

# Bilinear Mixed Effects Models for Dyadic Data

Peter D. Hoff \*

Technical Report no. 430

Department of Statistics

University of Washington

July 2, 2003

## Abstract

This article discusses the use of a symmetric multiplicative interaction effect to capture certain types of third-order dependence patterns often present in social networks and other dyadic datasets. Such an effect, along with standard linear fixed and random effects, is incorporated into a generalized linear model, and a Markov chain Monte Carlo algorithm is provided for Bayesian estimation and inference. In an example analysis of international relations data, accounting for such patterns improves model fit and predictive performance.

*KEY WORDS:* social network, balance, inner product scaling, generalized linear model.

## 1 Introduction

Dyadic data consist of measurements that are made on pairs of objects or under a pair of conditions, so that  $y_{i,j}$  denotes the value of the (possibly directed) measurement from  $i$  to  $j$ . Examples include social network analysis, “round robin” experiments in psychology, and comparative data in which  $y_{i,j}$  might be a measure of similarity between units  $i$  and  $j$ . In the social networks literature, modeling has focused on the binary case where  $y_{i,j}$  is either zero or one, indicating the presence or absence of a “link” from  $i$  to  $j$ . This has led to the development of data analysis tools based on directed graphs and the study of exponentially parameterized random graph models (Wasserman and Pattison 1996). For valued (non-binary) dyadic datasets, a perceived lack of statistical tools has sometimes led to ad-hoc reductions of valued responses to binary data. However, ANOVA methods are available for valued dyadic data: the so-called social relations model (Warner, Kenny, and Stoto 1979; Wong 1982) allows for the decomposition of the variance into sender and receiver

---

\*Peter D. Hoff is Assistant Professor of Statistics, Box 354322, University of Washington, Seattle WA 98195-4322, Email: hoff@stat.washington.edu, Web: www.stat.washington.edu/hoff. This research was supported by Office of Naval Research grant N00014-02-1-1011. The author thanks Mark Handcock and Michael Ward for helpful discussions.

specific effects, as well as allows for correlation between responses within a dyad. Such a model has been studied in the context of a linear group symmetry model by Li (2002), and advances in variance component analysis have been made by and Gill and Swartz (2001) and Li and Loken (2002). These models generally presume normally distributed data and additive effects, and thus the lack of any sort of dependence beyond those specified by second-order moments. In contrast, many observed dyadic datasets exhibit certain forms of third-order dependence, and often it is of scientific interest to quantify these higher order patterns.

In this article we propose a class of generalized additive models based on the social relations model, but incorporate third order dependence via a bilinear effect. The bilinear effect for a pair  $(i, j)$  is simply the inner product of unobserved characteristic vectors  $z_i$  and  $z_j$ , specific to units  $i$  and  $j$  respectively. This approach is similar in spirit to the latent variable methods proposed by Hoff, Raftery, and Handcock (2002) to capture transitivity in a social network dataset, but has some computational and conceptual advantages. The bilinear effect is also a type of multiplicative interaction (Gabriel 1978; Marasinghe and Johnson 1982; Oman 1991). The models presented in this article are similar to the generalized bilinear regression models studied by Gabriel (1998), who considered approximate maximum likelihood estimation in the context of factorial designs. In this article, we show how a bilinear effect can be used to represent certain forms of dependence often seen in dyadic data, and develop a Markov chain Monte Carlo algorithm based on Gibbs sampling, providing arbitrarily exact Bayesian inference. With some modifications, the algorithm can be used as a means of making Bayesian inference for a broad class of generalized bilinear regression models with mixed effects.

In the next section, we discuss the basic linear mixed effects model for dyadic data and the resulting dependence structure. In Section 3, we discuss types of third-order dependence often seen in network datasets and the use of a bilinear effect to capture such dependence. Section 4 gives a Markov chain Monte Carlo (MCMC) algorithm which can be used to obtain samples from the posterior distribution of the parameters. Issues such as model fit, model selection and interpretation are discussed in the context of a data analysis on international relations in Section 5. A discussion follows in Section 6.

## 2 Linear Mixed Effects Models for Exchangeable Dyadic Data

Suppose we are only interested in estimating the linear relationships between responses  $y_{i,j}$  and a possibly vector valued set of variables  $x_{i,j}$ , which could include characteristics of unit  $i$ , characteristics of unit  $j$ , or characteristics specific to the pair. In this case we might consider the regression model

$$y_{i,j} = \beta' x_{i,j} + \epsilon_{i,j}, \tag{1}$$

where  $y_{i,i}$  is typically not defined. The generalized least squares estimate  $\hat{\beta}$  and its covariance matrix depend on the joint distribution of the  $\epsilon_{i,j}$ 's only through their covariance. It is often

