimated coefficients tell a different story, but the assumptions under which the original results would hold are shown to be violated by the data.

Table 6. Comparison of Gibler (2017) Model 6 results with AME results. This was run over a chain of $10,000$ iterations.

	$\hat{\beta}$ Estimates	
Variable	`logit`	`amen`
Allied	0.142	0.023
Joint Democracy	**-0.507**	0.045
Peace Years	-0.260	**-0.060**
Spline 1	**-0.001**	**0.00**
Spline 2	**-0.000**	**0.00**
Spline 3	-0.000	**0.00**
Contiguity	**2.412**	**0.640**
Parity	0.075	-0.013
Parity at entry year	**0.868**	0.002
Rivalry	**2.031**	**0.721**
Constant	**-5.526**	**-2.581**

Note: **Bold** indicates conventional statistical significance at $p < 0.05$ or less in Gibler (2017). For comparative purposes, only, we have employed the same criterion to the results from the `amen` estimation.

Figure 16. The left panel is a plot of the dyadic dependencies in the Gibler model 6; the right panel shows the triad dependencies. Note that the red line represents the null hypothesis of no dependencies, indicating that the standard logistic approach is far from what is uncovered with this analysis.



6. Conclusion

It is no longer necessary to assume that the interesting, innate interdependencies in relational data can be ignored. Nor do they have to be approximated with ad hoc, incomplete solutions that purport to control for dependencies by modifying the post-estimation standard errors of the estimated coefficients. Instead, the interdependencies may be addressed directly with additive and multiplicative effects in the context of a generalized linear model that provides more reliable inferences and better out-of-sample predictive performance, along with new substantive insights.

APPENDIX

Additive and Multiplicative Effects Gibbs Sampler. To estimate, the effects of our exogenous variables and latent attributes we utilize a Bayesian probit model in which we sample from the posterior distribution of the full conditionals until convergence. Specifically, given observed data **Y** and **X** – where **X** is a design array that includes our sender, receiver, and dyadic covariates – we estimate our network of binary ties using a probit framework where: $y_{ij,t} = 1(\theta_{ij,t} > 0)$ and $\theta_{ij,t} = \boldsymbol{\beta}^\top \mathbf{X}_{ij,t} + a_i + b_j + \mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j + \epsilon_{ij}$. The derivation of the full conditionals is described in detail in Hoff (2005) and Hoff (2008), thus here we only outline the Markov chain Monte Carlo (MCMC) algorithm for the AME model that we utilize in this paper.

- Given initial values of $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and $\sigma_\epsilon^2\}$, the algorithm proceeds as follows:
  - sample $\boldsymbol{\theta} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - sample $\boldsymbol{\beta} \mid \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - sample $\mathbf{a}, \mathbf{b} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - sample $\Sigma_{ab} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \rho,$ and σ_ϵ^2 (Inverse-Wishart)
  - update ρ using a Metropolis-Hastings step with proposal $p^*|p \sim$ truncated normal$_{[-1,1]}(\rho, \sigma_\epsilon^2)$
  - sample $\sigma_\epsilon^2 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab},$ and ρ (Inverse-Gamma)
  - For each $k \in K$:
    - Sample $\mathbf{U}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}_{[,-k]}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
    - Sample $\mathbf{V}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}_{[,-k]}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
    - Sample $\mathbf{D}_{[k,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)[4]

[4]Subsequent to estimation, **D** matrix is absorbed into the calculation for **V** as we iterate through K.

References

Adamic, Lada A., & Glance, Natalie. 2005. The political blogosphere and the 2004 US election: Divided they blog. *Pages 36–43 of: Proceedings of the 3rd international workshop on Link discovery*. ACM.

Anderson, Carolyn J., Wasserman, Stanley, & Faust, Katherine. 1992. Building stochastic blockmodels. *Social Networks*, **14**(1), 137–161.

Aronow, Peter M., Samii, Cyrus, & Assenova, Valentina A. 2015. Cluster–Robust Variance Estimation for Dyadic Data. *Political Analysis*, **23**(4), 564–577.

Barabási, Albert-László, & Réka, Albert. 1999. Emergence of Scaling in Random Networks. *Science*, **286**(October 15), 509–510.

