

TAKING DYADS SERIOUSLY

ABSTRACT. International relations scholarship concerns dyads, yet standard modeling approaches fail to adequately capture the data generating process behind dyadic events and processes. As a result, they suffer from biased coefficients and poorly calibrated standard errors. We show how a regression-based approach, the Additive and Multiplicative Effects (AME) model, can be used to account for the inherent dependencies in dyadic data and glean substantive sights in the interrelations between actors. First, we conduct a simulation to highlight how the model captures dependencies and show that accounting for these processes improves our ability to conduct inference on dyadic data. Second, we compare the AME model to approaches used in three prominent studies from recent international relations scholarship. For each study, we find that compared to AME, the modeling approach used performs notably worse at capturing the data generating process. Further, conventional methods misstate the effect of key variables and the uncertainty in these effects. Finally, AME outperforms standard approaches in terms of out-of-sample fit. In sum, our work shows the consequences of failing to take the dependencies inherent to dyadic data seriously. Most importantly, by better modeling the data generating process underlying political phenomena, the AME framework improves scholars' ability to conduct inferential analyses on dyadic data.

Word count: 9,177

1. INTRODUCTION

Aronow et al. (2015) estimate that during the period 2010 to 2015, over sixty articles were published in the *American Political Science Review*, *American Journal of Political Science*, and *International Organization* using dyadic data.¹ Most of these studies use a generalized linear model (GLM). However, this approach to studying dyadic data increases the chance of faulty inferences by assuming data are conditionally independent and identically distributed (iid). Most standard approaches assume that the problems raised by non-iid relational data can be addressed by recalculating the standard errors of estimated parameters to reflect the potential clustering of cases. In practice, such strategies rarely work because they do not directly address the fundamental data generating process. This is important to consider since the inferential problems caused by non-iid observations affect more than just diagonals of the variance co-variance matrix (Beck, 2012; Franzese and Hays, 2007; King and Roberts, 2014).

In this article, we discuss and evaluate how well the Additive and Multiplicative Effects (AME) model can be used to account for the interdependencies underlying the data generating process of dyadic structures (Hoff, 2008; Minhas et al., 2018). We focus on three types of interdependencies that can complicate dyadic analyses. First, dependencies may arise within a set of dyads if a particular actor is more likely to send or receive actions such as conflict.² Additionally, if the event of interest has a clear sender and receiver, we are likely to observe dependencies within a dyad; for example, if a rebel group initiates a conflict against a government, the government will likely reciprocate that behavior. We capture these effects, often referred to as first- and second-order dependencies, respectively, within the additive effects portion of the model. Third-order dependencies capture relationships of transitivity, balance, and clusterability between different dyads. For example, we can only understand why Poland was involved in a dyadic conflict with Iraq in 2003 if we understand that the United States invaded Iraq in 2003 and that Poland often participates in US-led coalitions. The multiplicative effects capture these sorts of dependencies, specifically, those that result because the specified model has not accounted for a latent set of shared attributes that affect actors' probability of interacting with one another.

¹In 2017, *International Studies Quarterly* published a special issue on Dyadic Research Designs along with an online symposium to discuss the papers.

²In the case of undirected data where there is no clear sender or receiver, it is still essential to take into account the variance in how active actors are in the system.

We begin with a discussion of these dependencies and an introduction to the AME model. Next, we conduct a simulation study to show how the AME approach can recover unbiased and well-calibrated regression coefficients in the context of dyadic data. Last, to highlight the utility of this approach, we apply the AME model to three recent studies in the international relations (IR) literature. Our comparison reveals that by accounting for observational dependence, AME produces more precise estimates and better-calibrated confidence intervals for key variables. Consequently, AME generates results that, at times, differ from those found in the original study as well as from the broader literature. Moreover, we demonstrate that the AME latent factor approach offers substantive insights that are often occluded by ignoring the interdependencies found in the relational data widely employed in IR. Finally, we show that for each replication our network-based approach provides substantially more accurate out-of-sample predictions than the models used in the original studies.

The AME approach advances statistical analysis of dyadic data by accounting for observational dependence while allowing scholars to test the substantive effect of variables of interest. Thus, the AME allows scholars to achieve a two-fold goal: to continue to generate meaningful, substantive insights about political phenomena without abandoning a regression based approach, while at the same time accounting for the data generating processes behind such events of interest. Most importantly, the AME approach concentrates on the relational aspect of the field of international relations through a statistical framework that is familiar to most scholars.

2. DEPENDENCIES IN DYADIC DATA

Scholars working with dyadic data typically begin by stacking observations associated with each dyad on top of one another. This makes sense if each observation is independent of the others. For example, in this design, a conflict initiated from the United States against Japan is assumed to be independent of any conflictual 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. While most scholars begin with the assumption that each dyadic interaction takes place in isolation of the others, we know this assumption to be false both in theory and in practice. Relational data comes with an explicit structure that generally leads to particular types of dependencies. The importance of accounting for

the underlying structure of our data has been a lesson well understood, at least when it comes to time-series cross-sectional data (TSCS) within political science (Beck and Katz, 1995; Beck et al., 1998). As a result, it is now standard practice to take explicit steps to account for the complex data structures that emerge in TSCS applications and the unobserved heterogeneity that they cause.

To uncover the underlying structure of relational data, it is helpful to restructure dyadic data in the form of a matrix—often referred to as an adjacency matrix—as shown in Figure 1. Rows designate the senders of an event and columns the receivers. The cross-sections in this matrix represent the actions that were sent by an actor in the row to those designated in the columns. Thus y_{ij} designates an action y , such as a conflictual event or trade flow, that is sent from actor i to actor j . In many applications, scholars are interested in studying undirected (i.e., symmetric) outcomes in which there is no clear sender or receiver, these type of outcomes still can, and we argue should, be studied using the type of framework we discuss below.

Using the structure of an adjacency matrix, Figure 1 visualizes the types of first- and second-order dependencies that can complicate the analysis of relational data in traditional GLMs. The adjacency matrix on the top left highlights a particular row to illustrate that these values may be more similar to each other than other values because each has a common sender. Interactions involving a common sender also manifest heterogeneity in how active actors are across the network when compared to each other. 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 that are much more active than others (Barabási and Réka, 1999). For example, in an analysis of international trade certain countries (e.g., China) export much larger volumes than other countries for a variety of structural, contextual, and idiosyncratic reasons. Unless one is able to develop a model that can account for the variety of explanations that may play a role in determining why a particular actor is more active than others, parameter estimates from standard statistical models will be biased.³

For similar reasons one also needs to take into account the dependence between observations that share a common receiver. The bottom-left panel in Figure 1 illustrates that sender and receiver type dependencies can also blend together. Specifically, actors who are more likely to send

³In an undirected setting instead of studying sender and receiver heterogeneity, we would just be concerned with actor heterogeneity in general.



Figure 1. Nodal and dyadic dependencies in relational data.

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 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).

For most relational data, however, dependencies do not simply manifest at the nodal or dyadic 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 set of latent attributes between actors that affects their probability of interacting with one another (Zinnes, 1967; Wasserman and Faust, 1994). In Figure 2 we provide a visualization of a simulated relational dataset wherein the nodes designate actors and edges between the nodes indicate that

an interaction between the two took place. To highlight third-order dependence patterns, nodes with similar latent attributes are colored similarly.