assumed in regression problems that the regressors  $x_{i,j}$  contain enough information so that the distribution of the errors is invariant under permutations of the unit labels. This assumption is equivalent to the  $n \times n$  matrix of errors (with an undefined diagonal) having a distribution that is invariant under identical row and column permutations, so that  $\{\epsilon_{i,j} : i \neq j\}$  is equal in distribution to  $\{\epsilon_{\pi(i),\pi(j)} : i \neq j\}$  for any permutation  $\pi$  of  $\{1, \dots, n\}$ . This condition is called weak row-and-column exchangeability of an array. For undirected data, such exchangeability implies a “random effects” representation of the errors, in that  $\epsilon_{i,j}$  is equal in distribution to  $f(\mu, a_i, a_j, \gamma_{i,j})$  where  $\mu, a_i, a_j, \gamma_{i,j}$  are independent random variables and  $f$  is a function to be specified (Aldous 1985, Theorem 14.11). If in addition to the above invariance assumption we also model the errors as Gaussian, then the joint distribution can be represented in terms of a linear random effects model. In the more general case of directed observations, we can represent the joint distribution of the  $\epsilon_{i,j}$ ’s as follows:

$$\begin{aligned} \epsilon_{i,j} &= a_i + b_j + \gamma_{i,j} & (2) \\ (a_i, b_i)' &\sim \text{multivariate normal}(0, \Sigma_{a,b}), \quad \Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \\ (\gamma_{i,j}, \gamma_{j,i})' &\sim \text{multivariate normal}(0, \Sigma_\gamma), \quad \Sigma_\gamma = \begin{pmatrix} \sigma_\gamma^2 & \rho\sigma_\gamma^2 \\ \rho\sigma_\gamma^2 & \sigma_\gamma^2 \end{pmatrix}, \end{aligned}$$

with effects otherwise being independent. The covariance structure of the errors (and thus the observations) is as follows:

$$\begin{aligned} E(\epsilon_{i,j}^2) &= \sigma_a^2 + 2\sigma_{ab} + \sigma_b^2 + \sigma_\gamma^2 & E(\epsilon_{i,j}\epsilon_{i,k}) &= \sigma_a^2 \\ E(\epsilon_{i,j}\epsilon_{j,i}) &= \rho\sigma_\gamma^2 + 2\sigma_{ab} & E(\epsilon_{i,j}\epsilon_{k,j}) &= \sigma_b^2 \\ E(\epsilon_{i,j}\epsilon_{k,l}) &= 0 & E(\epsilon_{i,j}\epsilon_{k,i}) &= \sigma_{ab} \end{aligned}$$

and so  $\sigma_a^2$  represents the dependence of observations having a common sender,  $\sigma_b^2$  that of observations having a common receiver, and  $\rho$  represents the correlation of observations within a dyad (often interpreted as “mutuality” or “reciprocity”). This has been called the “social relations” or “round robin” model (Warner et al. 1979; Wong 1982), and is related to a model for diallel cross data used by Cockerham and Weir (1977). The model is a special case of a linear group symmetry model (Andersson and Madsen, 1998), and has been studied in this context by Li (2002). Recent advances in variance component estimation have been made by Gill and Swartz (2001) and Li and Loken (2002).

To analyze responses in particular sample spaces, the error structure described above can be added to a linear predictor in a generalized linear model:

$$\begin{aligned} \theta_{i,j} &= \beta'x_{i,j} + a_i + b_j + \gamma_{i,j} & (3) \\ E(y_{i,j}|\theta_{i,j}) &= g(\theta_{i,j}) \\ p(y_{1,2}, \dots, y_{n,n-1}|\theta_{1,2}, \dots, \theta_{n,n-1}) &= \prod_{i \neq j} p(y_{i,j}|\theta_{i,j}). \end{aligned}$$

This is a generalized linear mixed-effects model with inverse-link function  $g(\theta)$ , in which the observations are modeled as conditionally independent given the random effects, but are unconditionally dependent. The covariance pattern for the observations is given approximately as

$$\begin{aligned} \text{Cov}(y_{i_1,j_1}, y_{i_2,j_2}) &= E[\text{Cov}(y_{i_1,j_1}, y_{i_2,j_2} | \theta_{i_1,j_1}, \theta_{i_2,j_2})] + \text{Cov}[E(y_{i_1,j_1} | \theta_{i_1,j_1}), E(y_{i_2,j_2} | \theta_{i_2,j_2})] \\ &= E[0] + \text{Cov}[g(\theta_{i_1,j_1}), g(\theta_{i_2,j_2})] \\ &\approx \text{Cov}(\theta_{i_1,j_1}, \theta_{i_2,j_2}) \times g'(\beta' x_{i_1,j_1}) g'(\beta' x_{i_2,j_2}), \end{aligned}$$

where the pattern for  $\text{Cov}(\theta_{i_1,j_1}, \theta_{i_2,j_2})$  is the same as that for the  $\epsilon_{i,j}$ 's given above. However, unlike the linear regression case,  $\hat{\beta}$  is not given by linear combinations of the observations, and  $E(\hat{\beta})$  and  $\text{Cov}(\hat{\beta})$  are not functions of only the first and second order moments of the data. Model lack of fit, or third and higher order dependence, will affect our inference on  $\beta$ . Many dyadic datasets exhibit certain forms of third order dependence. Indeed, it is these higher order patterns of dependence that are often of interest, and may also provide information useful for predictive inference.

