

## HOW TO FIX IR

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<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup>-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) claim 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*.<sup>1</sup> Most of these studies use a generalized linear model (GLM), either OLS or logistic regression, to estimate regression coefficients. However, we know that extant approaches to studying dyadic data increase the chance of faulty inferences by treating data as independent and identically distributed (iid) when they are highly dependent and intertwined. 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, in order 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 which remains a threat to inference because of the interdependence of observations or measurements. It is not just the diagonals of the variance-covariance matrix this affects.

In this article, we introduce a Bayesian approach for directly modeling relational data to reflect the data generating process that yields interdependencies in the data. This model has two parts, one of which uses a set of additive random effects capturing 1<sup>st</sup>- and 2<sup>nd</sup>-order dependencies in the relational data. The second part is a multiplicative component that captures 3<sup>rd</sup>-order dependencies, i.e., phenomena that leads to triads such as homophily and stochastic equivalence. We apply this model, called the Additive and Multiplicative effects (AME) model, to five prominent studies in the international relations literature, and compare results from the current state-of-the-art approach (a generalized linear model with robust standard errors) to those obtained with the AME framework.

This latent factor approach is able to better capture 1<sup>st</sup>-, 2<sup>nd</sup>-, and 3<sup>rd</sup>-order interdependencies than the standard approaches. It also produces results that are more precise, and often different, than those found in these literatures, and as such 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 out-of-sample predictions (cross-validations) are substantially improved with this approach.

---

<sup>1</sup>Aronow et alia (2015) provide a solution to simple clustering via covariance estimation but do not address the broader issues of interdependence focused on herein.

## 2. LIMITATIONS OF STANDARD FRAMEWORK

Generalized linear models are typically expressed via a stochastic and systematic component. The stochastic component reflects assumptions about the probability distribution from which the data are generated:  $y_{ij} \sim P(Y|\theta_{ij})$ , with a probability density or mass function such as the normal, binomial, or Poisson. Each dyad in the sample is independently drawn from a particular distribution, given  $\theta_{ij}$ . The systematic component characterizes the model for the parameters of that distribution and describes how  $\theta_{ij}$  varies as a function of a set of nodal and dyadic covariates,  $\mathbf{X}_{ij}$ :  $\theta_{ij} = \beta^T \mathbf{X}_{ij}$ . A fundamental assumption we make when applying this modeling technique is that given  $\mathbf{X}_{ij}$  and the parameters of the distribution, each of the dyadic observations is conditionally independent.

The importance of this assumption becomes clearer in the process of estimating a GLM via maximum likelihood. After having chosen a set of covariates and specifying a distribution, the joint density function over all dyads using the observations from Table 1 is given:

$$(1) \quad P(y_{ij}, y_{ik}, \dots, y_{lk} | \theta_{ij}, \theta_{ik}, \dots, \theta_{lk}) = P(y_{ij} | \theta_{ij}) \times P(y_{ik} | \theta_{ik}) \times \dots \times P(y_{lk} | \theta_{lk})$$

$$P(\mathbf{Y} | \boldsymbol{\theta}) = \prod_{\alpha=1}^{n \times (n-1)} P(y_{\alpha} | \theta_{\alpha})$$

The joint probability is converted into a likelihood:  $\mathcal{L}(\boldsymbol{\theta} | \mathbf{Y}) = \prod_{\alpha=1}^{n \times (n-1)} P(y_{\alpha} | \theta_{\alpha})$ .

The parameters are estimated by maximizing the likelihood. The likelihood as defined above is only valid if  $y_{ij}$  is independent of  $y_{ji}$  and  $y_{ik}$  given the set of covariates we specified, or the values of  $\theta_{ij}$ .

## 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). Ignoring the assumption of independence is a widely-noted shortcoming of many studies, but it continues to be the standard practice in much scholarship. Fortunately, it is easy to empirically examine dependencies in relational data.

The additive and multiplicative effects model (AME) can help to eliminate this color-blindness to interdependence. It was introduced in Hoff (2008) and elaborated in Hoff (2015); the open-source computer package associated with this approach is Hoff *et al.* (2017). We describe it below.

Parameter interpretation is straightforward because it has a simple regression basis as shown in Equation 2 which portrays directed matrix

$$\begin{aligned}
 y_{ij} &= g(\theta_{ij}) \\
 \theta_{ij} &= \beta^T \mathbf{X}_{ij} + e_{ij} \\
 e_{ij} &= a_i + b_j + \epsilon_{ij} + \alpha(\mathbf{u}_i, \mathbf{v}_j), \text{ where} \\
 \alpha(\mathbf{u}_i, \mathbf{v}_j) &= \mathbf{u}_i^T \mathbf{D} \mathbf{v}_j = \sum_{k \in K} d_k u_{ik} v_{jk}.
 \end{aligned}
 \tag{2}$$

With this framework, it is straightforward to model dyadic observations as conditionally independent given  $\theta$ , where  $\theta$  depends on the the unobserved random effects modeled to account for the potential 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup>-order dependencies. In Equation 3,  $a_i + b_j + \epsilon_{ij}$  represent the additive random effects in this framework and account for sender, receiver, and within-dyad dependence. The multiplicative effects,  $\mathbf{u}_i^T \mathbf{D} \mathbf{v}_j$ , capture higher-order dependence patterns in  $\theta$  after accounting for any known covariate information. A Bayesian procedure in which parameters are iteratively updated using a Gibbs sampler is available in the **amen** R package to estimate this type of generalized linear mixed effects model from Gaussian, binary, ordinal, and other relational data types.

**Addressing Dependencies in Dyadic Data.** Relational, or dyadic, data provide measurements of how sets of actors relate to one another. Typically, pairs of actors are examined. These pairs are called *dyads*. Such data can be arranged as a directed dyadic design in which the unit of analysis is some set of  $n$  actors that have been paired together to form a matrix of dyads, i. e., represented as a  $n \times n$  matrix,  $\mathbf{Y}$ , with diagonals typically undefined. The  $ij^{th}$  entry defines the relationship of  $i$  and  $j$  and can be continuous or discrete. It can also be symmetric (known as undirected) or asymmetric (called directed). This is not the stacked dyad structure normally found in social science, but it is equivalent and presents a format in which the interdependencies are exposed and transparent. The standard statistical approach to these data uses a linear or generalized linear model (GLM) which assumes that each of the dyadic observations is conditionally independent on observed regressors, even though events and relationships between these actors is most often *interdependent*.

One important dependency is nodal dependency, whereby each actor has similar interactions with many of its partners, leading to significant heterogeneity in activity levels across nodes. An important implication of this across-node heterogeneity is within-node homogeneity of ties: the values across a row, say  $\{y_{ij}, y_{ik}, y_{il}\}$ , are similar to each other because each has a common sender. If a sender  $i$  is more active in the network than other senders, it will have greater homogeneity of its linkages. More popular targets show greater similarity across linkages in a given column,  $\{y_{ji}, y_{ki}, y_{li}\}$ . Actors which are more likely to send ties in a network are also more likely to receive them. As a result the rows and columns in an adjacency matrix may be correlated. Another structural interdependence is reciprocity. This is a second-order, or dyadic dependency whereby values of  $y_{ij}$  and  $y_{ji}$  are correlated. If the data are simply stacked dyads in some arbitrary order, this kind of dependency is harder to visualize, but exists nonetheless.

The presence of these interdependencies in relational data complicates 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?