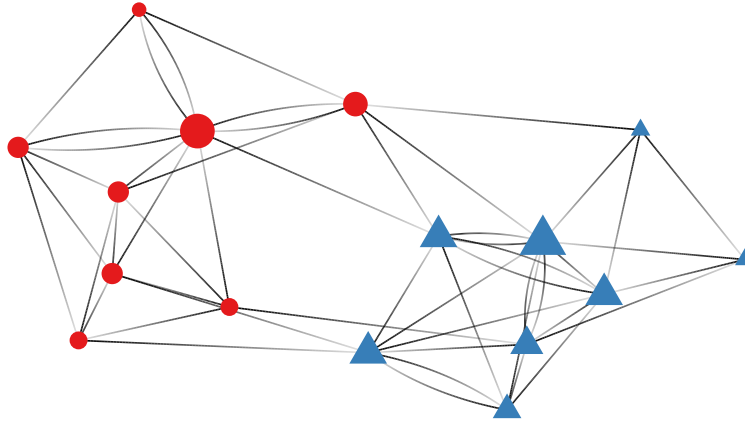


Figure 2. Higher-order dependence patterns in a network.

The visualization illustrates that actors belonging to the same group have a higher likelihood of interacting with each other, whereas interactions across groups are rarer. A prominent example of a network with this type of structure is discussed by Adamic and Glance (2005), who visualize linkages between political blogs preceding the 2004 United States Presidential Election. Adamic and Glance find that the degree of interaction between right- and left-leaning blogs is minimal and that most blogs are linked to others that are politically similar. This showcases the types of higher-order dependencies that can emerge in relational data. First, the fact that interactions are determined by a shared attribute, in this case political ideology, is an example of *homophily*. Homophily explains 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 meso-scopic features that emerge in relational data is community structure (Mucha et al., 2010), which is often formalized through the concept of stochastic equivalence (Anderson et al., 1992). Stochastic equivalence refers to a type of pattern in which actors can be divided into groups such that members of the same group have similar patterns of relationships. In the example above, each of the left leaning blogs would be considered stochastically equivalent to one another because any given left-leaning blog is more likely to interact with a blog of a similar political position and less likely to interact with one of a divergent political position.

These types of patterns frequently emerge in IR contexts.⁴ For example, a perennial finding in the interstate trade literature emphasizes the role that geography plays in determining trade flows. Geographic proximity in the network context is an example of homophily—a shared attribute between actors that corresponds to a greater likelihood of the event of interest taking place. Alternatively, in the interstate conflict literature, we may find that actors who are each a member of a particular (formal or informal) alliance are likely to act similarly in the conflict network. Specifically, they will tend to initiate conflictual events with actors that their fellow alliance members initiate conflict with, and they will be unlikely to initiate conflict with members of their alliance—an example of stochastic equivalence. In both these examples, we are able to explicitly parameterize the attribute that might explain the emergence of higher order dependence patterns. While sometimes the conditions driving these patterns, such as geography, are easy to identify, at other times it can be difficult to describe exactly why higher order dependence patterns in networks may develop.

3. ADDITIVE AND MULTIPLICATIVE EFFECT MODELS FOR NETWORKS

To account for the dependencies that are prevalent in dyadic data, we turn to the AME model. 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 the social sciences. It accounts for nodal and dyadic dependence patterns, as well as higher-order dependencies such as homophily and stochastic equivalence.⁵ The AME model combines the social relations regression model (SRRM) to account for nodal and dyadic dependencies and the latent factor model (LFM) for third-order dependencies. For details on the SRRM see Li and Loken (2002); Hoff (2005); Dorff and Minhas (2017).⁶ The AME model is specified as follows:

⁴For example, see: Manger et al. (2012); Kinne (2013); Chyzh (2016).

⁵Minhas et al. (2018) detail how this framework contrasts with alternative network-based approaches.

⁶An earlier version of the LFM used in AME is presented as the general bilinear mixed effects (GBME) model in Hoff (2005). The GBME model is more limited in the types of dependence patterns that it can capture due to the formulation of the matrix decomposition procedure.

$$\begin{aligned}
y_{ij} &= f(\theta_{ij}), \text{ where} \\
\theta_{ij} &= \boldsymbol{\beta}_d^\top \mathbf{X}_{ij} + \boldsymbol{\beta}_s^\top \mathbf{X}_i + \boldsymbol{\beta}_r^\top \mathbf{X}_j && \text{(Exogenous parameters)} \\
(1) \quad &+ a_i + b_j + \epsilon_{ij} && \text{(SRRM parameters)} \\
&+ \mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j && \text{(LFM parameters)}
\end{aligned}$$

where $y_{ij,t}$ captures the interaction between actor i (the sender) and j (the receiver) at time t . We model a latent variable, θ_{ij} , first using a set of exogenous dyadic ($\boldsymbol{\beta}_d^\top \mathbf{X}_{ij}$), sender ($\boldsymbol{\beta}_s^\top \mathbf{X}_i$), and receiver covariates ($\boldsymbol{\beta}_r^\top \mathbf{X}_j$). f is typically a mapping function, and can be one that applies to dichotomous, ordinal, or continuous distributions.

Next, to account for the dependencies that emerge in dyadic data that may complicate inference on the parameter associated with exogenous covariates, we add parameters from the SRRM and LFM. a_i and b_j in Equation 1 represent sender and receiver random effects incorporated from the SRRM framework:

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

The sender and receiver random effects are modeled jointly from a multivariate normal distribution to account for correlation in how active an actor is in sending and receiving ties. Heterogeneity in the sender and receiver effects is captured by σ_a^2 and σ_b^2 , respectively, and σ_{ab} describes the linear relationship between these two effects (i.e., whether actors who send [receive] a lot of ties also receive [send] a lot of ties). Beyond these first-order dependencies, second-order dependencies are described by σ_ϵ^2 and a within dyad correlation, or reciprocity, parameter ρ .

The LFM contribution to the AME is in the multiplicative term: $\mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j = \sum_{k \in K} d_k u_{ik} v_{jk}$. K denotes the dimensions of the latent space. The construction of the LFM here is actually quite similar to the recommender systems that companies like Amazon and Netflix use to model customer behavior (Resnick and Varian, 1997; Bennett and Lanning, 2007). This model posits a latent vector of characteristics \mathbf{u}_i and \mathbf{v}_j for each sender i and receiver j . The similarity or dissimilarity of these vectors will then influence the likelihood of activity, and provides a representation of third-order interdependencies. The LFM parameters are estimated by a process similar to computing the singular value decomposition (SVD) of the observed network. When computing the SVD we factorize our observed network into the product of three matrices: \mathbf{U} , \mathbf{D} , and \mathbf{V} . This provides us with a low-dimensional representation of our original network.⁷ Values in \mathbf{U} provide a representation of how stochastically equivalent actors are as senders in a network or, for example, how similar actors are in terms of who they initiate conflict with. $\hat{\mathbf{u}}_i \approx \hat{\mathbf{u}}_j$ would indicate that actor i and j initiate battles with similar third actors. \mathbf{V} provide a similar representation but from the perspective of how similar actors are as receivers. The values in \mathbf{D} , a diagonal matrix, represent levels of homophily in the network.⁸ Note that this model easily generalizes to the case, common in IR, where interactions are undirected (for example the presence of conflict or a bilateral investment treaty). In the case of the SRRM, ρ is constrained to be one and instead of separate sender and receiver random effects a single actor random effect is utilized. For the LFM, an eigen-decomposition scheme is used to capture higher-order dependence patterns. In the application section, we show the applicability of the AME approach to both directed and undirected dyadic data.⁹

Non-iid observations in relational data result from the fact that there is a complex structure underlying the dyadic events or processes we observe, and that accounting for this structure is necessary if we are to adequately represent the data generating process. If one can specify each of the nodal, dyadic, and triadic set of attributes that influence interactions then the conditional independence assumption underlying standard approaches will be satisfied. However, it is rarely

⁷The dimensions of \mathbf{U} and \mathbf{V} are $n \times K$ and \mathbf{D} is a $K \times K$ diagonal matrix.

⁸Unlike traditional SVD, in the latent factor model the singular values are not restricted to be positive, thus allowing us to account for both positive and negative homophily.