### 3 Modeling Third Order Dependence Patterns

Some dependence patterns commonly seen in dyadic datasets have been given the descriptive titles of transitivity, balance, and clusterability. In the context of binary data, graph theoretic definitions of these concepts appear in Wasserman and Faust (1994, chapter 6) and are as follows:

**Transitivity:** For directed binary data, an ordered triad  $i, j, k$  is transitive if whenever  $y_{i,j} = 1$  and  $y_{j,k} = 1$ , we have  $y_{i,k} = 1$ , i.e. “a friend of a friend is a friend.”

**Balance:** For signed unordered relations, a triad  $i, j, k$  is said to be balanced if  $y_{i,j} \times y_{j,k} \times y_{k,i} > 0$ . The idea is that if the relationship between  $i$  and  $j$  is “positive” then they will relate to another unit  $k$  in an identical fashion, so that if  $y_{i,j} > 0$  then  $y_{j,k}$  and  $y_{k,i}$  are either both positive or both negative.

**Clusterability:** This is a relaxation of the concept of balance. A triad is clusterable if it is balanced or the relations are all negative. The idea is that a clusterable triad can be divided into groups where the measurements are positive within groups and negative between groups.

In a statistical sense, a dataset will display varying degrees of transitivity, balance, or clusterability. Often it is found that there are more transitive, balanced, or clusterable triads than would be expected under models (2) or (3). Another indication of third order dependence would be if after fitting a regression model and obtaining the residuals  $\hat{\epsilon}_{i,j}$ , the average value of  $\hat{\epsilon}_{i,j} \times \hat{\epsilon}_{j,k} \times \hat{\epsilon}_{k,i}$  is substantially larger than zero, the expected value presumed by model (2).

Hoff et al. (2002) used simple functions of latent characteristic vectors in a fixed effects setting to capture some forms of transitivity, balance, and clusterability. For example, they considered

models in which  $\theta_{i,j} = \beta'x_{i,j} + f(z_i, z_j)$  where  $f(z_i, z_j) = -|z_i - z_j|$  (“the distance model”) or  $f(z_i, z_j) = z_i'z_j/|z_j|$  (“the projection model”). In what follows, we consider a similar approach using the inner product kernel  $f(z_i, z_j) = z_i'z_j$ , and give random and fixed effects interpretations.

Adding the bilinear effect  $z_i'z_j$  to the linear random effects in models (2) and (3) gives

$$\begin{aligned}\epsilon_{i,j} &= a_i + b_j + \gamma_{i,j} + \xi_{i,j} \\ \xi_{i,j} &= z_i'z_j\end{aligned}\tag{4}$$

where the random effects  $a_i, b_j$  and  $\gamma_{i,j}$  are modeled with the multivariate normal distributions described above. We have written  $\xi_{i,j} = z_i'z_j$  to suggest the interpretation of  $z_i'z_j$  as a mean-zero random effect: If the  $z$ ’s are modeled as independent  $k$ -dimensional multivariate normal random vectors with mean zero and covariance matrix  $\Sigma_z$ , then the resulting distribution for the  $\xi$ ’s has the following moment properties:

- $E(\xi_{i,j}) = 0$ ;
- $E(\xi_{i,j}^2) = \text{trace } \Sigma_z^2$ ;
- $E(\xi_{i,j}\xi_{j,k}\xi_{k,i}) = \text{trace } \Sigma_z^3$ ;

with all other second and third order moments equal to zero. Note that an orthogonal transformation of the  $z$ ’s leaves  $z_i'z_i$  invariant, so we can assume  $\Sigma_z$  is a diagonal matrix (otherwise, the off-diagonal terms are non-identifiable). For simplicity we focus on the case  $\Sigma_z = \sigma_z^2 I_{k \times k}$ , for which the above moments are  $0, k\sigma_z^4$ , and  $k\sigma_z^6$  respectively. With  $\xi_{i,j}$  added to the error term, the nonzero second and third order moments are

$$\begin{aligned}E(\epsilon_{i,j}^2) &= \sigma_a^2 + 2\sigma_{ab} + \sigma_b^2 + \sigma_\gamma^2 + k\sigma_z^4 & E(\epsilon_{i,j}\epsilon_{i,k}) &= \sigma_a^2 \\ E(\epsilon_{i,j}\epsilon_{j,i}) &= \rho\sigma_\gamma^2 + 2\sigma_{ab} + k\sigma_z^4 & E(\epsilon_{i,j}\epsilon_{k,j}) &= \sigma_b^2 \\ E(\epsilon_{i,j}\epsilon_{j,k}\epsilon_{k,i}) &= k\sigma_z^6 & E(\epsilon_{i,j}\epsilon_{k,i}) &= \sigma_{ab}.\end{aligned}$$