**3.1. Social Relations Model: Additive Part of AME.** Warner *et al.* (1979) developed a social relations model (SRM), using ANOVA decomposition, that focuses on such dependencies. SRM provides the error structure for the additive effects component of the AME framework. SRM decomposes the variance of dyadic observations in terms of a) heterogeneity across row means (often called out-degree), b) column means (or, in-degree), c) correlation between row and column means, and d) correlations within dyads. Wong (1982); Li & Loken (2002) provide a random effects representation of the SRM:

$$\begin{aligned}
 y_{ij} &= \mu + e_{ij} \\
 e_{ij} &= a_i + b_j + \epsilon_{ij} \\
 \{(a_1, b_1), \dots, (a_n, b_n)\} &\stackrel{\text{iid}}{\sim} N(0, \Sigma_{ab}) \\
 \{(\epsilon_{ij}, \epsilon_{ji}) : i \neq j\} &\stackrel{\text{iid}}{\sim} N(0, \Sigma_{\epsilon}), \text{ where} \\
 \Sigma_{ab} &= \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \quad \Sigma_{\epsilon} = \sigma_{\epsilon}^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},
 \end{aligned}
 \tag{3}$$

wherein  $\mu$  is the density of a network, and  $e_{ij}$  represents residual variation in each cell (or dyad). The residual variation has three separate parts: a) row/sender effects ( $a_i$ ), b) column/receiver effects ( $b_j$ ), and c) within-dyad effects ( $\epsilon_{ij}$ ). The row and column effects are modeled jointly to account for correlation in how active an actor is in sending and receiving ties. Heterogeneity in the row and column means is modeled by  $\sigma_a^2$  and  $\sigma_b^2$ , respectively.  $\sigma_{ab}$  describes whether actors which send a lot of ties also receive a lot. Second-order dependencies are described by  $\sigma_{\epsilon}^2$  and a reciprocity, parameter  $\rho$ .

The SRM covariance structure described in Equation 3 can be incorporated into the systematic component of a GLM framework to produce the social relations regression model (SRRM):  $\beta^T \mathbf{X}_{ij} + a_i + b_j + \epsilon_{ij}$ , where  $\beta^T \mathbf{X}_{ij}$  accommodates the inclusion of dyadic, sender, and receiver covariates (Hoff, 2005). This is a regression framework with additive, random effects, but it models variances

and covariances often seen in relational data. Furthermore, this accommodates a diversity of outcome distributions. In the case of binary data, this can be done by utilizing a random effects logistic (or probit) latent variable representation.

**3.2. Latent Factor Model: Multiplicative Part of AME.** The additive part described above does not handle third-order dependency. A third-order dependency is the dependency between triads, not dyads. Third-order effects can arise from the presence of some set of shared attributes between actors that affects their probability of interacting with one another. Individuals try to close triads because this is thought to be a more stable or preferable social situation (Wasserman & Faust, 1994; Zinnes, 1967).

For example, a common finding is that democracies are more likely to form trade agreements with one another. Both geographic proximity and a country's political system are examples of homophily, which captures the idea that the relationships between actors with similar characteristics in a network are likely to be stronger than nodes with different characteristics. 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 have a long history in international relations. A binary network where actors tend to form ties based on shared characteristics often leads to a network with a high number of "transitive triads" in which sets of actors  $\{i, j, k\}$  are each linked to one another. Thus, unless we specify a list of exogenous variables that determine this prevalence of triads, the probability of  $j$  and  $k$  forming a tie is not independent of the ties that already exist between those actors and others.

Another third-order dependence pattern is stochastic equivalence. Actors  $i, j$  are stochastically equivalent if the probability of  $i$  relating to every other actor is similar to the probability for  $j$ . Equivalence implies the existence of groups of nodes in a network with similar relational patterns. Recent work has found stochastic equivalence to be important (Manger *et al.*, 2012). Latent variables can account for third-order dependence patterns within the context of the SRRM. These models assume that relationships between nodes are mediated by a small number ( $K$ ) of node-specific unobserved latent variables.

A number of latent variable approaches have been developed to represent third-order dependencies in relational data, we focus on the latent factor model (LFM). This expands the SSRM

to include an additional term,  $y_{ij}$ ,  $\alpha(u_i, u_j)$ , that captures latent third-order characteristics of a network—where  $u_i$  and  $u_j$  are node-specific latent variables for symmetric data. General definitions for how  $\alpha(u_i, u_j)$  is defined for these latent variable models are shown in Equation 4.

$$(4) \quad \begin{aligned} \alpha(\mathbf{u}_i, \mathbf{u}_j) &= \mathbf{u}_i^\top \Lambda \mathbf{u}_j \\ \mathbf{u}_i &\in \mathbb{R}^K, i \in \{1, \dots, n\}, \text{ where} \\ \Lambda &\equiv K \times K \text{ diagonal matrix} \end{aligned}$$

Many observed networks exhibit varying degrees of stochastic equivalence and homophily. In the latent factor model, each actor has an unobserved vector of characteristics,  $\mathbf{u}_i = \{u_{i,1}, \dots, u_{i,K}\}$ , describing each node's behavior in the network. The probability of a tie between  $i$  and  $j$  depends on the extent to which  $\mathbf{u}_i$  and  $\mathbf{u}_j$  are “similar” and on whether the entries of  $\Lambda$  are greater or less than zero. Similarity in the latent factors,  $\mathbf{u}_i \approx \mathbf{u}_j$ , corresponds to stochastic equivalence; the eigenvalue determines whether the network exhibits positive or negative homophily. Accordingly, this approach can represent both homophily and stochastic equivalence.

In directed (asymmetric) networks there are two sets of latent positions, denoted  $\mathbf{u}$  and  $\mathbf{v}$ , representing the sending and receiving positions separately. Thus, actors in the network have a vector of latent characteristics describing their behavior as a sender, denoted by  $\mathbf{u}$ , and as a receiver,  $\mathbf{v}$ :  $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^K$ . This can alter the probability of an interaction between  $ij$  additively:  $\mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j$ , where  $\mathbf{D}$  is a  $K \times K$  diagonal matrix.

The LFM captures a wide range of interdependence patterns that arise in relational data such as often studied in international relations.

#### 4. SIMULATION STUDY

Here we consider inference for a regression parameter  $\beta$  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 sociomatrix,  $x = \{x_i\} \in \mathbb{R}^n$  is a vector of observed node-specific characteristics, and  $w = \{w_i\} \in \mathbb{R}^n$  is a vector of unobserved node-specific characteristics.