Beck, Nathaniel, Katz, Jonathan N., & Tucker, Richard. 1998. Taking Time Seriously: Time-series-cross-section Analysis with a Binary Dependent Variable. *American Journal of Political Science*, **42**(2), 1260–1288.

Bennett, D. Scott, & Stam, Allan C. 2000. Research Design and Estimator Choices in the Analysis of Interstate Dyads: When Decisions Matter. *Journal of Conflict Resolution*, **44**(5), 653–685.

Gibler, Douglas M. 2017. State Development, Parity, and International Conflict. *American Political Science Review*, **111**(1), 21–38.

Greenhill, Brian, Ward, Michael D., & Sacks, Audrey. 2011. The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Data. *American Journal of Political Science*, **55**(4), 991–1002.

Hoff, Peter D. 2005. Bilinear Mixed-Effects Models for Dyadic Data. *Journal of the American Statistical Association*, **100**(4690), 286–295.

Hoff, Peter D. 2008. Modeling homophily and stochastic equivalence in symmetric relational data. *Pages 657–664 of:* Platt, John C., Koller, Daphne, Singer, Yoram, & Roweis, Sam T. (eds), *Advances in Neural Information Processing Systems 20*. Processing Systems 21. Cambridge, MA, USA: MIT Press.

Keohane, Robert O. 1989. Reciprocity in international relations. *International Organization*, **40**(1).

Maoz, Zeev. *Dyadic Militarized Interstate Disputes Dataset Version 2.0*.

McDonald, Patrick J. 2004. Peace through Trade or Free Trade? *Journal of Conflict Resolution*, **48**(4), 547–572.

Minhas, Shahryar, Hoff, Peter D., & Ward, Michael D. 2016 (October). *Let's Say Amen for Latent Factor Models*. Working paper.

Organski, A.F.K. 1958. *World Politics: the Stages of Political Development*. New York: Alfred A. Knopf.

Reiter, Dan, & Stam, Allan C. 2003. Identifying the Culprit: Democracy, Dictatorship, and Dispute Initiation. *American Political Science Review*, **97**(2), 333–337.

Richardson, Lewis F. 1960. *Arms and Insecurity*. Chicago and Pittsburgh, PA: Quadrangle/Boxwood.

Rose, Andrew K. 2004. Do We Really Know That the WTO Increases Trade? *American Economic Review*, **94**(1), 98–114.

Snijders, Tom A.B. 2011. Statistical Models for Social Networks. *Annual Review of Sociology*, **37**, 131–53.

Tomz, Michael, Goldstein, Judith, & Rivers, Douglas. 2007. Do We Really Know That the WTO Increases Trade? *American Economic Review*, **97**(5), 2005–2018.

Ward, Michael D., Metternich, Nils W., Dorff, Cassy L., Gallop, Max, Hollenbach, Florian M., Schultz, Anna, & Weschle, Simon. 2013. Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction. *International Studies Review*, **16**(4), 473–644.

Wasserman, Stanley, & Faust, Katherine. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Weeks, Jessica L. 2012. Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict. *American Political Science Review*, **106**(2), 326–347.

Zinnes, Dina A. 1967. An Analytical Study of the Balance of Power Theories. *Journal of Peace Research*, **3**, 270–288.

SHAHRYAR MINHAS: DEPARTMENT OF POLITICAL SCIENCE

Current address: Michigan State University

E-mail address, Corresponding author: minhassh@msu.edu

CASSY L. DORFF

Current address: Department of Political Science, University of New Mexico

E-mail address: cdorff@unm.edu

MARGARET FOSTER

Current address: Department of Political Science, Duke University

E-mail address: margaret.foster@duke.edu

MAX GALLOP

Current address: Department of Government and Public Policy, University of Strathclyde

E-mail address: max.gallop@strath.ac.uk

HOWARD LIU

Current address: Department of Political Science, Duke University

E-mail address: hao.liu@duke.edu

JUAN TELLEZ

Current address: Department of Political Science, Duke University

E-mail address: juan.tellez@duke.edu

MICHAEL D. WARD

Current address: Department of Political Science, Duke University

E-mail address: michael.d.ward@duke.edu