⁹Parameter estimation in the AME takes place within the context of a Gibbs sampler in which we iteratively sample from the posterior distribution of the full conditionals for each parameter. The algorithm is provided in the Appendix.

the case that this is possible even for TSCS data and thus modeling decisions must account for underlying structure. Failing to do so in either TSCS or dyadic data leads to a number of well-known challenges: a) biased estimates of the effect of independent variables, b) uncalibrated confidence intervals, and c) poor predictive performance. Additionally, by ignoring these potential interdependencies, we often ignore substantively interesting features of the phenomena we investigate. 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?

4. SIMULATION STUDY

Through a simulation study, we highlight the utility of AME as an inferential tool for dyadic analysis.¹⁰ Most scholars working with dyadic data are primarily concerned with understanding the effect of a particular independent variable on a dyadic dependent variable. The goal of our simulation is to assess how well AME can provide unbiased and well-calibrated estimates of coefficient parameters in the presence of unobserved dependencies. Assume that the true data-generating process for a particular Y is given by:¹¹

$$(3) \quad y_{i,j} \sim \mu + \beta x_{i,j} + \gamma w_{i,j} + \epsilon_{i,j}$$

where $Y = \{y_{i,j}\} \in \mathbb{R}^{n \times n}$ is an observed sociomatrix, $X = \{x_{i,j}\} \in \mathbb{R}^{n \times n}$ is a matrix of observed dyad-specific characteristics, and $W = \{w_{i,j}\} \in \mathbb{R}^{n \times n}$ is a matrix of unobserved dyad-specific characteristics. Y can be thought of as a dyadic dependent variable, X and W are both dyadic covariates that are a part of the data-generating process for Y , but W is not observed. We compare inference for μ and β —the latter parameter would be of primary concern for applied scholars—using three models:

¹⁰Alternative network based approaches for dyadic data are exponential random graph models (ERGMs) and the related stochastic actor oriented model (SAOM). While both these models have led to numerous contributions to a variety of literatures, the applicability of these approaches may be limited to certain types of networks and individual level characteristics. Specifically, Block et al. (2017) note that these types of models may not be appropriate in situations where network and behavioral data depend on unobserved latent variables, which is explicitly the focus of our analysis here.

¹¹Note that with the data generating process specified in Equation 3 edge values in Y will be normally distributed.

-
- the “standard” international relations approach estimated through a typical generalized linear model;
 - the AME approach outlined in the previous section with a unidimensional latent factor space ($K = 1$);¹²
 - and an “oracle” regression model that assumes we have measured all sources of dependencies and thus includes both $x_{i,j}$ and $w_{i,j}$.

The first model corresponds to the “standard” approach in which little is explicitly done to account for dependencies in dyadic data. In the second model, we use the AME framework described in the previous section. For both the first and second models, we are simply estimating a linear model of X on Y , and assessing the extent to which inference on the regression parameters are complicated by the presence of unobserved dependencies, W . In the last model, we provide an illustration of the ideal case in which we have observed and measured W and include it in our specification for Y . The oracle case provides an important benchmark for the standard and AME approaches.

For the simulation we set the true value of μ (the intercept term) to -2 and β (the effect of X on Y) to 1.¹³ We conduct two sets of simulations, one in which the number of actors in the network is set to 50 and the other at 100. In total, we run 1,000 simulations where we begin by simulating Y from the specification given in Equation 3 and then for each simulated Y we estimate a standard, AME, and oracle model.

We first compare the performance of the models in terms of how well they estimate the true values of μ and β in Figure 3 μ and β estimates from the simulations for the three models. The panels on the left show the results for when the number of actors is set to 50 and on the right for 100 actors. The top pair of panels represents the estimates for μ while the bottom pair do the same for β . In each case, we find that the estimates for μ and β produced by the standard approach are notably off from their true values. On the other hand, the AME model performs just as well as the oracle at estimating the true parameter values.

¹²Results with higher values of K are similar.

¹³The value of γ is also set to 1, which corresponds to an example where the W character is associated with homophily.

Figure 3. Regression parameter estimates for the standard, AME, and oracle models from 1,000 simulations. Summary statistics are presented through a traditional box plot, and the estimates from each simulation are visualized as well as points.



Next, we estimate the 95% confidence interval for the three models in each of the simulations and estimate the proportion of times that the true value fell within those intervals. The results are summarized in Figure 4, and again we see that the AME approach performs as well as the oracle, while the standard approach performs poorly by comparison. The implication of the results presented in Figures 3 and 4 is that standard approaches can often fail at estimating parameter values and conducting inferential tasks in the presence of unobserved dependencies. The AME approach by comparison can be used as a tool for scholars working with dyadic data to still estimate the true effects of their main variables of interest, while accounting for dependencies that do often emerge in dyadic data.

Moreover, the AME approach allows scholars to better understand what parameters their model may be missing. In the case of the simulation here, W is set as an unobserved dyadic covariate that has a homophilous effect on Y . The effect of W is homophilous within this framework because it is a dyadic attribute involving i and j that positively affects the degree to which actors interact with one another, i.e., y_{ij} . This type of unobserved dependency will be captured through the multiplicative effects portion of the model, $\mathbf{U}^T \mathbf{D} \mathbf{V}$. To estimate how well the model performs, we

Figure 4. Proportion of times the true value fell within the estimated 95% confidence interval for the standard, AME, and oracle models from 1,000 simulations.



recover the multiplicative effects term for each simulation and calculate the correlation between it and the unobserved dependency, W .¹⁴ We visualize the distribution of the correlations from each of the 1,000 simulations in Figure 5 for the case where the number of actors is set to 100 (top pair of panels) and 50. Additionally, we calculate the median across the correlations and display the result using a vertical line. For both $n = 50$ and $n = 100$, we find that the multiplicative effects perform very well in capturing the unobserved dependency, which indicates that the AME does not simply capture noise but also works as a tool to estimate unobserved structure.

The simulation shows that beyond obtaining less biased and better-calibrated parameter estimates, a key benefit of the AME framework is to directly estimate unobserved dependencies through the random effects structure of the model. Scholars can use this framework in an iterative fashion: beginning with an estimated model, they can then empirically study the structure of the random effects to assess whether there are unobserved covariates that they want to include in their model. At the very least, this simulation demonstrates that the AME model can help conduct inference and estimate the true effect of independent variables that scholars operationalize to test theoretical propositions.

¹⁴Specifically, since both the multiplicative effects term and W are continuous dyadic variables, we calculate the Pearson correlation coefficient.

Figure 5. Distribution of correlation between missing variable and multiplicative random effect in AME across the 1,000 simulations. Vertical line through the distribution represents the median value across the simulations.



5. APPLICATIONS WITH AME

We apply AME to three recent IR studies: Reiter and Stam (2003); Weeks (2012); Gibler (2017). Each of these studies use relational data of state interactions and propose both dyadic, monadic, and structural explanations for the behavior of actors in the system. We demonstrate the capabilities of AME with reference to existing studies in order to highlight several features of our network based approach. First, we show that simply by using the AME framework scholars can better model the data generating process behind events of interest. Second, the results of AME estimation are interpretable alongside results using standard approaches. Third, through using this approach we can also quantify the degree to which first-, second-, and third-order dependencies are present.

We obtained the data for these studies from their replication archives.¹⁵ The chosen studies have each gained over 100 or more citations and were recently published.¹⁶ Each of these pieces is well-known and posited an hypothesis in which interdependencies are consequential. Reflecting the dominant approach in the literature, the studies tested their hypothesis by employing some form of a generalized linear model. Table 1 provides descriptive information for the studies that we replicated.

¹⁵Without exception the data was easy to retrieve thanks to the authors' transparency and an increasing norm in the social sciences of open data sharing.

¹⁶Note that we chose papers with at least 100 citations as an indicator of influence within the discipline. Of course, by the criteria of influence many other papers could have been considered for the replication.

Table 1. Descriptive information about the replicated studies.