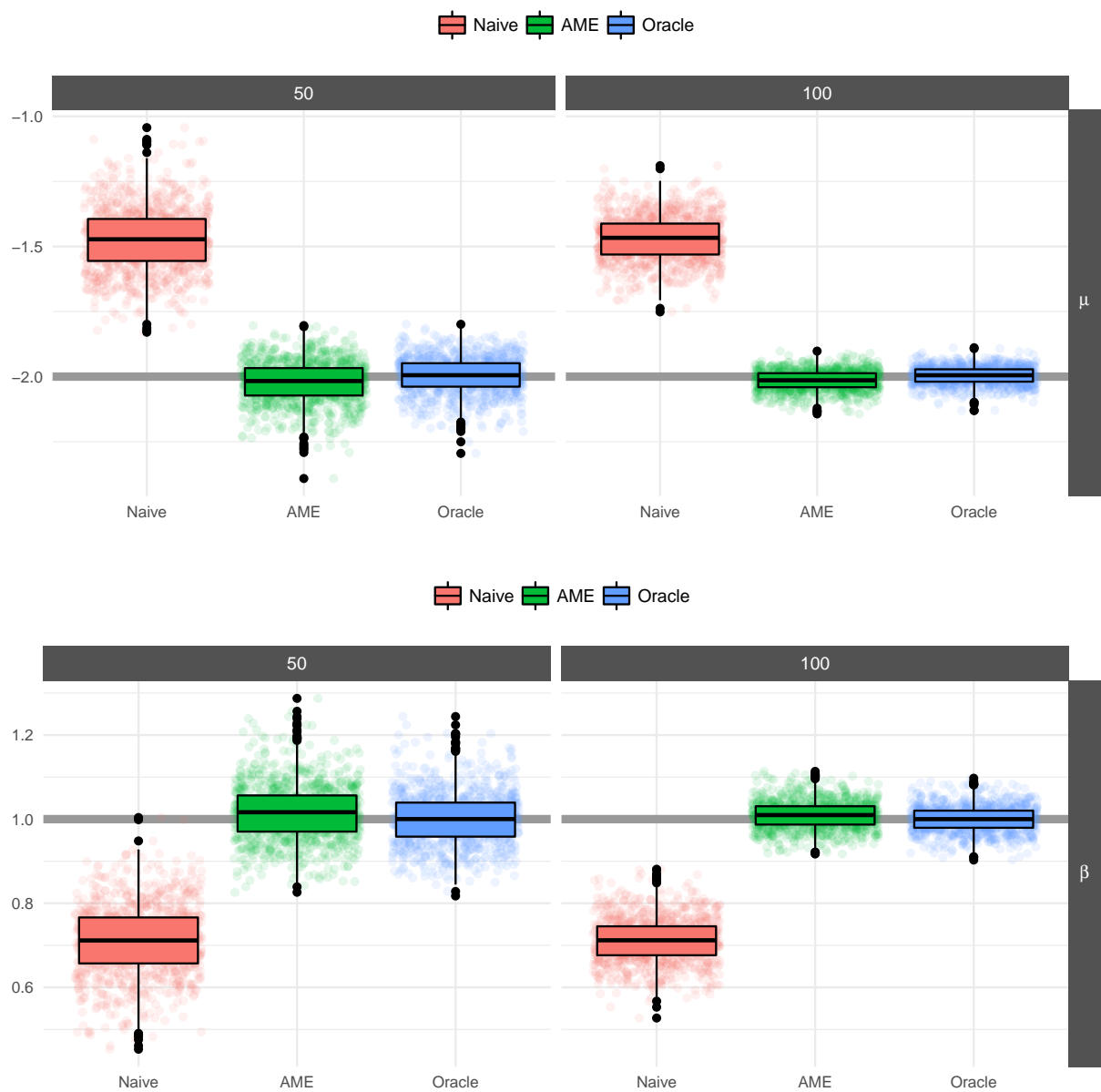
We compare inference for  $\beta$  using three models:

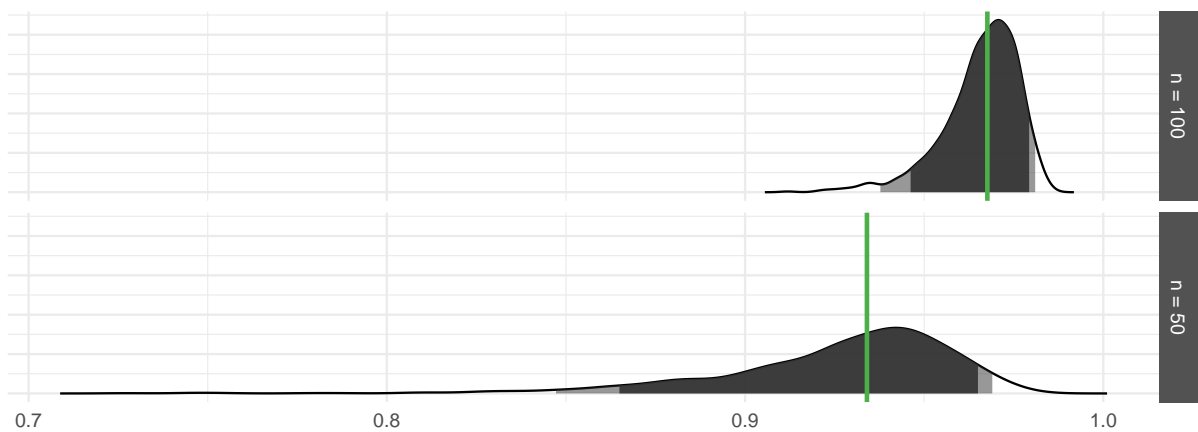
- A naive regression model assuming independent errors;
- A latent factor model;
- An "oracle" regression model that includes both  $x_{i,j} = x_i x_j$  and  $w_{i,j} = w_i w_j$ .

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

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 1.** Bias in parameter when homophily is ignored.

**Figure 2.** Coverage in parameter when homophily is ignored.**Figure 3.** 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.<sup>2</sup> 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.

<sup>2</sup>Without exception this was straightforward to accomplish, thanks to an increasing norm in the social sciences of open data sharing.

**Table 1.** 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<br>in a Network Model? |
|----------------------|--|--|
| Weeks (2012)         | Bosses, Juntas and Strongmen are more Aggressive, Machines are Not       | <b>Fails to Replicate</b>                |
| Reiter & Stam (2003) | Personalist Regimes Attack Democracies, Not Vice Versa                   | Replicates                               |
| McDonald (2004)      | Lower Trade Barriers and Higher Trade Lead to Peace                      | <b>Fails to Replicate</b>                |
| Gibler (2017)        | Power Parity at time of Entry to International System Increases Conflict | <b>Fails to Replicate</b>                |
| Rose (2004)          | WTO Membership Does not Effect Trade                                     | Replicates                               |

**Table 2.** 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  $s \in \{1, \dots, k\}$ , 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 criteria 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 | <b>0.97</b> |
|                      | Area Under PR Curve (AUC-PR)      | 0.00 | <b>0.15</b> |
| Reiter & Stam (2003) | AUC (ROC)                         | 0.92 | <b>0.96</b> |
|                      | AUC (PR)                          | 0.08 | <b>0.15</b> |
| McDonald (2004)      | AUC (ROC)                         | 0.92 | <b>0.99</b> |
|                      | AUC (PR)                          | 0.13 | <b>0.28</b> |
| Gibler (2017)        | AUC (ROC)                         | 0.52 | <b>0.91</b> |
|                      | AUC (PR)                          | 0.00 | <b>0.08</b> |
| Rose (2004)          | Root Mean Squared Error (RMSE)    | 3.23 | <b>1.99</b> |
|                      | Root Median Squared Error (RMDSE) | 2.01 | <b>1.06</b> |

