TAKING DYADS SERIOUSLY

ABSTRACT. Much of international relations scholarship concerns dyads: dyadic hypotheses and especially dyadic data. Yet, contemporary research brushes aside the complications that come with analyzing interdependent relational data. Dyadic observations do not typically satisfy the conditional independence criterion required for many statistical approaches. As a result, many studies often produce results with biased coefficient estimates and poorly calibrated standard errors. These biases have profound consequences for evaluating parametric models. We present an alternative, regression-based, approach that accounts for the dependencies complicating this type of analysis. We first present a simulation exercise highlighting the model's ability to account for the dependencies that emerge in relational data. In addition, we replicate four recent studies in recent international relations scholarship, comparing the standard approach to our alternative. For each study, we find that conventional methods overstate the effect of key variables in the study of interstate conflict (and trade), underestimate the uncertainty in these effects, and in some cases lead researchers to faulty conclusions about the statistical significance and substantive importance of their variables. Further, we show that our approach dominates in terms of out-of-sample cross-validations, rendering it more useful in forecasting applications and in modeling the data generating process behind outcomes of interest.

1. Introduction

? 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) to test hypothesized relationships. 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 having 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. (**???**).

In this article, we present the Additive and Multiplicative Effects (AME) model, a Bayesian approach for directly modeling relational data to better reflect the interdependencies underlying the data generating process of dyadic data structures (**??**). 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, especially 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.

[1]In 2017, *International Studies Quarterly* published a special issue on Dyadic Research Designs along with an online symposium to discuss the papers.
[2]In 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. The comparison reveals that in accounting for observational dependence, AME produces more precise estimates and better-calibrated confidence intervals for key variables in the literature. Consequently, AME produces results that, at times, differ from those found in the original study as well as from the broader literature. Moreover, we demonstrate the latent factor approach offers substantive insights that are often occluded by ignoring the interdependencies found in the relational data of IR studies. Finally, we show that for each replication our network-based approach provides substantively more accurate out-of-sample predictions than the models used in the original studies.

The framework that we present 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 approach can be used by scholars in the field to continue to generate substantive insights, while accounting for the data generating process behind political 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, 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 is taking 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 (**??**). 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 (**?**). 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 ties in a network tend to also be more likely to receive them. As a result, the rows and columns

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

SENDER HETEROGENEITY

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

RECEIVER HETEROGENEITY

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

SENDER-RECEIVER COVARIANCE

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

RECIPROCITY

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

Figure 1. Nodal and dyadic dependencies in relational data.

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 (**??**).

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 (**??**). 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.

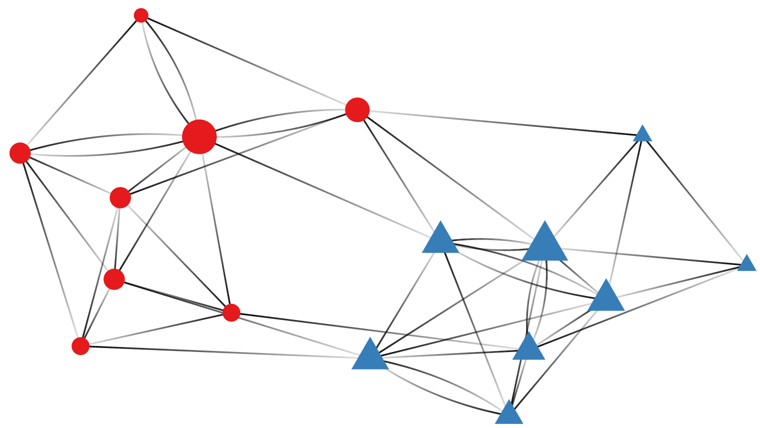


Figure 2. Higher-order dependence patterns in a network.

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 **?**, who visualize linkages between right and left leaning political blogs preceding the 2004 United States Presidential Election. **?** 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 (**?**), which is often formalized through the concept of stochastic equivalence (**?**). 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 with a divergent political position.

These types of patterns frequently emerge in IR contexts as well.[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 **???**.[6] The AME model is specified as follows:

[4]For example, see: **???**.

[5]**?** detail how this framework contrasts with alternative network-based approaches.

[6]An earlier version of the LFM used in AME is presented as the general bilinear mixed effects (GBME) model in **?**. 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.

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

where $y_{ij,t}$ captures the interaction between actor i (the sender) and j (the receiver) at time t. We use a Bayesian probit framework, in which 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 ($\boldsymbol{\beta}_r^\top \mathbf{X}_j$). 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:

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

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 (**??**). 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.