	Model	Date Range	N. Actors	Dyads Type	Clustering $\sigma_{\hat{\beta}}$
Reiter & Stam (2003)	Logit	1945–1995	193	Directed	Robust
Weeks (2012)	Logit	1946–1999	197	Directed	Robust
Gibler (2017)	Logit	1816–2008	193	Undirected	Robust

For each of the studies listed above we replicated the authors' key model using their original estimation procedure, a generalized linear model.¹⁷ Next, utilizing the same covariate specifications, we estimate the models with the AME framework.¹⁸ For Reiter and Stam (2003); Weeks (2012), we utilized a version of AME that accounts for the directed nature of the data and for Gibler (2017) we used an undirected version. In the directed formulation, separate random effects are used for senders and receivers in both the additive and multiplicative portions of the model.

A key claim that we have made is that by accounting for dependencies inherent to relational data we can better capture the data generating process behind events of interest. To assess whether or not the AME approach successfully does this, we turn to an out-of-sample cross validation strategy. An out-of-sample approach is essential since relying on in-sample procedures would enable models with more parameters, such as AME, to simply overfit the data. The cross validation procedure is executed as follows. For each study we randomly divide the data 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 estimate model parameters with $\{y_{ij,t} : s_{ij,t} \neq s\}$, the data not in set s , 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 a subset of the data that does not include $y_{ij,t}$. Next we conduct a series of analyses to discern whether or not the AME model provides any benefit for each study. These analyses are summarized in Figure 6. The left-most plot in each of the panels evaluates performance using Receiver Operating Characteristic (ROC) curves. Models that have a better fit according to this metric will follow the upper-left corner border of the ROC space. In addition to ROC curves we also use separation plots (Greenhill et al., 2011). Separation plots visualize each of the

¹⁷Replicating the key models from each study was straightforward to do because of the authors' assembled replication scripts.

¹⁸In estimating AME, we show results when setting $K = 2$. Results with alternative values of K are similar.

observations in the dataset according to their predicted probability. In this graph, the shaded panels correspond to the occurrence of an event. Darker panels are events, lighter panels are non events. If a model performs well, then the events that actually occur would stack to the far right of the graph, where the predicted probabilities generated by the model are highest. The right-most plot in each of the panels evaluates performance using precision-recall (PR) curves. PR curves are useful in situations where correctly predicting events is more interesting than simply predicting non-events (Davis and Goadrich, 2006). This is especially relevant in the context of our applications here, as they each are trying to model conflict within dyads, which is an infrequent occurrence.

For each of the replications, we find that the AME approach substantially outperforms the original models in terms of out-of-sample predictive performance. This indicates that switching to the AME framework—even when using the same covariate specification as the original studies—enables scholars to better represent the data generating process of their events of interest. The fact that this analysis is done in an out-of-sample context ensures that the AME framework is not simply overfitting with more parameters. Instead, it suggests that the additive and multiplicative effects are capturing underlying structure previously missed by the exogenous covariates in the models.

Ignoring this underlying structure has consequences for inferential analysis.¹⁹ The fact that there is such a divergence in the performance of AME versus the original estimation procedures highlights that there are unobserved sources of bias in each of these studies. We hone in on the main finding of each study to draw into focus the potential consequences for ignoring these sources of bias and the inferential benefits of the AME estimation procedure. In Table 2, we present the overall results; the term *Unconfirmed* indicates that the statistical significance of the crucial finding in the original study is not found to hold in the AME estimation.²⁰

An important takeaway here is that many scholars are forced to make knowledge claims based on the statistical significance of a small set of covariates or the differences between these covariates. These differences may change dramatically once interdependencies are taken into account.

¹⁹As shown in the simulation section, inference in a dyadic setting can become complicated when there are unobserved dependencies.

²⁰Full tabular results for each of the original and re-estimated models are presented in the Appendix.

Reiter & Stam (2003)



Weeks (2012)



Gibler (2017)



Figure 6. Assessments of out-of-sample predictive performance using ROC curves, PR curves, and separation plots.

Table 2. Here we provide a brief summary of the key variable in each of the replications and a note about whether or not the highlighted finding remains when using our network-based approach.

Study	Central Finding	Confirmed after accounting for dependencies?
Reiter & Stam (2003)	Personalist Regimes Attack Democracies, Not Vice Versa	Confirmed
Weeks (2012)	Bosses, Juntas, and Strongmen are more Aggressive, Machines are Not	Unconfirmed
Gibler (2017)	Entry to the International System, Not Power Parity, Determines Conflict	Partially Confirmed

This outcome follows from AME's ability to better account for the dependencies discussed in the previous section, whereas GLM approaches explicitly assume observational independence conditional on the specified covariates. As this is a widely-known limitation of GLM approaches, scholars often attempt to account for clustering of observations by including additional variables and adjusting the standard errors of the resulting estimates. At best, this method introduces noise and imprecision into results and at worst can produce misleading outcomes.

Next we discuss each of the replications in more detail and highlight the substantive insights drawn from the AME framework.

5.1. Reiter & Stam (2003). Reiter and Stam (2003) examine the relationship between democracy, dictatorship and the initiation of militarized disputes. Their work contests prior scholarship claiming that interstate dyads containing democracies and personalist dictatorships were particularly prone to conflict because of aggression on the part of the democratic state (Peceny et al., 2002). Using a directed dyadic dataset of almost a million observations, they find evidence against this hypothesis: dictators are in fact more likely to challenge democracies, but not the other way around. In addition, military regimes and single-party regimes are more prone to initiate disputes with democracies, but the opposite is not true.²¹ As is prevalent in this literature, Reiter & Stam employ a logistic regression that includes an indicator of the time since the last dispute as well as three

²¹Independent variables focus on various encodings of regime types, contiguity, alliance, and capability measures. For a full tabular display of the results see the Appendix.

cubic splines. Based on their statistical analysis, they conclude that institutional constraints affect the propensity of democratic and non-democratic leaders to engage in military conflict.

The key variables in the original model measure whether or not the sender in the directed dyad is a personalist regime and the target a democratic regime ("Pers/Democ directed dyad") or whether the opposite is true ("Democ/Personalist directed dyad"). The authors find that coefficient of the Pers/Democ Directed Dyad indicator is positive, while the Democ/Personalist directed dyad is too imprecisely measured to indicate a direction. In our re-estimation using the AME framework, we confirm these results, indicating that dictators are likely to initiate or engage in conflict with democratic regimes but not vice versa.

Even though we are able to confirm the original results, employing the AME model offers clear benefits in this case. As already shown in Figure 6, our approach performs notably better in reflecting the data generating process. The reason for this is that there is still underlying structure within this conflict system that the Reiter & Stam model does not fully capture. To highlight this we visualize the estimated sender random effects (a_i) from the SRRM portion of the AME framework in Figure 7.