**Table 3.** 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.

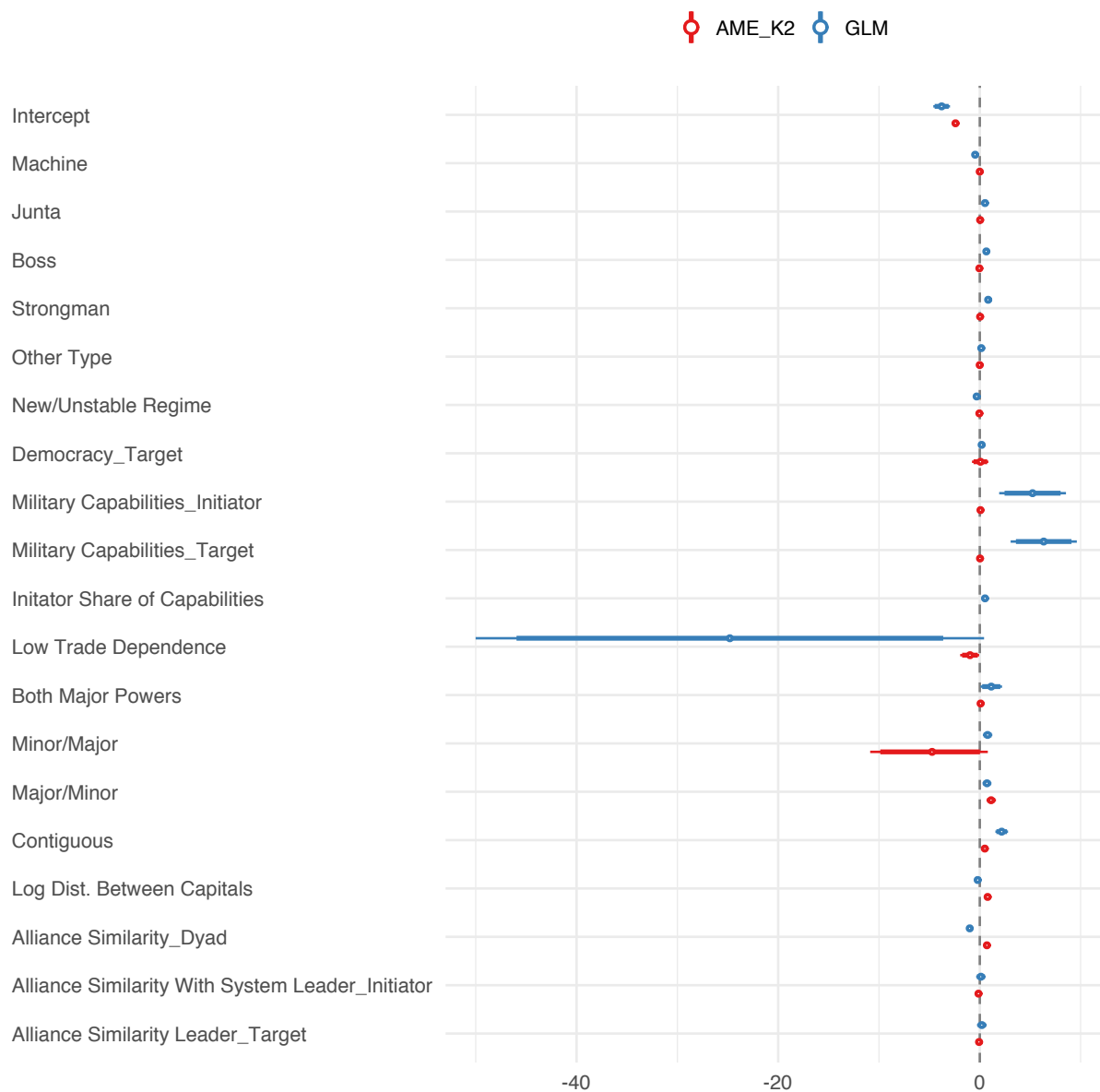
**5.2. Replication of Weeks (2012).** Weeks (2012) examines the influence of domestic institutions on the initiation of military conflicts by autocratic leaders. She argues that in some circumstances autocrats are held accountable for their foreign policy decisions. She adds the nuance that autocratic 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 4 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 6 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 5(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.

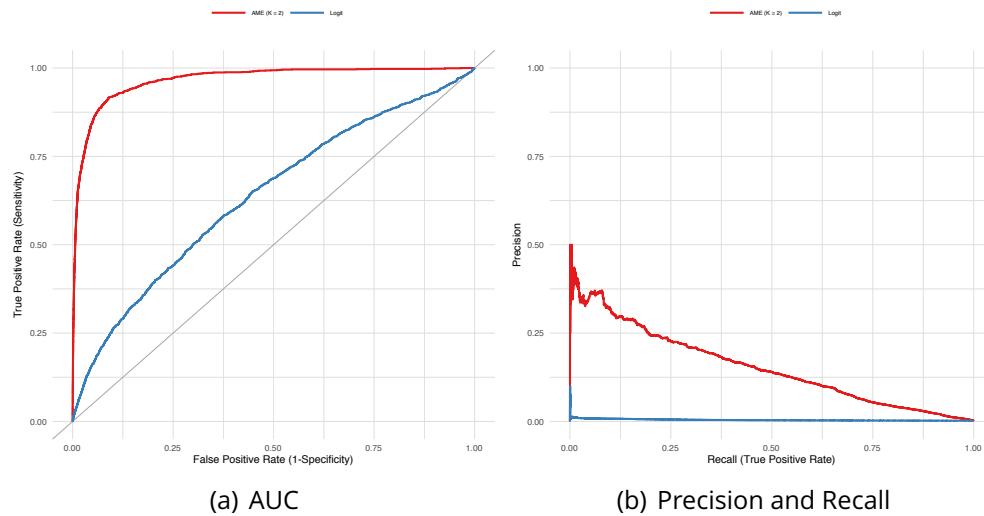
**5.3. Replication of Reiter & Stam (2003).** Reiter & Stam (2003) examine the relationship between democracy, dictatorship and the initiation of militarized disputes. They use directed dyads and find that dyads involving a democratic leader on the one hand and a personalist dictator on the other tend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





**Figure 4.** Coefficient plot of Weeks' (2012) original model (blue) compared to AME model (red).

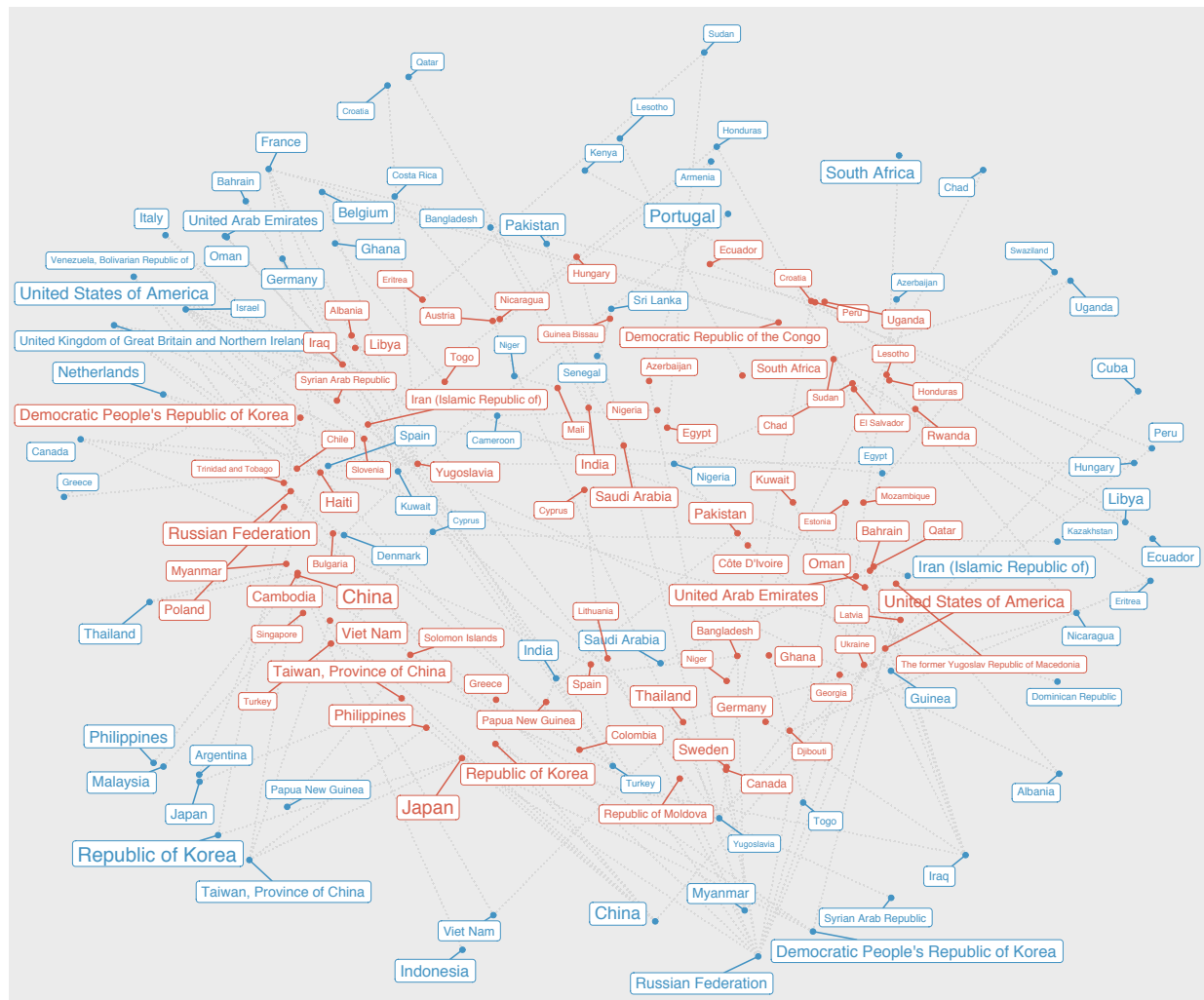
this is not reciprocal. In addition, military regimes and single-party regimes are more prone to initiate disputes with democracies, than the other way around. They use the MID data, but note that "We code a state as having initiated a dispute if it is on 'side A' of a MID, the conventional approach 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