By integrating the SRRM and LFM into a Bayesian probit framework, we can account for the underlying structure in dyadic data that, if left un-estimated, would complicate any inferences we might wish to draw for the exogenous parameters. Parameter 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. Specifically, given initial values of $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and $\sigma_\epsilon^2\}$, the algorithm proceeds as follows until convergence:

- sample $\boldsymbol{\theta} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
- sample $\boldsymbol{\beta} \mid \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
- sample $\mathbf{a}, \mathbf{b} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
- sample $\Sigma_{ab} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \rho,$ and σ_ϵ^2 (Inverse-Wishart)
- update ρ using a Metropolis-Hastings step with proposal $\rho^* | \rho \sim$ truncated normal$_{[-1,1]}(\rho, \sigma_\epsilon^2)$

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

[8]Unlike 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.

- sample $\sigma_\epsilon^2 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab},$ and ρ (Inverse-Gamma)
- For each $k \in K$:
  - Sample $\mathbf{U}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}_{[,-k]}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - Sample $\mathbf{V}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}_{[,-k]}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - Sample $\mathbf{D}_{[k,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)[9]

The presence of these dependencies in relational data points to the fact that there is a complex structure underlying the dyadic events that we observe, and that accounting for this structure is necessary if we are to adequately represent the data generating process. Of course, 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 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

We utilize a simulation study to 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 the simulation is to assess how well AME can provide unbiased and well-calibrated estimates of coefficient parameters in the presence of unobserved dependencies. Specifically, we are concerned with conducting inference on regression parameters of a linear model for a network in the case where

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

[10]Alternative 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, **?** 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.

there is an unaccounted for higher order dependence pattern. For instance, assume that the true data-generating process for a particular Y is given by:

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

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 primarily theoretical concern for applied scholars—using three models:

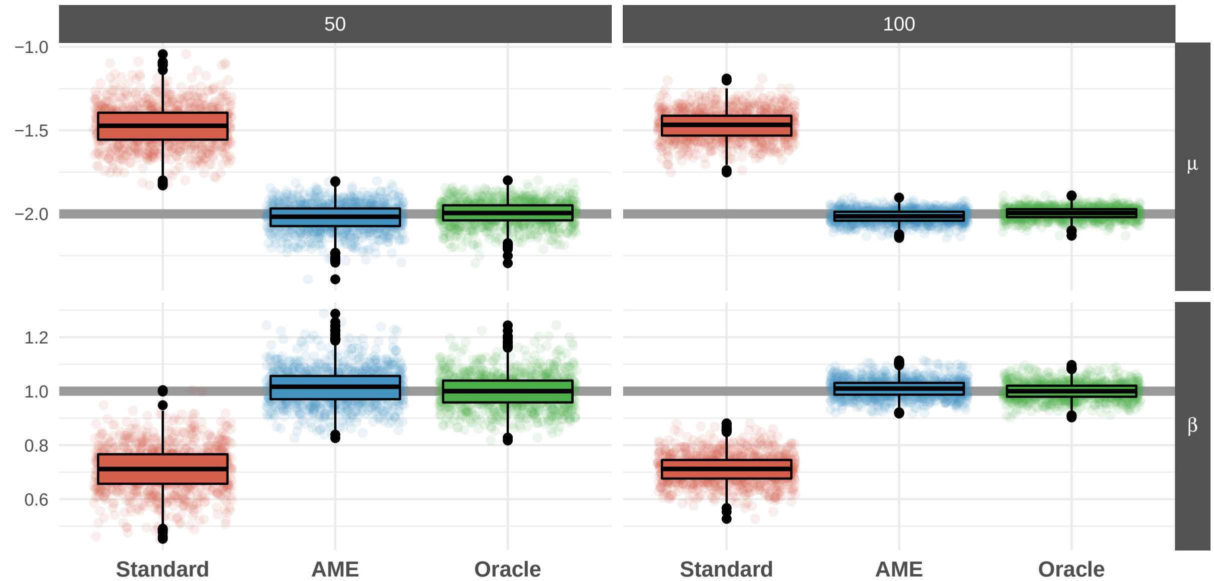
- the "standard" international relations approach estimated through a typical generalized linear model;
- the AME approach outlined in the previous section with a unidimensional latent factor space ($K = 1$);[11]
- and an "oracle" regression model that assumes we have measured all sources of dependencies 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 in 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.[12] We conduct two sets of simulations, one in which the number of actors in the network is

[11]Results with higher values of K are similar.
[12]The 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.