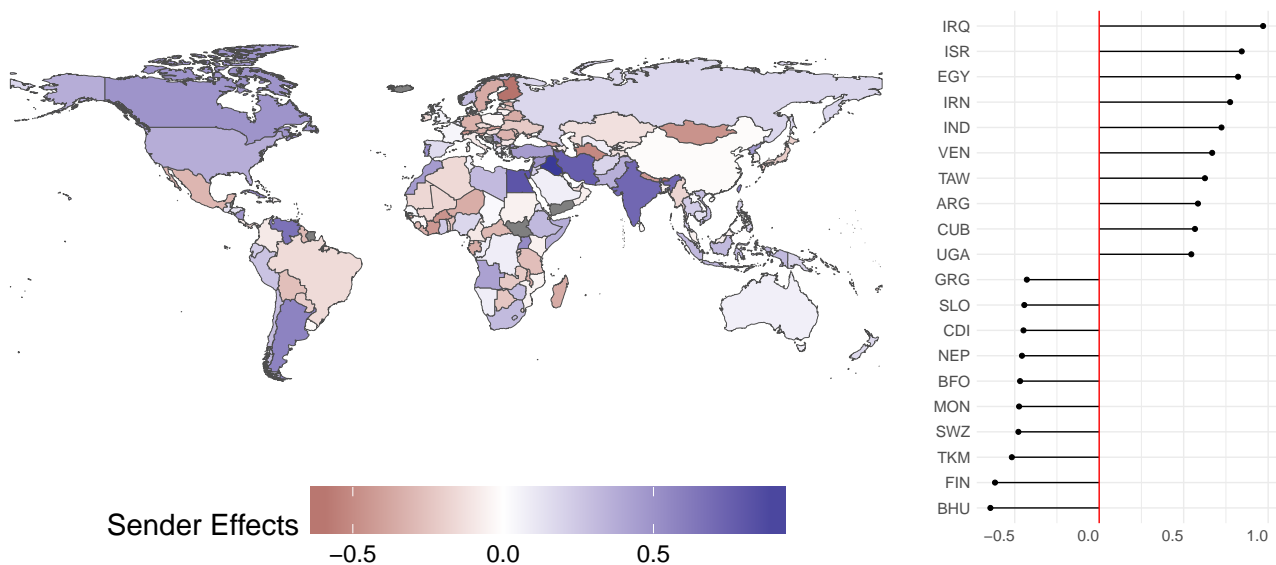


Figure 7. Estimates of sender random effects (a_i) from AME for the Reiter & Stam (2003) model. Positive values indicate that the particular country is more likely to be involved in conflict than predicted by the covariates in the model. Negative values indicate that the country is less likely to be involved in conflict. The left-panel shows a_i estimates for all countries and right-panel highlights the top ten countries in terms of positive values of a_i .

The visualization of the sender random effects highlights that the behavior of countries does not fully accord with predictions from the covariates specified by Reiter & Stam. Specifically, we can see that countries such as Iraq, Israel, Egypt, and Iran are more likely to be involved in initiating or continuing conflict with other countries than the model would predict. Further, other countries such as Bhutan, Finland, and Turkmenistan are less likely to engage in conflict than the exogenous covariates in the model would suggest. None of these findings change the key conclusion from Reiter & Stam's work, but by using the AME framework we are able to better understand the limitations of their model.

5.2. **Weeks (2012).** Weeks 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, and that this is dependent on the audiences of autocrats. When the autocratic regime is nonmilitary, domestic audiences 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 are more likely to employ military force in their foreign policy. To study this question, Weeks employs a directed dyad design of conflict similar to that used by Reiter & Stam.

The major innovation in her study resides in the nuanced way she conceptualizes and codes regimes into four types: a) Machine, b) Junta, c) Boss, and d) Strongmen.²² She uses a logistic regression, following Beck et al. (1998) and includes splines to capture temporal covariation in the dependent variable along with dyad clustered standard errors. The key findings of her work are that a) juntas, boss, and strongmen regimes are more likely to initiate conflict than machine-type regimes and b) machine-type regimes are no more belligerent than democracies.²³ In the empirical analysis, Weeks finds that machines are less prone to initiate conflict than the reference category, whereas juntas, bosses, and strongmen are more conflict-prone. When analyzing the results using AME, however, we find that the parameters on each of her autocratic regime type variables are too imprecisely measured to draw any inference about their putative causal effects. Consequentially,

²²Weeks also includes a variety of control variables focusing on capabilities for both sides of the dyad, alliances, geography, trade dependence, regime instability, and the regime type of "side B." For a full tabular display of the results see the Appendix.

²³These insights are mainly emphasized in the paper by the parameter estimates depicted in Tables 1 and 2 (pages 339–340) from the paper.

none of the findings from her original analysis are confirmed once known dependencies among the data are taken into account via AME.

There is also a striking difference between Weeks' original model and our estimation using AME in terms of capturing the data generating process. Specifically, the area under the ROC curve for AME and GLM models are 0.97 and 0.64, respectively, which is a dramatic difference given that both models use the same exogenous covariates. As with Reiter & Stam, the divergence is a result of the GLM framework's inability to account for underlying structure generating the event of interest. To uncover this structure, we illustrate another beneficial aspect of the AME framework, which is the multiplicative random effects estimated through the LFM portion of the model – $\mathbf{u}_i^T \mathbf{D} \mathbf{v}_j$. These random effects account for higher order dependencies that manifest as a result of homophily and stochastic equivalence in dyadic data.

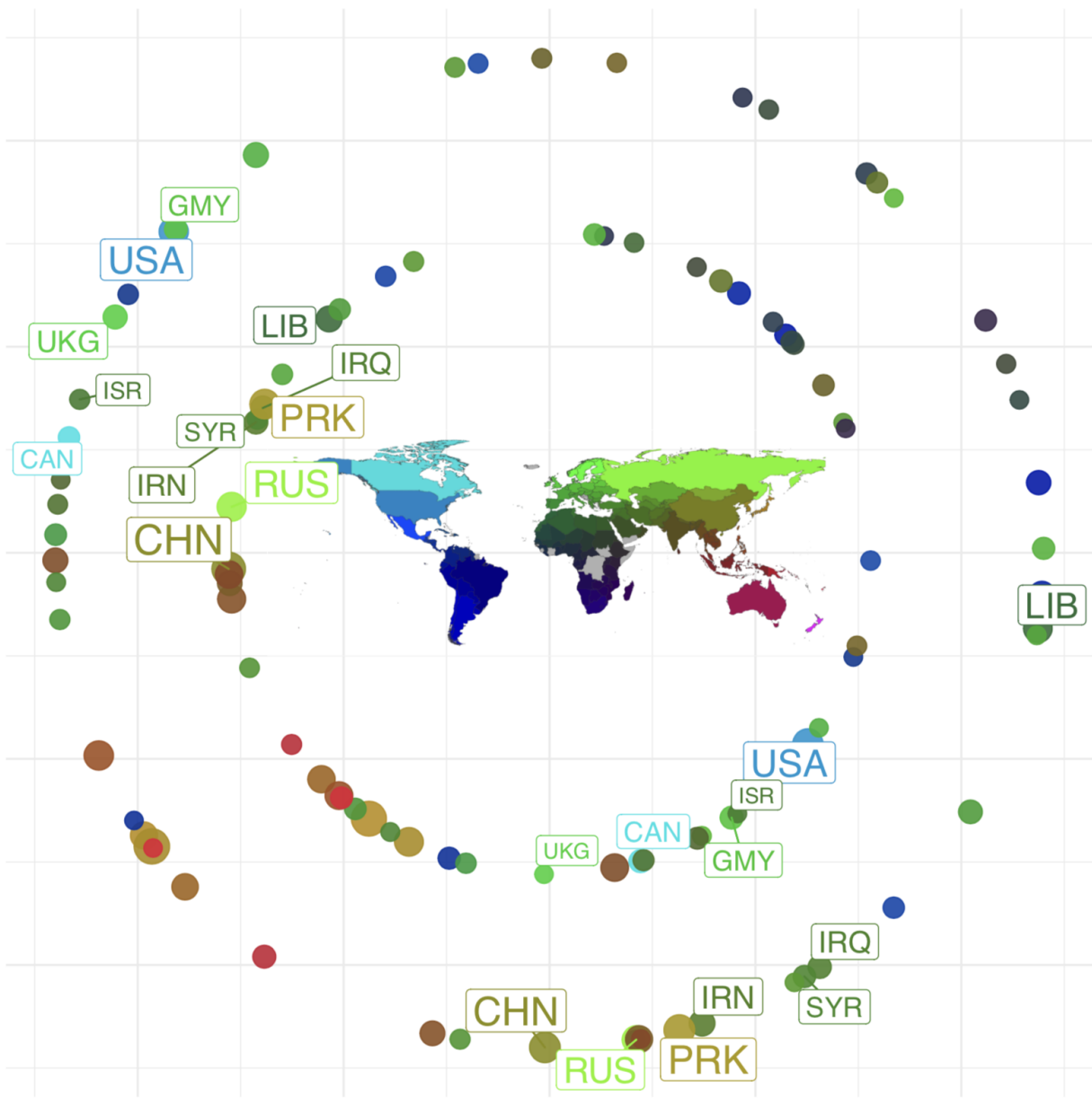


Figure 8. Visualization of multiplicative effects for Weeks (2012). Each circle designates a country and the color corresponds to the legend at the center of the visualization. Countries that cluster together in the outer ring are those that were found by the model to have similar sending patterns, meaning that they tend to send conflict to similar sets of countries. The inner ring clusters countries by the similarity of who they receive conflict from.