**Figure 5.** Assessments of out-of-sample predictive performance for Weeks (2012) using ROC curves and PR curves.

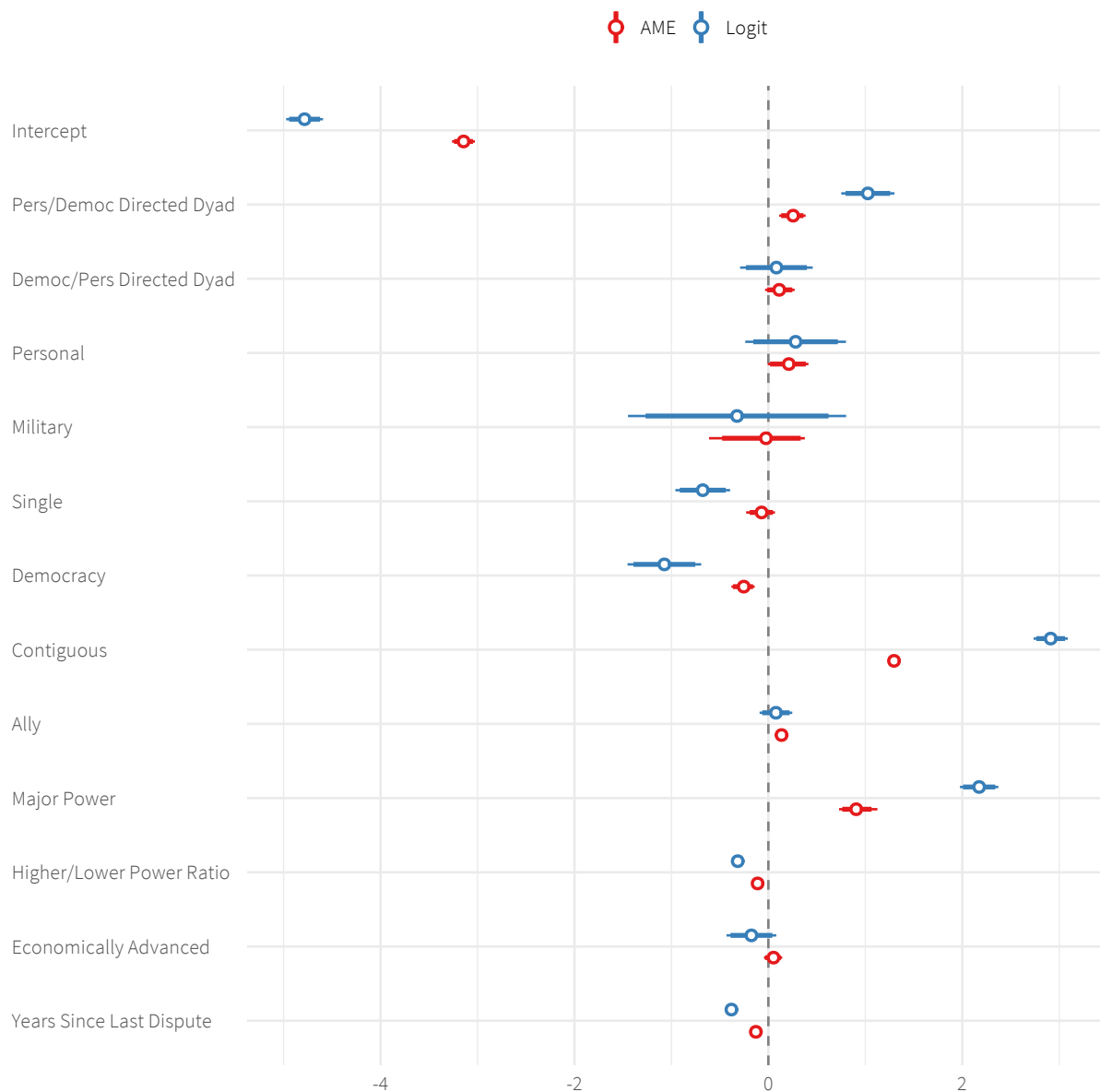
initiators or targets, though the results do not change if we do not code joiners as initiators or targets. ... Though coding initiation will always be difficult, the ‘side A’ variable has been widely used in past conflict scholarship (page 334).” Independent variables are largely taken from an earlier 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 → 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



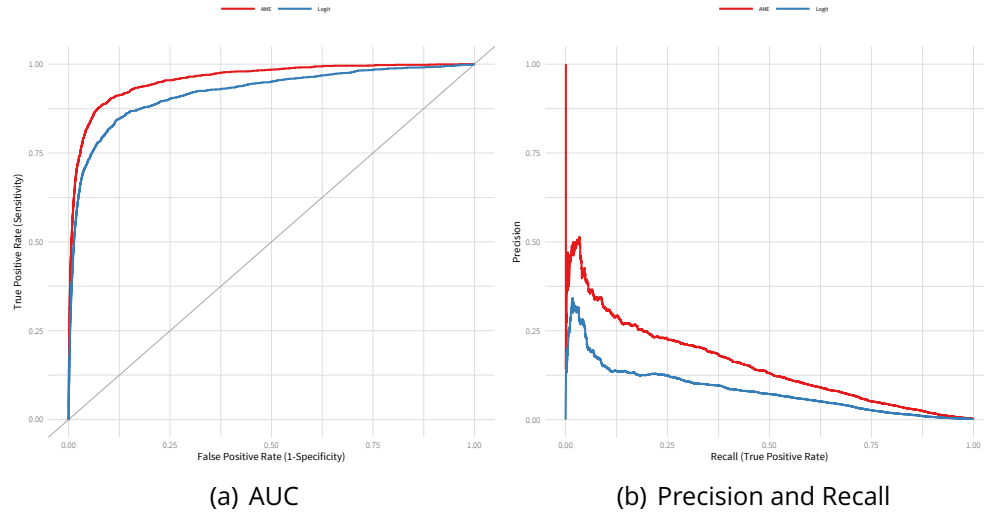
**Figure 6.** Visualization of multiplicative effects for Weeks (2012). Blue represents groups with common sending patterns and red represents groups with common receiving patterns.