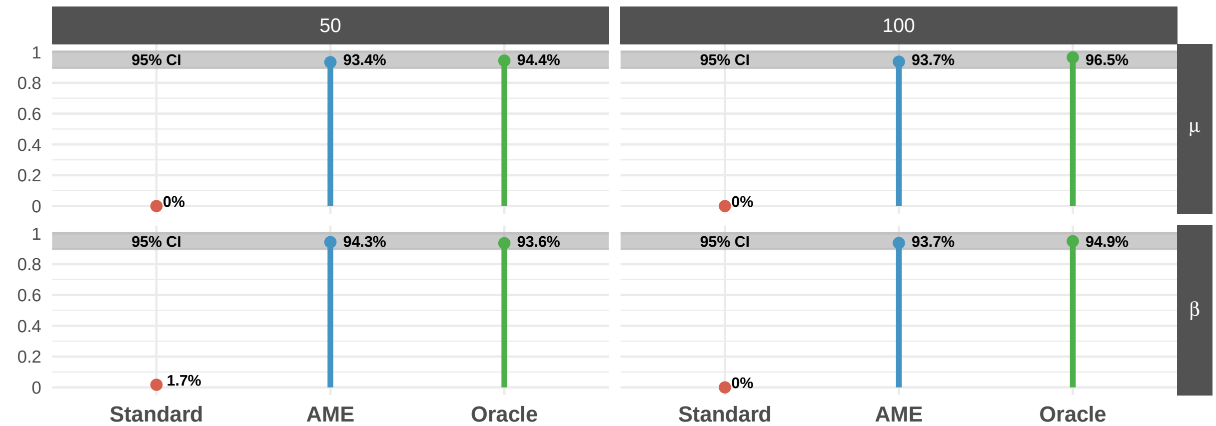


set to 50 and the other at 100. In total, we ran 1,000 simulations where we begin by simulating Y from the specification given in Equation 3 and then for each simulated Y we estimated a standard, AME, and oracle model.

We compare the performance of the models first in terms of how well they estimate the true values of μ and β in Figure 3 by depicting the average μ and β estimates from the simulations for the three models. The panels in the left show the results for when the number of actors is set to 50 and those on the right for 100; and 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 case in estimating the true parameter values.

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

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.

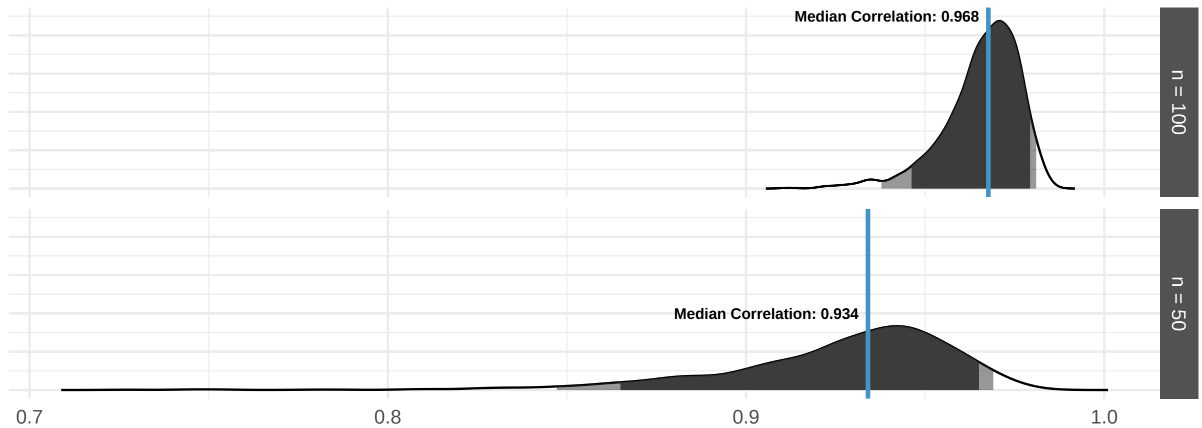


and conducting inferential tasks in the presence of unobserved dependencies. The AME approach by comparison can be used as a tool for scholars working with dyadic data to still estimate the true effects of their main variables of interest, while accounting for dependencies that do often emerge in dyadic 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 had a homophilous effect on Y. Homophilous because W within this framework is simply an example of a dyadic attribute involving i and j that positively affects the degree to which they will interact with one another, i.e., y_{ij}. This type of unobserved dependency will be captured through the multiplicative effects portion of the model, $\mathbf{U}^\top \mathbf{D} \mathbf{V}$. To estimate how well the model performs, we recover the multiplicative effects term for each simulation and calculate the correlation between it and the unobserved dependency, W.[13] 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 this

[13]Specifically, 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.



framework is not simply capturing noise but can be used as a tool to estimate unobserved structure.

What this simulation has shown is 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 a 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 models. At the very least, what this simulation section shows is that the AME model can help in estimating the true effect and conducting inference on the independent variables that scholars have operationalized to test their theoretical propositions.

5. Applications with AME

5.1. **Design.** We apply AME to four recent IR studies: **????**. Each of these studies are representative of broader trends in the field in that they use relational data of state interactions and propose both dyadic, monadic, and structural explanations for behavior of actors in the system. We choose to demonstrate the capabilities of AME with reference to existing studies in order to highlight several features of the AME approach. First, the results of AME estimation are interpretable alongside results using standard approaches, but, as shown in the simulation section, have the

Table 1. Features of the Studies Re-estimated.