To visualize the multiplicative effects, we display the circular diagram shown in Figure 8. The nodes throughout the diagram represent countries and are colored according to their geographic position – a legend is provided in the center. The outer ring visualizes higher order dependence

patterns through countries' sender relationships – country positions here are estimated in the \mathbf{u}_i random effects described in Equation 1. Countries that are more proximate to each other in this outer ring are more likely to send or initiate conflicts with similar targets. The inner ring, based on estimates of \mathbf{v}_i , is constructed such that countries closer together are more likely to receive conflict from the same sender countries. Last, the distance between countries in the inner and outer rings proportionately reflects who a country is more likely to be in conflict with.

Figure 8 reveals a number of notable clusters. For example, in the top left corner we see the US, UK, Germany, Canada, and Israel. These states cluster together in the outer ring of this visualization because they tend to send conflicts to similar targets. Conversely, in the bottom right of the outer ring, we see a cluster of authoritarian countries: Iraq, Russia, Syria, North Korea, and China. We observe similar clusters in the inner ring. Specifically, we again see the US, UK, Germany, Canada, and Israel clustering together indicating that they are more likely to receive conflict from the same countries. The cluster of democratic and authoritarian countries facing each other in the inner and outer rings indicate that they are more likely to engage in conflict with one another.

Perhaps most critically, an evaluation of this visualization highlights Weeks' original expectation of how states behave in the conflict system. Specifically, Iraq, Syria, Libya, and North Korea all fell under Weeks' "boss" category, and each of these states tends to cluster together in the inner and outer rings. This indicates that even though we do not find support for Weeks' assertion that certain authoritarian regime types are more likely to initiate conflict, we do find that these regimes are more likely to behave similarly in terms of who they target and receive conflict from.

5.3. Gibler (2017). The last replication we conduct with the AME model considers a study by Gibler (2017). Gibler argues that the long-standing relationship between the relative parity of capabilities and initiation of militarized interstate disputes (MIDs) is almost completely mediated by the initial conditions for the members of the dyad when they joined the international system as sovereign members. In most specifications, when you take into account the initial conditions for dyadic entry, the statistical significance of power parity vanishes. This finding calls into question many IR theories about the role of balance in generating international conflict (Organski, 1958).

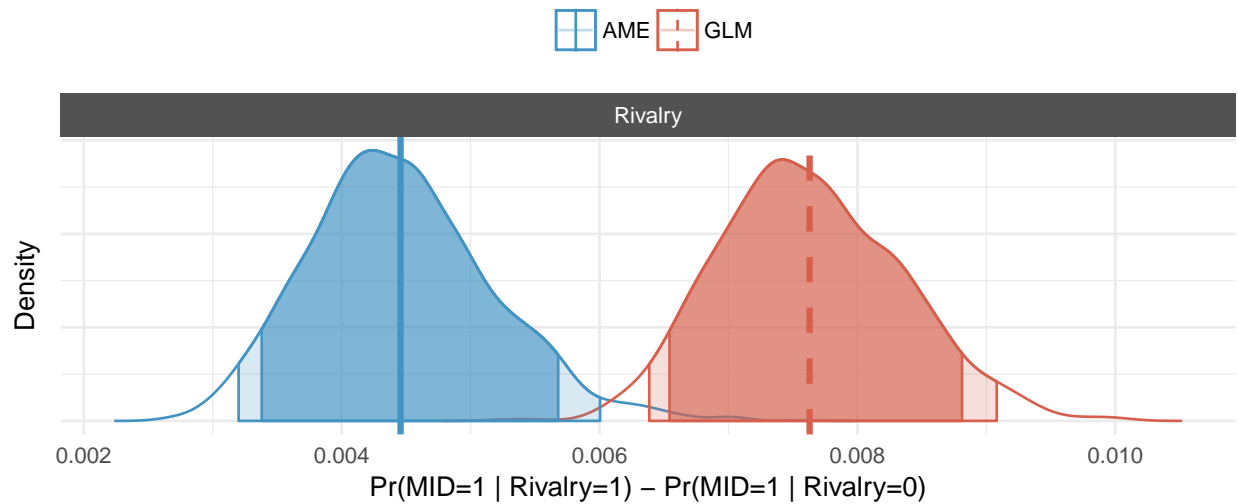
To test this hypothesis, Gibler employs an undirected dyadic design and estimates his model using a GLM with dyad-clustered standard errors.²⁴ With this design, Gibler finds support for both the insignificance of contemporary power parity, and the effect of the initial conditions for entry on driving conflict behavior. When we reestimate using AME power parity still has a small, imprecisely measured effect, supporting Gibler's argument that contemporary power parity is an artifact of other aspects of state relations, and does not drive contemporary conflict behavior. At the same time, the AME approach finds that the effect of power parity when a country enters into the international system is too imprecisely measured to draw any inferences about its effect on the initiation of MIDs. Not only is our estimate of the effect of this variable small, but it has a very large relative standard error—over a magnitude larger than the parameter itself. Thus, while we can confirm the argument that the effect of power parity vanishes when accounting for both initial conditions and network dependencies, we find less support for the argument that initial conditions drive the initiation of MIDs.

In the two previous replications, we showed how to parse apart underlying structure using the SRM and LFM portions of the AME framework. Here we turn our focus to parameter interpretation in a substantive context. Specifically, even when the GLM and AME frameworks produce results that may seem to be in accordance with one another, the substantive interpretations of the effects of covariates can differ notably since the AME model uses a set of random effects to account for unobserved factors. To explore this, we focus on the effect of rivalry on MID initiation. Both the GLM and AME estimations find that rivalry has a positive effect on MID initiation, but the expected effects between the two models differs greatly.

To clarify the difference, we turn to a simulation based approach. We employ mean or modal values for all independent variables, except we change the rivalry variable to indicate that there was a rivalry when the actual data suggest there is none. This provides us with two scenarios, one in which rivalry is set to one and the other zero, while in both scenarios all other parameters are set to their measure of central tendency. The expected values of this scenario are essentially a first difference plot comparing results with the model when estimated in two different ways: the GLM estimation and our AME approach.

²⁴Full tabular results are shown in the appendix, we focus on model 6 from Table 6 (2017, 34).

Figure 9. Marginal effects of a change in the Rivalry variable for both the AME and the Gibler estimation.



As Figure 9 illustrates, the substantive results differ notably. First, the expected value of the dependent variable—the probability of the onset of a militarized interstate dispute, is considerably lower once interdependencies are taken into account with the AME model. These are rare events, so the probabilities are low, but the difference is a factor of almost two. Thus, we find substantially different expected values from these two models.

5.4. Lessons Learned. First, utilizing the AME framework enables scholars to better model the underlying structure inherent to dyadic data. In each of the replications, the AME’s predictive performance out-of-sample is notably superior. Specifically, not only does AME perform better at correctly identifying cases in which the dependent variable takes a value of 0 (via the ROC curves), but it also dominates at correctly identifying occurrences of the dependent variable in the data (seen via the PR curves).

Second, in both the simulation and replication sections we have shown that results can change notably when interdependencies are not taken into account. Not only are coefficients biased in the GLM approaches, but they are often imprecisely measured with poorly calibrated standard errors. This means that significance testing (for better or worse) is compromised when network effects are ignored.