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 7. Finally, our modeling approach outperforms the original model by better, and more accurately, predicting MIDs out-of-sample (Figure 8(a) and Figure 8(b)).



**Figure 7.** 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



**Figure 8.** Assessments of out-of-sample predictive performance for Reiter & Stam (2003) using ROC curves and PR curves.

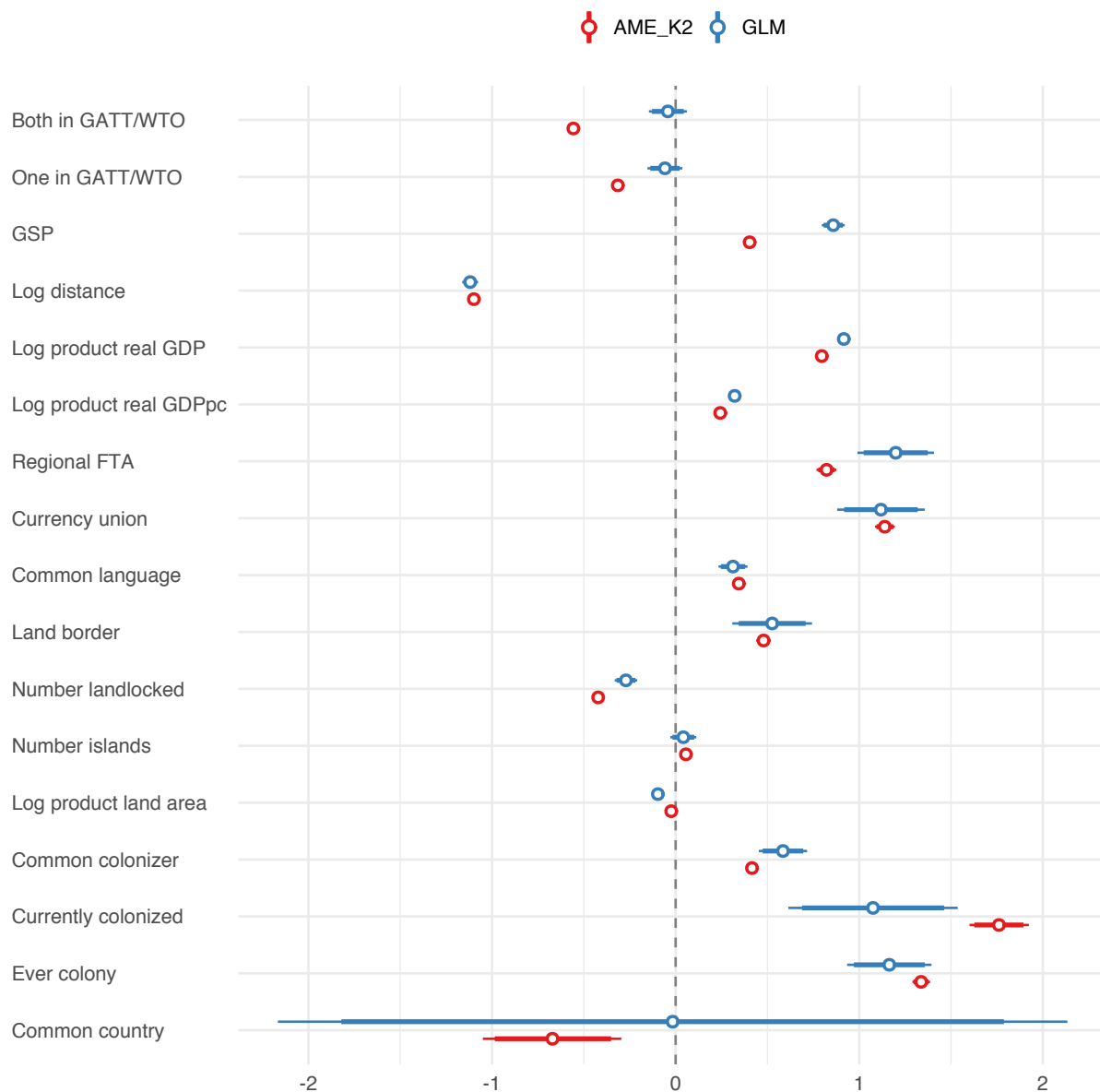
that WTO membership fosters greater cooperation and thereby more trade among its members, the empirics do not bear out such claims. He uses a standard gravity model with dyadic data on bilateral merchandise trade (not services) for 175 countries over a period of five decades. Estimating this model using 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 patterns 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 9, 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 10 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.<sup>3</sup> 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.

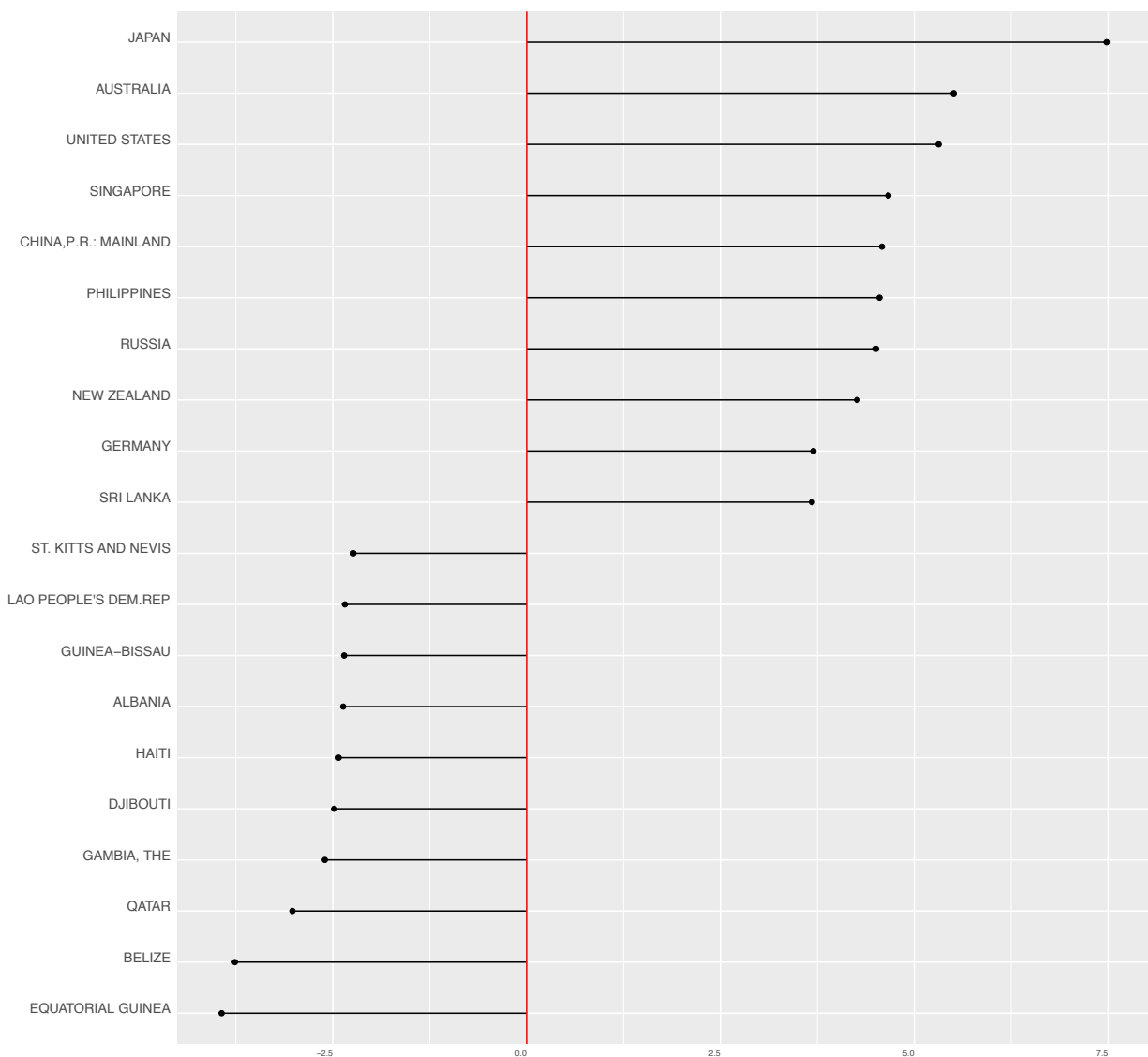
---

<sup>3</sup>Note: Qatar exhibits strongly negative random effects.