	Model	Date Range	N. Actors	N. Dyads	Dyads Type	Clustering $\sigma_{\hat{\beta}}$
Reiter & Stam (2003)	Logit	1945–1995	193	753,456	Directed	Robust
McDonald (2004)	Logit	1959–2002	198	92,354	Undirected	Robust
Weeks (2012)	Logit	1946–1999	197	901,540	Directed	Robust
Gibler (2017)	Logit	1816–2008	193	650,557	Undirected	None

additional benefit of being able to take into account dependencies that may complicate inference. Second, through using this approach, we can also quantify the degree to which first, second, and third order dependencies are present in events of interest. Third, we show that by using the AME framework scholars can better model the data generating process behind their events of interest.

We obtained the data for each of these studies from their replication archives and replicated the main results of each article.[14] The four studies used below were selected based on how recent they were and whether they had more then 100 citations.[15] Each of these pieces, published in prominent journals well-known in their respective literatures, posited a theory in which interdependencies are consequential. Reflecting the dominant approach in the literature, each of the authors tested their hypothesis by employing some form of a general linearized model.[16]

Each of the four studies has a crucial finding that we hone in on to further draw into focus the analytical power of the AME estimation procedure. In Table 2, we present the overall results; the term *Unconfirmed* indicates only that the sign and/or significance of the putatively crucial finding in the original study is not found to hold in the AME estimation.[17]

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 when interdependencies are taken into account

[14] Without exception this was straightforward to accomplish, thanks to the authors' transparency and an increasing norm in the social sciences of open data sharing.
[15] Note 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. However, as described in the following sections, the AME model has specific data requirements, which limited the scope of potential papers.
[16] It is important to note that the AME framework is applicable only where a full set of dyads for an outcome of interest is observable—for example it is unsuitable for studying onsets, where dyad-years representing conflict continuations are removed from the sample. The AME is also not applicable to analyses of case-level data, for example, studies that examine the decision to go to war by particular states.
[17] Full tabular results for each of the original and reestimated models are presented in the Appendix.

Table 2. Here we provide a brief summary of the key variable in each of the four 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	Partially Confirmed
McDonald (2004)	Lower Trade Barriers and Higher Trade Lead to Peace	Confirmed
Weeks (2012)	Bosses, Juntas, and Strongmen are more Aggressive, Machines are Not	Unconfirmed
Gibler (2017)	Power Parity at Time of Entry to International System Increases Conflict	Unconfirmed

directly. 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.

Beyond just comparing parameter estimates, we examine how well each approach can represent the data generating process using an out-of-sample cross validation strategy. Specifically, 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 $s \in \{1, \ldots, k\}$, we:

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

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

The result of this procedure is a set of sociomatrices $\hat{Y}$, in which each entry $\hat{y}_{ij,t}$ is a predicted value obtained from using a subset of the data that does not include $y_{ij,t}$. We summarize the performance of the various models in Table 3 below. For the binary models we provide the area under the Receiver Operator Characteristic (ROC) and Precision Recall (PR) curves. Only one of the studies here had a continuous dependent variable and for this we provide the root mean squared error (RMSE) and root median squared error (RMDSE).[18] 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 is important as it indicates that switching to the AME framework—even when using the exact same 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, rather the dependence parameters we include are capturing underlying structure previously missed by the exogenous covariates in the models.

		GLM	AME
Reiter & Stam (2003)	Area Under ROC Curve, AUC-ROC	0.92	0.96
	Area Under PR Curve, AUC-PR	0.08	0.15
McDonald (2004)	AUC-ROC	0.92	0.99
	AUC-PR	0.13	0.28
Weeks (2012)	AUC-ROC	0.64	0.97
	AUC-PR	0.00	0.15
Gibler (2017)	AUC-ROC	0.52	0.91
	AUC-PR	0.00	0.08

Table 3. Here we provide a summary of the out-of-sample performance based on our cross-validation strategy for each of the four replications when using the standard dyadic approach and our network-based approach. Each of the studies involved a binary dependent variable and area under the curve (AUC) statistics are reported.

6. Conclusion

International relations is generally about the interactions and dependencies among a set of countries or other important actors such as international governmental organizations (IGOs). This is particularly true of scholarship in the tradition of the Correlates of War Project, but it is by no means limited to it.[19] Many scholars have debated the use and abuse of dyadic data.[20] It is clear from a survey of the literature and from work in this area published as recently as 2017 that many find dyadic data to be an important touchstone in the study of international relations (**??**).

[18] More details on the performance of each of these models can be found in the Appendix.

[19] See **?** 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/.