Third, even when the results from the AME estimation conform with those found in an OLS or logistic regression, new insights can emerge from the structure uncovered by the random effects of the AME framework. In particular, the AME provides information about the dependencies so that clusters can be identified. This information can then be used to generate new hypotheses.

Fourth, it is evident that the actual results—not the estimated coefficients and their covariances—generated by the models differ greatly in expectations. This implies that policy experimentations with the models, as well as scenario-based simulations and forecasting of GLM models are likely to give misleading results compared to the AME approach.

6. CONCLUSION

International relations is about the interactions, relationships, and dependencies among countries or other important international actors. This is particularly true of scholarship in the tradition of the Correlates of War Project, but it is by no means limited to it.²⁵ Many scholars have debated the use and abuse of dyadic data.²⁶ Research published as recently as 2017 and a broad survey of the IR literature makes it clear that scholars find dyadic data to be an essential touchstone in the study of international relations (Erikson et al., 2014; Aronow et al., 2015).

At the same time, we know that research designs focusing on the statistical analysis of dyadic data quickly go astray if the dyadic data are assumed to be iid. Virtually all of the standard statistical models—ordinary least squares and logistic regressions, to name a few—fail if the data are not conditionally independent. This fact has been accepted as it relates to temporal dependencies, but adoption of methods to account for network dependencies have seen less progress. By definition dyadic data are not iid and thus the standard approaches cannot be used cavalierly to analyze these data. Signorino (1999) showed why this is true of models of strategic interaction, but it is more broadly true of models that employ dyadic data. We show that the AME framework can be employed to account for the statistical issues that arise when studying dyadic data.

To explore this approach in the context of international relations we have presented two analyses. The first is a simulation where the characteristics of the network are known. This shows

²⁵See Singer (1972) for an early description of the project and also see the project's Web site for an history and more recent efforts <http://www.correlatesofwar.org/>.

²⁶One recent on-line symposium can be found at <http://bit.ly/2wB2hab>.

that when there are unobserved dependencies, the AME approach is less biased in terms of parameter estimation compared to the standard approach employed in international relations to study dyadic data (i.e., GLM models). The second analysis is a replication of recent studies that use a broad range of dyadic data to draw inferences about international relations. These studies have been replicated with the original research designs, each of which used a statistical method that assumes the dyadic data are all independent from one another. We then re-analyzed each study using the AME model. In every case, we found that the AME approach provided a) increased precision of estimation, b) better out-of-sample fit, and c) evidence of 1st-, 2nd-, and 3rd-order dependencies that were overlooked in the original studies.²⁷

²⁷The Appendix contains performance data on all of these replications, as well as sample code illustrating how to undertake AME analysis using `amen`.

REFERENCES

- Adamic, Lada A., and Glance, Natalie. 2005. The Political Blogosphere and the 2004 US Election: Divided they Bloglog. Pages 36–43 of: *Proceedings of the 3rd International Workshop on Link Discovery*. New York, N.Y.: ACM, for ACM.
- Anderson, Carolyn J., Wasserman, Stanley, and Faust, Katherine. 1992. Building Stochastic Block-models. *Social Networks*, **14**(1), 137–161.
- Aronow, Peter M., Samii, Cyrus, and Assenova, Valentina A. 2015. Cluster-Robust Variance Estimation for Dyadic Data. *Political Analysis*, **23**(4), 564–577.
- Barabási, Albert-László, and Réka, Albert. 1999. Emergence of Scaling in Random Networks. *Science*, **286**(October 15), 509–510.
- Beck, Nathaniel. 2012. Sweeping Fewer Things under the Rug: Tis Often (Usually?) Better to Model than be Robust. *The Society for Political Methodology POLMETH XXIX*.
- Beck, Nathaniel, and Katz, Jonathan N. 1995. What To Do (and Not To Do) with Pooled Time-Series Cross-Section Data. *American Political Science Review*, **89**(3), 634–647.
- Beck, Nathaniel, Katz, Jonathan N., and 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, James, and Lanning, Stan. 2007. The Netflix Prize. In: *Proceedings of KDD cup and workshop*.
- Block, Per, Stadtfeld, Christoph, and Snijders, Tom A.B. 2017. Forms of Dependence: Comparing SAOMs and ERGMs from Basic Principles. *Sociological Methods & Research*.
- Chyzh, Olga. 2016. Keeping up with which Joneses: Spatial Diffusion of Rule of Law through Economic International Organizations. *Foreign Policy Analysis*, **13**(1), 28–49.
- Davis, Jesse, and Goadrich, Mark. 2006. The Relationship Between Precision-Recall and ROC Curves. Pages 233–240 of: *Proceedings of the 23rd International Conference on Machine Learning*. ICML 2006. New York, NY, USA: ACM.
- Dorff, Cassy, and Minhas, Shahryar. 2017. When Do States Say Uncle? Network Dependence and Sanction Compliance. *International Interactions*, **43**(4), 563–588.
- Erikson, Robert S., Pinto, Pablo M., and Rader, Kelly T. 2014. Dyadic Analysis in International Relations: A Cautionary Tale. *Political Analysis*, **22**(4), 457–463.

-
- Franzese, Robert, and Hays, Jude C. 2007. Spatial Econometric Models for the Analysis of TSCS Data in Political Science. *Political Analysis*, **15**(15), 2.
- Gibler, Douglas M. 2017. State Development, Parity, and International Conflict. *American Political Science Review*, **111**(1), 21–38.
- Greenhill, Brian, Ward, Michael D., and 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, and 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).
- King, Gary, and Roberts, Margaret E. 2014. How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do about It. *Political Analysis*, **23**(2), 159–179.
- Kinne, Brandon J. 2013. Network Dynamics and the Evolution of International Cooperation. *American Political Science Review*, **107**(04), 766–785.
- Li, Heng, and 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., and 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.
- Minhas, Shahryar, Hoff, Peter D., and Ward, Michael D. 2018. Inferential Approaches for Network Analysis: AMEN for Latent Factor Models. *Political Analysis*.
- Mucha, Peter J., Richardson, Thomas, Macon, K., Porter, Mason A., and Onnela, J. P. 2010. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science*, **328**(5980), 876ff.
- Organski, A.F.K. 1958. *World Politics: The Stages of Political Development*. New York: Alfred A. Knopf.

-
- Peceny, Mark, Beer, Caroline C., and Sanchez-Terry, Shannon. 2002. Identifying the Culprit: Democracy, Dictatorship, and Dispute Initiation. *American Political Science Review*, **96**(1), 15–26.
- Reiter, Dan, and Stam, Allan C. 2003. Identifying the Culprit: Democracy, Dictatorship, and Dispute Initiation. *American Political Science Review*, **97**(2), 333–337.
- Resnick, Paul, and Varian, Hal R. 1997. Recommender Systems. *Communications of the ACM*, **40**(3), 56–58.
- Richardson, Lewis F. 1960. *Arms and Insecurity*. Chicago and Pittsburgh, PA: Quadrangle/Boxwood.
- Signorino, Curtis. 1999. Strategic Interaction and the Statistical Analysis of International Conflict. *American Political Science Review*, **92**(2), 279–298.
- Singer, J. David. 1972. The Correlates of War Project: Continuity, Diversity, and Convergence. Chap. 2 of: Singer, J. David (ed), *Quantitative International Politics: An Appraisal*. Special Studies in International Politics and Government, vol. VI. Praeger.
- Wasserman, Stanley, and 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.