**Figure 9.** 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 10.** Nodal Random Effects for AME estimation of Rose (2004).

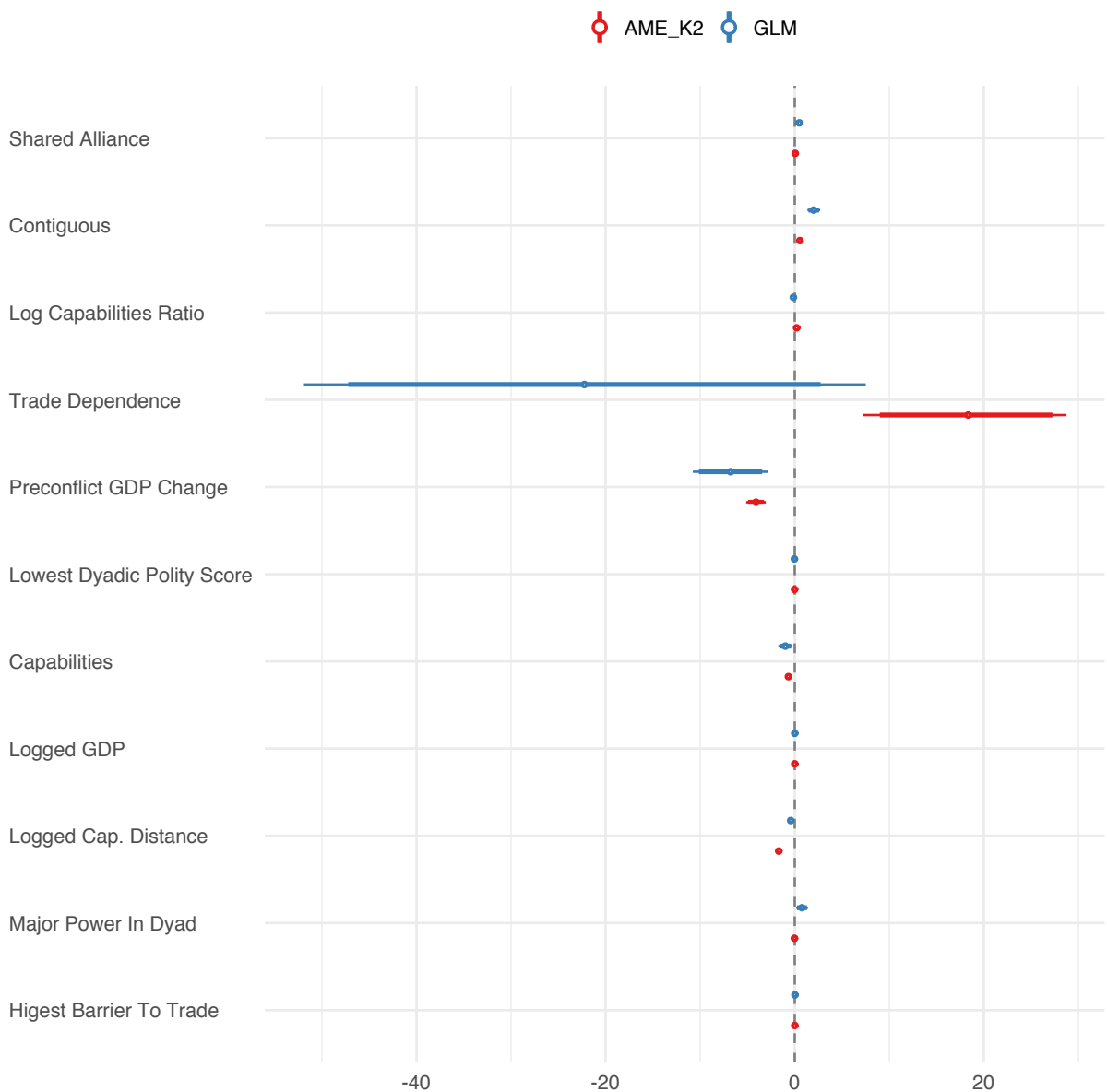
argument—that interdependence between states “makes conflict less likely because of its efficiency over conquest in acquiring resources... (547)” in his overview of underdeveloped hypotheses.

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



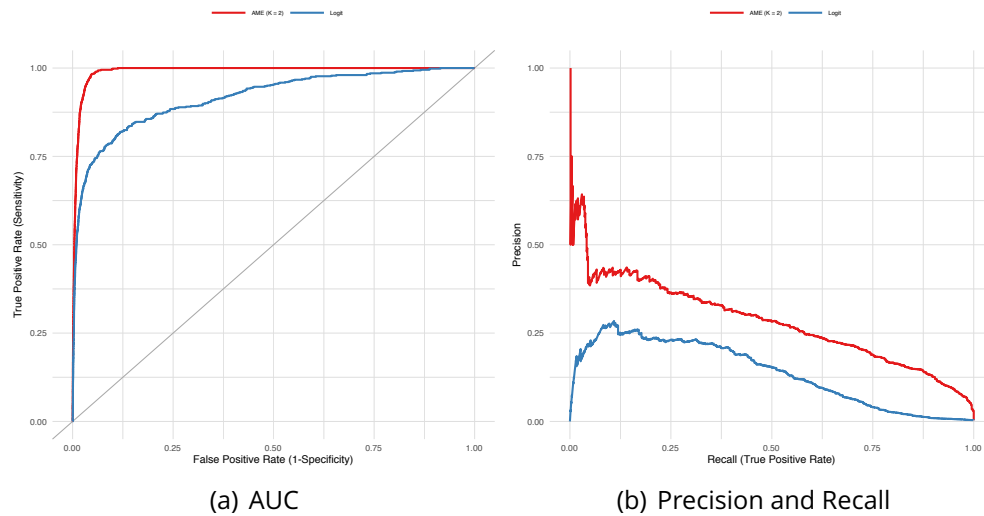
work of classic liberalism and connects free trade to the power of domestic audiences. McDonald (2004) measures free trade in two ways. The first captures the idea that larger protected sectors generate greater societal pressures resulting in pockets of support for war. This protection variable measures 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 11 shows the coefficients for both the original and replicated model and Figure 12 demonstrates the predictive performance of each model. Figure 13 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 11.** 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