[20] One recent on-line symposium can be found at http://bit.ly/2wB2hab.

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 when it comes 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 can not be used cavalierly to analyze these data. **?** 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 broad analyses. The first is a simulation where the characteristics of the network are known. This shows that when there are unobserved dependencies, the AME approach is less biased in terms of parameter estimation compared with standard approach employed in international relations to study dyadic data (i.e., GLM models). The second analysis is a replication of four recent studies that have been published recently using a broad range of dyadic data to draw inferences about international relations. These four 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.[21] In several cases, the new approach overturns the basic findings of the original research. This leads us to speculate that many of the findings in the international relations literature may be fragile in the sense that they can only be obtained under stringent assumptions that are not valid.

It is no longer necessary to assume that the interesting, innate interdependencies in relational data about international relations can be ignored. Nor do they have to be approximated with *ad hoc*, incomplete solutions that purport to control for dependencies (such as 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

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

model that provides more reliable inferences, better out-of-sample predictive performance, and new substantive insights.

Appendix

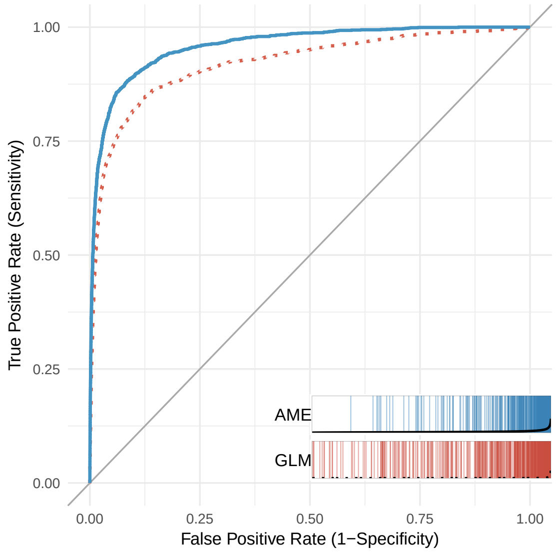
Additional Replication Information. For each of the replications involving a binary dependent variable we provide a table of coefficient estimates that includes the original GLM estimation with a logit link, a GLM estimation with a probit link, and the AME model. The GLM estimation with a probit link function is provided so as to ease comparison between the AME model, which is also based on the probit link.

Additionally, for each replication we provide a more detailed visualization illustrating the results of our out-of-sample performance analysis.

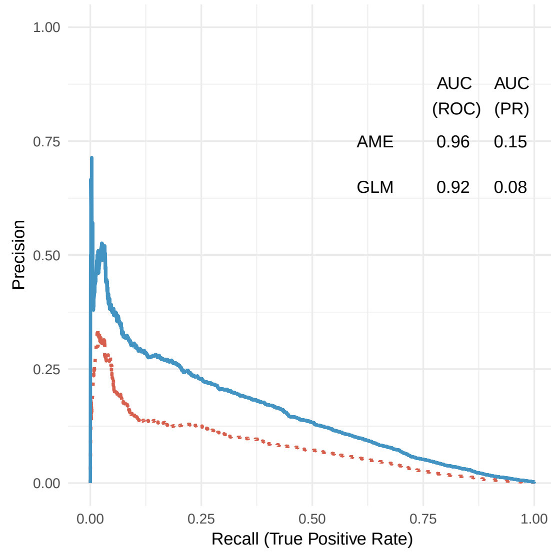
Reiter & Stam (2003). Additional information for the Reiter & Stam (2003) re-estimation.

Variable	GLM (Logit)	GLM (Probit)	AME
Intercept	-4.784^{**}	-2.339^{**}	-3.144^{**}
	(0.097)	(0.034)	(0.06)
Pers/Democ Directed Dyad	1.026^{**}	0.378^{**}	0.255^{**}
	(0.14)	(0.051)	(0.068)
Democ/Pers Directed Dyad	0.083	0.033	0.112
	(0.191)	(0.066)	(0.079)
Personal	0.281	0.15	0.211^{*}
	(0.265)	(0.099)	(0.11)
Military	-0.323	-0.105	-0.025
	(0.574)	(0.204)	(0.249)
Single	-0.677^{**}	-0.261^{**}	-0.07
	(0.144)	(0.062)	(0.073)
Democracy	-1.073^{**}	-0.428^{**}	-0.254^{**}
	(0.194)	(0.07)	(0.063)
Contiguous	2.912^{**}	1.147^{**}	1.296^{**}
	(0.09)	(0.031)	(0.033)
Major Power	2.174^{**}	0.919^{**}	0.906^{**}
	(0.101)	(0.037)	(0.093)
Ally	0.078	-0.003	0.136^{**}
	(0.086)	(0.035)	(0.037)
Higher/Lower Power Ratio	-0.316^{**}	-0.122^{**}	-0.111^{**}
	(0.027)	(0.01)	(0.011)
Economically Advanced	-0.175	-0.054	0.053
	(0.131)	(0.051)	(0.05)
Years Since Last Dispute	-0.381^{**}	-0.149^{**}	-0.129^{**}
	(0.023)	(0.009)	(0.008)
Cubic Spline 1	-0.004^{**}	-0.001^{**}	-0.001^{**}
	(0.000)	(0.000)	(0.000)
Cubic Spline 2	0.002^{**}	0.001^{**}	0.001^{**}
	(0.000)	(0.000)	(0.000)
Cubic Spline 3	-0.001^{**}	0.000^{**}	0.000^{**}
	(0.000)	(0.000)	(0.000)