APPENDIX A. AME MONTE CARLO MARKOV CHAIN ALGORITHM

Given initial values of $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2\}$, the algorithm proceeds as follows until convergence:

- sample $\boldsymbol{\theta} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)
- sample $\boldsymbol{\beta} \mid \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)
- sample $\mathbf{a}, \mathbf{b} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)
- sample $\Sigma_{ab} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \rho, \text{ and } \sigma_\epsilon^2$ (Inverse-Wishart)
- update ρ using a Metropolis-Hastings step with proposal $p^* \mid \rho \sim \text{truncated normal}_{[-1,1]}(\rho, \sigma_\epsilon^2)$
- sample $\sigma_\epsilon^2 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \text{ and } \rho$ (Inverse-Gamma)
- For each $k \in K$:
 - Sample $\mathbf{U}_{[:,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}_{[:, -k]}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)
 - Sample $\mathbf{V}_{[:,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}_{[:, -k]}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)
 - Sample $\mathbf{D}_{[k,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)²⁸

²⁸Subsequent to estimation, \mathbf{D} matrix is absorbed into the calculation for \mathbf{V} as we iterate through K .

APPENDIX B. ADDITIONAL REPLICATION INFORMATION

For each of the replications involving a binary dependent variable we provide a table of coefficient estimates that includes the original GLM estimation and the results from the AME model. We have adopted for presentation purposes only the use of significance testing on these observational data for facile comparison with the replicated studies. However, note that we do provide predictive heuristics for these models as well.

Reiter & Stam (2003). Additional information for the Reiter & Stam (2003) re-estimation.

Variable	GLM (Logit)	AME
Intercept	-4.784 (0.097)	-3.144 (0.06)
Pers/Democ Directed Dyad	1.026 (0.14)	0.255 (0.068)
Democ/Pers Directed Dyad	0.083 (0.191)	0.112 (0.079)
Personal	0.281 (0.265)	0.211 (0.11)
Military	-0.323 (0.574)	-0.025 (0.249)
Single	-0.677 (0.144)	-0.07 (0.073)
Democracy	-1.073 (0.194)	-0.254 (0.063)
Contiguous	2.912 (0.09)	1.296 (0.033)
Major Power	2.174 (0.101)	0.906 (0.093)
Ally	0.078 (0.086)	0.136 (0.037)
Higher/Lower Power Ratio	-0.316 (0.027)	-0.111 (0.011)
Economically Advanced	-0.175 (0.131)	0.053 (0.05)
Years Since Last Dispute	-0.381 (0.023)	-0.129 (0.008)
Cubic Spline 1	-0.004 (0.000)	-0.001 (0.000)
Cubic Spline 2	0.002 (0.000)	0.001 (0.000)
Cubic Spline 3	-0.001 (0.000)	0.000 (0.000)

Table B1. Parameter comparison for Reiter & Stam (2003). Standard errors in parentheses.

Weeks (2012). Additional information for the Weeks (2012) re-estimation.

Variable	GLM (Logit)	AME
(Intercept)	-3.784 (0.423)	-2.409 (0.132)
Machine	-0.459 (0.174)	-0.006 (0.04)
Junta	0.515 (0.169)	0.034 (0.046)
Boss	0.649 (0.153)	-0.044 (0.044)
Strongman	0.832 (0.132)	0.032 (0.044)
Other Type	0.147 (0.132)	-0.01 (0.034)
New/Unstable Regime	-0.312 (0.092)	-0.043 (0.031)
Democracy Target	0.185 (0.115)	0.024 (0.026)
Military Capabilities Initiator	5.234 (1.69)	0.071 (0.412)
Military Capabilities Target	6.34 (1.675)	-0.969 (0.48)
Low Trade Dependence	-24.794 (12.866)	-4.733 (3.017)
Both Major Powers	1.136 (0.547)	1.122 (0.241)
Minor/Major	0.772 (0.239)	0.496 (0.118)
Major/Minor	0.711 (0.225)	0.778 (0.16)
Contiguous	2.172 (0.32)	0.705 (0.06)
Log Dist. Between Capitals	-0.209 (0.038)	-0.129 (0.01)
Alliance Similarity Dyad	-0.999 (0.144)	-0.073 (0.065)
Alliance Similarity With System Leader Initiator	0.11 (0.24)	0.068 (0.057)
Alliance Similarity Leader Target	0.203 (0.244)	0.08 (0.056)
Time Since Last Conflict	-0.229 (0.018)	-0.067 (0.007)
Spline1	-0.001 (0.000)	0.000 (0.000)
Spline2	0.000 (0.000)	0.000 (0.000)
Spline3	0.000 (0.000)	0.000 (0.000)

Table B2. Parameter comparison for Weeks (2012). Standard errors in parentheses.

Gibler (2017). Additional information for the Gibler (2017) re-estimation.

Variable	GLM (Logit)	AME
(Intercept)	-5.826 (0.366)	-2.758 (0.045)
Allied	0.133 (0.102)	0.078 (0.021)
Joint Democracy	-0.527 (0.099)	0.005 (0.022)
Peace Years	-0.261 (0.016)	-0.058 (0.004)
Spline 1	-0.001 (0.000)	0.000 (0.000)
Spline 2	0.000 (0.000)	0.000 (0.000)
Spline 3	0.000 (0.000)	0.000 (0.000)
Contiguity	2.427 (0.196)	0.66 (0.023)
Parity	-0.77 (0.551)	-0.067 (0.057)
Parity at Entry Year	2.034 (0.617)	-0.05 (0.065)
Rivalry	2.034 (0.213)	0.655 (0.028)

Table B3. Parameter comparison for Gibler (2017). Standard errors in parentheses.

APPENDIX C. AME TUTORIAL

Using the AMEN function requires formatting data into a particular structure. The primary distinction in data formatting is whether the outcome of interest represents a directed or undirected network.

If undirected, the AMEN function has three main inputs:

- **Y**: a T length **list** of $n \times n$ adjacency matrices, where T = number of years in the dataset and n = number of nodes in the network.
- **Xdyad**: a T length **list** of $n \times n \times p$ arrays, where p = number of dyadic covariates in dataset.
- **Xrow**: a T length **list** of $n \times p$ matrices, where p = number of monadic (nodal) covariates in dataset.

If directed, AMEN further requires:

- **Xrow**: a T length list of $n \times p$ matrices, where p = number of sender (nodal) covariates in dataset.
- **Xcol**: a T length list of $n \times p$ matrices, where p = number of receiver (nodal) covariates in dataset.

Beyond the data inputs, the AMEN function requires additional specification:

- **model**: how to model the outcome variable, e.g., 'logit'
- **symmetric**: whether the input network is symmetric
- **intercept**: whether to estimate an intercept
- **nscan**: number of iterations of the Markov chain
- **burn**: burn-in period
- **odens**: thinning interval
- **R**: dimension of the multiplicative effect (referred to as K in the paper)
- **gof**: whether to calculate goodness of fit statistics

There is often little theoretical reason to choose a particular value of R (above). One strategy is to estimate models at different values of R and compare goodness of fit statistics across models.

Given the computational intensity needed for parameter estimates to converge, parallelization strategies are recommended to speed up analysis. In addition, providing AMEN function with

starting values, either dictated by theory, previous research, or previous runs can also help speed up convergence time.

The code below presents an example of an AME model running in parallel across 4 different levels of R. Note also that the model is using starting values from a previous run, defined in *startValso*.

```
# running in parallel varying k
imps = 10000 ; brn = 25000 ; ods = 10 ; latDims = 0:3

# Run amen in parallel
library(doParallel) ; library(foreach) ; cl=makeCluster(4) ; registerDoParallel(cl)
foreach(ii=1:length(latDims), .packages=c("amen")) %dopar% {

  # load previous model run
  load(prevModelFiles[ii])
  # extract start vals
  startValso = ameFit$startVals
  # dump rest
  rm(ameFit)

  ameFit = ame_repL(
    Y=yList ,Xdyad=xDyadList ,Xrow=NULL,Xcol=NULL,
    model="bin",symmetric=FALSE,intercept=TRUE,R=latDims[ii],
    nscan=imps, seed=1, burn=brn, odens=ods,
    plot=FALSE, print=FALSE, gof=TRUE, startVals=startValso ,
    periodicSave=TRUE )
  save(ameFit, file=pasteo('model_k', latDims[ii], '_v2.rda') )
}

stopCluster(cl)
```