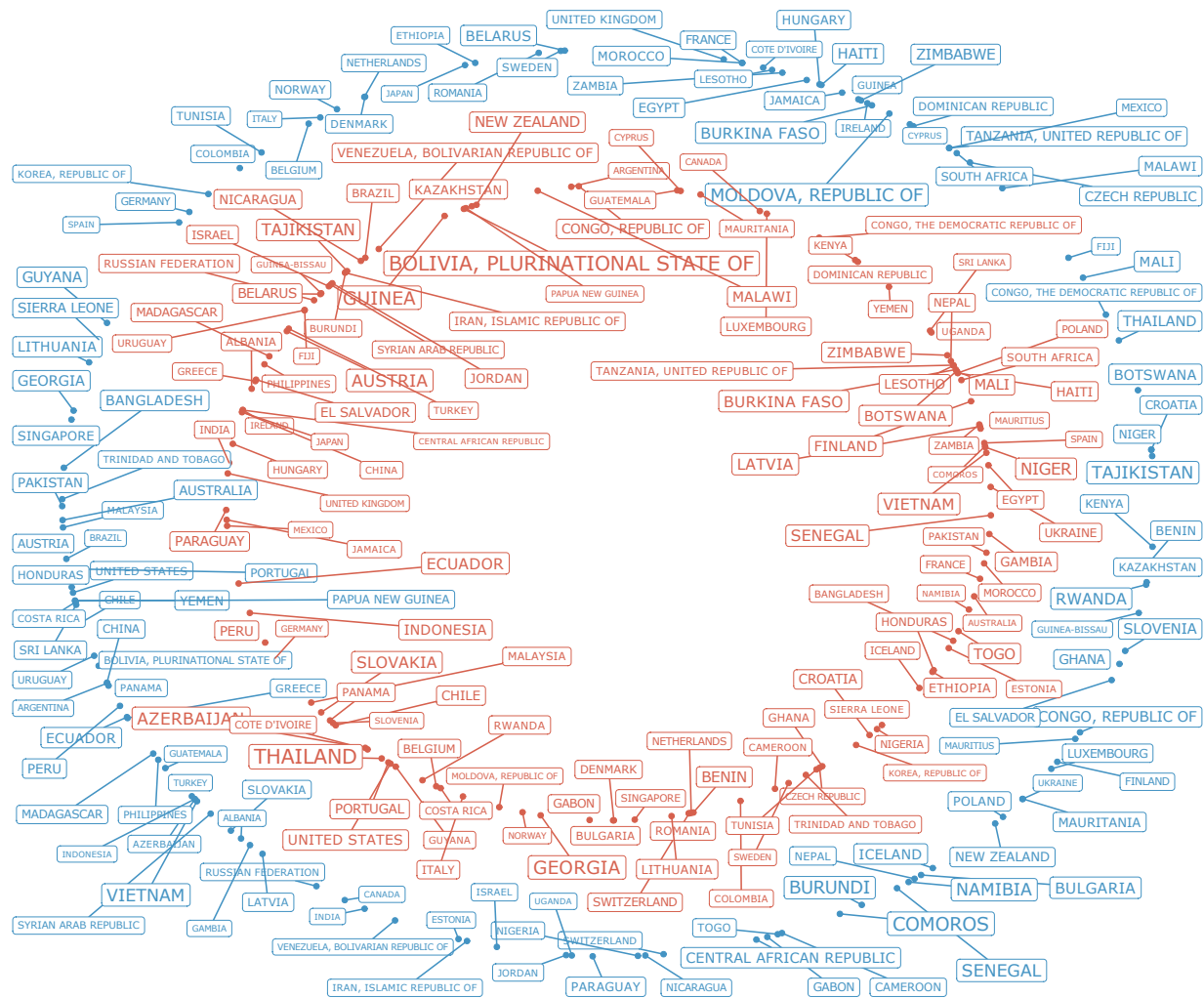


**Figure 12.** Assessments of out-of-sample predictive performance for McDonald (2004) using ROC curves and PR curves.

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 4 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 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 13.** Visualization of multiplicative effects. Blue represents groups with common sending patterns and red represents groups with common receiving patterns.

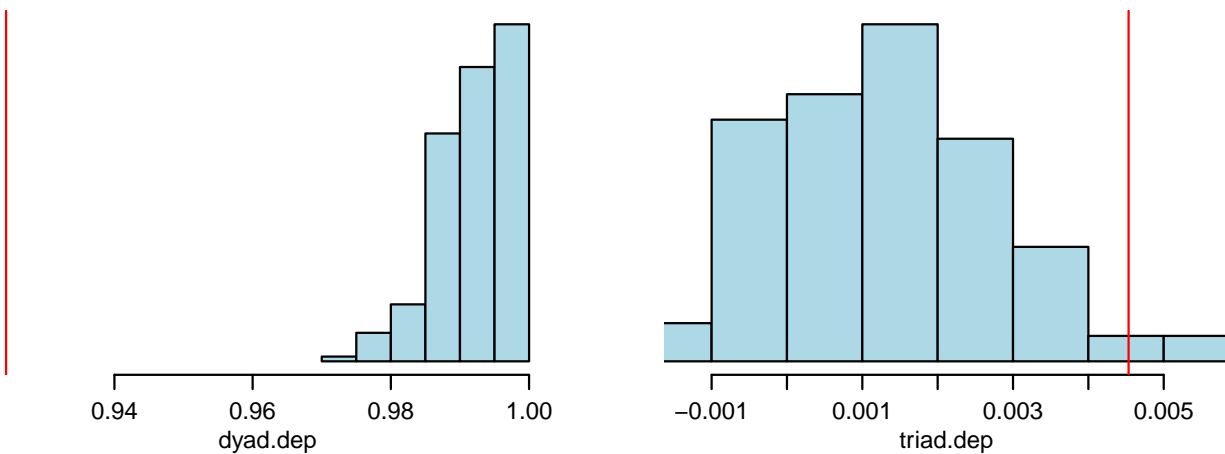
and triadic dependencies which are presented in Figure 14, which further reveal that the assumption of independence among the dyadic data in this study can be strongly rejected. Perhaps most importantly, the main substantive conclusions of the 2017 Gibler study do not seem warranted from the perspective of the results obtained with the `amen` estimation that explicitly models 1<sup>st</sup>-, 2<sup>nd</sup>-, and 3<sup>rd</sup>-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 4.** Comparison of Gibler (2017) Model 6 results with AME results. This was run over a chain of 10,000 iterations.

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

**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 14.** 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.

## REFERENCES

- Aronow, Peter M., Samii, Cyrus, & Assenova, Valentina A. 2015. Cluster-Robust Variance Estimation for Dyadic Data. *Political Analysis*, **23**(4), 564–577.
- 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.*
- Hoff, Peter D. 2015. Multilinear Tensor Regression for Longitudinal Relational Data. *The Annals of Applied Statistics*, **9**(3), 1169–1193.
- Hoff, Peter D., Fosdick, Bailey, Volfovsky, Alex, He, Yanjun, & Minhas, Shahryar. 2017 (July). *Additive and Multiplicative Effects Models for Networks and Relational Data, version 1.4*. Comprehensive R-Archive Network.
- Li, Heng, & Loken, Eric. 2002. A Unified Theory of Statistical Analysis and Inference for Variance Component Models For Dyadic Data. *Statistica Sinica*, **12**(2), 519–535.
- Manger, Mark S., Pickup, Mark A., & Snijders, Tom A.B. 2012. A Hierarchy of Preferences: A Longitudinal Network Analysis Approach to PTA Formation. *Journal of Conflict Resolution*, **56**(5), 852–877.
- 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.
- 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.
- 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.
- Warner, Rebecca, Kenny, David, & Stoto, Michael. 1979. A New Round Robin Analysis of Variance for Social Interaction Data. *Journal of Personality and Social Psychology*, **37**, 1742–1757.
- 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.
- Wong, George Y. 1982. Round robin analysis of variance via maximum likelihood. *Journal of the American Statistical Association*, **77**(380), 714–724.
- Zinnes, Dina A. 1967. An Analytical Study of the Balance of Power Theories. *Journal of Peace Research*, **3**, 270–288.



ALL ABOUT US

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