Table A.1. Parameter comparison for Reiter & Stam (2003). Standard errors in parentheses. ** and * indicate significance at $p < 0.05$ and $p < 0.10$, respectively.



(a) AUC



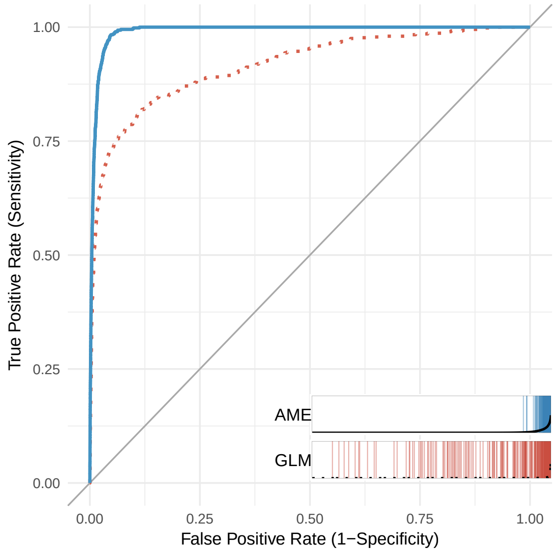
(b) Precision and Recall

Figure A1. Assessments of out-of-sample predictive performance for Reiter & Stam (2003) using ROC curves, PR curves, and separation plots.

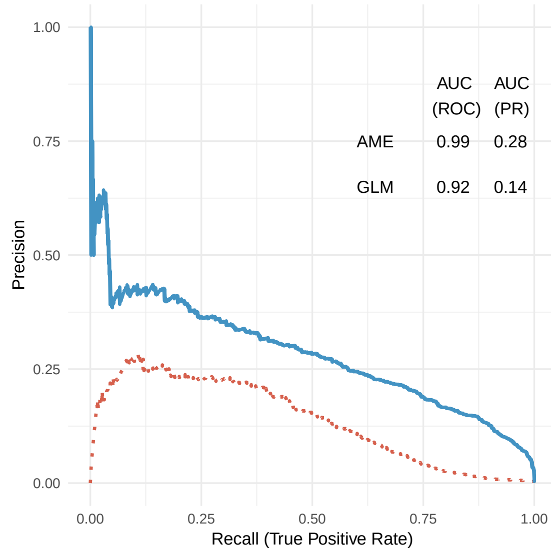
McDonald (2004). Additional information for the McDonald (2004) re-estimation.

Variable	GLM (Logit)	GLM (Probit)	AME
(Intercept)	0.054	0.085	-1.171^{**}
	(1.179)	(0.409)	(0.096)
Spline0	-0.438^{**}	-0.222^{**}	-0.145^{**}
	(0.061)	(0.026)	(0.019)
Spline1	-0.003^{**}	-0.002^{**}	-0.001^{**}
	(0.001)	(0.000)	(0.000)
Spline2	0.001	0.001^{**}	0.000^{*}
	(0.001)	(0.000)	(0.000)
Spline3	0.000	0.000	0.000^{**}
	(0.000)	(0.000)	(0.000)
Shared Alliance	0.483^{**}	0.155	0.342^{**}
	(0.233)	(0.095)	(0.069)
Contiguous	2.011^{**}	0.789^{**}	0.988^{**}
	(0.343)	(0.118)	(0.066)
Log Capabilities Ratio	-0.146^{**}	-0.054^{**}	0.029^{**}
	(0.072)	(0.026)	(0.013)
Trade Dependence	-22.244	-7.051	-13.134^{**}
	(15.184)	(5.536)	(4.938)
Preconflict GDP Change	-6.79^{**}	-3.155^{**}	-2.651^{**}
	(2.033)	(0.788)	(0.574)
Lowest Dyadic Polity Score	-0.036^{**}	-0.014^{**}	-0.026^{**}
	(0.015)	(0.006)	(0.002)
Capabilities	-0.995^{**}	-0.349^{**}	0.022
	(0.377)	(0.14)	(0.079)
Logged GDP	0.000^{**}	0.000^{**}	0.000^{**}
	(0.000)	(0.000)	(0.000)
Logged Cap. Distance	-0.425^{**}	-0.224^{**}	-0.275^{**}
	(0.14)	(0.047)	(0.012)
Major Power In Dyad	0.769^{**}	0.312^{**}	0.212^{**}
	(0.322)	(0.122)	(0.098)
Highest Barrier To Trade	0.024^{**}	0.011^{**}	0.004^{**}
	(0.008)	(0.003)	(0.001)

Table A.2. Parameter comparison for McDonald (2004). Standard errors in parentheses. ** and * indicate significance at $p < 0.05$ and $p < 0.10$, respectively.



(a) AUC



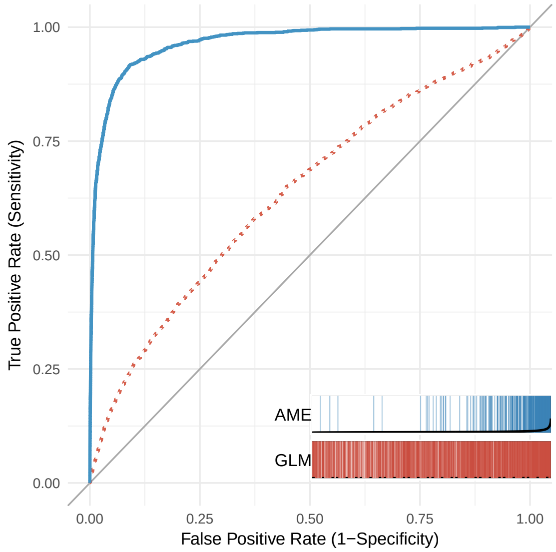
(b) Precision and Recall

Figure A2. Assessments of out-of-sample predictive performance for McDonald (2004) using ROC curves, PR curves, and separation plots.

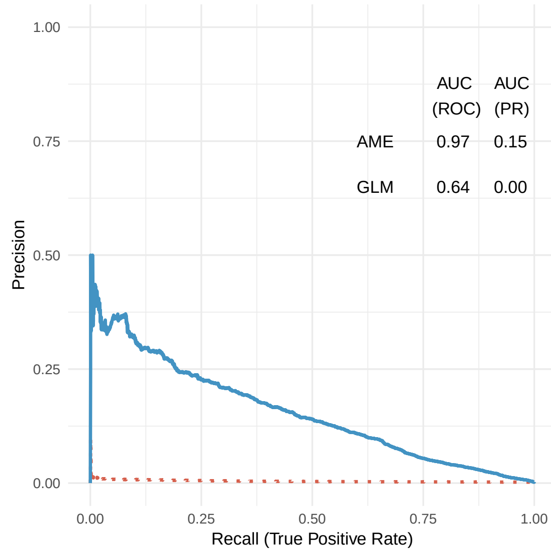
Weeks (2012). Additional information for the Weeks (2012) re-estimation.

Variable	GLM (Logit)	GLM (Probit)	AME
(Intercept)	−3.784**	−1.797**	−2.409**
	(0.423)	(0.159)	(0.132)
Machine	−0.459**	−0.162**	-0.006
	(0.174)	(0.062)	(0.04)
Junta	0.515**	0.194**	0.034
	(0.169)	(0.062)	(0.046)
Boss	0.649**	0.281**	-0.044
	(0.153)	(0.05)	(0.044)
Strongman	0.832**	0.295**	0.032
	(0.132)	(0.048)	(0.044)
Other Type	0.147	0.051	-0.01
	(0.132)	(0.046)	(0.034)
New/Unstable Regime	−0.312**	−0.123**	-0.043
	(0.092)	(0.033)	(0.031)
Democracy Target	0.185	0.052	0.024
	(0.115)	(0.04)	(0.026)
Military Capabilities Initiator	5.234**	2.136**	0.071
	(1.69)	(0.554)	(0.412)
Military Capabilities Target	6.34**	2.865**	−0.969**
	(1.675)	(0.573)	(0.48)
Low Trade Dependence	−24.794*	-8.197	-4.733
	(12.866)	(5.582)	(3.017)
Both Major Powers	1.136**	0.687**	1.122**
	(0.547)	(0.183)	(0.241)
Minor/Major	0.772**	0.292**	0.496**
	(0.239)	(0.086)	(0.118)
Major/Minor	0.711**	0.332**	0.778**
	(0.225)	(0.075)	(0.16)
Contiguous	2.172**	0.738**	0.705**
	(0.32)	(0.125)	(0.06)
Log Dist. Between Capitals	−0.209**	−0.095**	−0.129**
	(0.038)	(0.015)	(0.01)
Alliance Similarity Dyad	−0.999**	−0.386**	-0.073
	(0.144)	(0.05)	(0.065)
Alliance Similarity With System Leader Initiator	0.11	0.011	0.068
	(0.24)	(0.082)	(0.057)
Alliance Similarity Leader Target	0.203	0.032	0.08
	(0.244)	(0.081)	(0.056)
Time Since Last Conflict	−0.229**	−0.089**	−0.067**
	(0.018)	(0.007)	(0.007)
Spline1	−0.001**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)
Spline2	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)
Spline3	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)

Table A.3. Parameter comparison for Weeks (2012). Standard errors in parentheses. ** and * indicate significance at $p < 0.05$ and $p < 0.10$, respectively.



(a) AUC



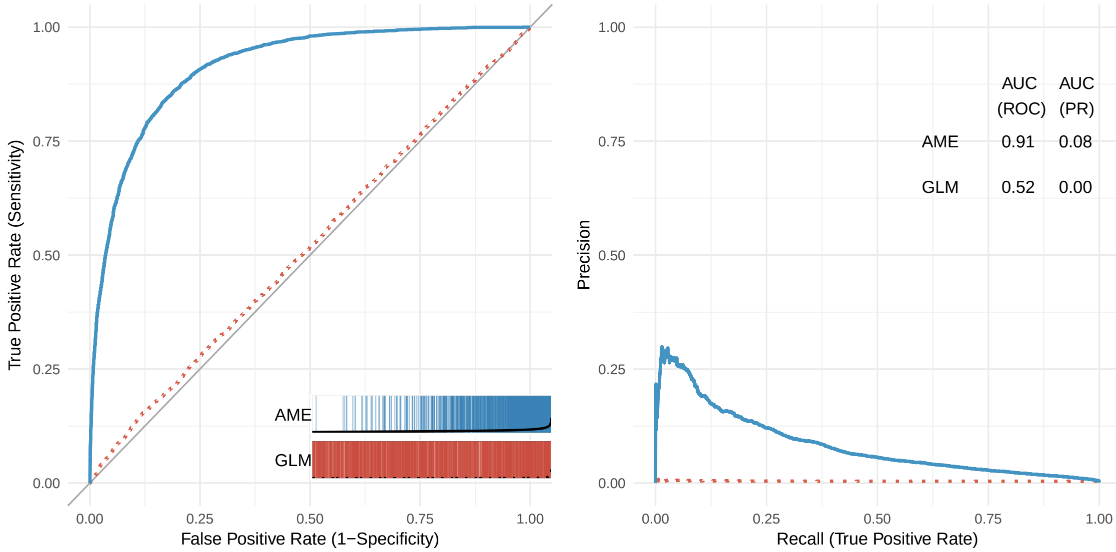
(b) Precision and Recall

Figure A3. Assessments of out-of-sample predictive performance for Weeks (2012) using ROC curves and PR curves

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

Variable	GLM (Logit)	GLM (Probit)	AME
(Intercept)	-5.826^{**}	-2.793^{**}	-2.758^{**}
	(0.366)	(0.366)	(0.045)
Allied	0.133	0.067	0.078^{**}
	(0.102)	(0.102)	(0.021)
Joint Democracy	-0.527^{**}	-0.186^{*}	0.005
	(0.099)	(0.099)	(0.022)
Peace Years	-0.261^{**}	-0.099^{**}	-0.058^{**}
	(0.016)	(0.016)	(0.004)
Spline 1	-0.001^{**}	0.000^{**}	0.000^{**}
	(0.000)	(0.000)	(0.000)
Spline 2	0.000^{**}	0.000^{**}	0.000^{**}
	(0.000)	(0.000)	(0.000)
Spline 3	0.000	0.000	0.000^{**}
	(0.000)	(0.000)	(0.000)
Contiguity	2.427^{**}	0.95^{**}	0.66^{**}
	(0.196)	(0.196)	(0.023)
Parity	-0.77	-0.228	-0.067
	(0.551)	(0.551)	(0.057)
Parity at Entry Year	2.034^{**}	0.739	-0.05
	(0.617)	(0.617)	(0.065)
Rivalry	2.034^{**}	1.035^{**}	0.655^{**}
	(0.213)	(0.213)	(0.028)

Table A.4. Parameter comparison for Gibler (2017). Standard errors in parentheses. ** and * indicate significance at $p < 0.05$ and $p < 0.10$, respectively.



(a) AUC

(b) Precision and Recall

Figure A4. Assessments of out-of-sample predictive performance for Gibler (2017) using ROC curves, PR curves, and separation plots.

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 *startVals0*.

```
# 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
  startVals0 = 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=startVals0,
    periodicSave=TRUE )
  save(ameFit, file=paste0('model_k', latDims[ii],'_v2.rda') )
}
stopCluster(cl)
```

References