Thus the effect  $\xi_{i,j} = z_i'z_j$  can be interpreted as a mean-zero random effect able to induce a particular form of third-order dependence often found in dyadic datasets. Marginally, as  $k$  increases the distribution of  $\xi_{i,j}$  will converge to a normal distribution, due to the central limit theorem. Jointly, the Markov dependence graph for the  $\xi$ ’s has two dyads as neighbors if they have at least one unit in common.

Considered as fixed effects, the  $\xi$ ’s can be viewed as interaction terms that are highly constrained due to the functional dependence on the  $z$ ’s. The constraint is easy to visualize in terms of the  $z$ ’s: If  $z_i$  and  $z_j$  are vectors of similar direction and magnitude, then  $z_i'z_k$  and  $z_j'z_k$  will not be too different. This feature can be related to transitivity, which is conceptually a measure of how  $\xi_{i,k}$  is a function of  $\xi_{i,j}$  and  $\xi_{j,k}$ . Considering for the moment  $z$ ’s scaled to have unit length so that

$|z_i - z_j| = \sqrt{2(1 - z'_i z_j)}$ , by the triangle inequality we have

$$1 - z'_i z_k \leq 1 - z'_i z_j + 1 - z'_j z_k + 2\sqrt{(1 - z'_i z_j)(1 - z'_j z_k)}, \quad \text{or}$$

$$\xi_{i,k} \geq \xi_{i,j} + \xi_{j,k} - \left[ 1 + 2\sqrt{(1 - \xi_{i,j})(1 - \xi_{j,k})} \right],$$

which gives a lower bound for  $\xi_{i,k}$  in terms of  $\xi_{i,j}$  and  $\xi_{j,k}$ .

Balance and clusterability describe how similar  $\xi_{i,k}$  and  $\xi_{j,k}$  are as a function of  $\xi_{i,j}$ . For scaled  $z$ 's, we have

$$|\xi_{i,k} - \xi_{j,k}| = |z'_k(z_i - z_j)| \leq |z_k| \times |z_i - z_j| = |z_i - z_j|.$$

Noting that  $z'_i z_j = \cos(\phi_i - \phi_j)$ , where  $\phi_i$  is the angle of  $z_i$  from a fixed axis, we have

$$\begin{aligned} |z_i - z_j| &= 2 \sin[(\phi_i - \phi_j)/2] \\ &= 2 \sin \frac{1}{2} \cos^{-1}(z'_i z_j) \\ &= \sqrt{2(1 - \xi_{i,j})}, \end{aligned}$$

and so  $|\xi_{i,k} - \xi_{j,k}| \leq \sqrt{2(1 - \xi_{i,j})}$ . If  $\xi_{i,j}$  is large, the difference between  $\xi_{i,k}$  and  $\xi_{j,k}$  must be small. If  $\xi_{i,j}$  is negative one, the difference is unconstrained and could range from zero to a maximum of two (in this scaled case).

## 4 Parameter Estimation

In the frequentist setting, approximate estimation for generalized linear mixed effects models often proceeds via Taylor expansions and iteratively reweighted least squares for the fixed effects, along with approximate restricted maximum likelihood estimation for the variance components (Schall 1991; Breslow and Clayton 1993; Wolfinger and O'Connell 1993; McGilchrist 1994). The accuracy of these approximate methods is generally dependent on the sample size, see Booth and Hobert (1998) for a discussion. Gabriel (1998) suggests an algorithm along these lines for the generalized bilinear mixed effects model. Alternatively, Zeger and Karim (1991), Gelfand, Sahu and Carlin (1996), and Natarajan and Kass (2000) have proposed Gibbs sampling approaches to parameter estimation for generalized linear mixed effects models. However, estimation is more difficult for the complicated dependence structure of the random effects in the invariant normal model (2). Gill and Swartz (2001) have proposed a Gibbs sampling scheme for estimation of random effects in the linear case with the identity link, although we have found that their algorithm does not mix well when covariates are included, due to a weak identifiability of the unit level random effects and certain regression coefficients: As discussed in Gelfand, Sahu, and Carlin (1995) the random effects  $a$  and  $b$  will be confounded to a degree with each other and to regression parameters associated with predictors that do not vary across receivers (i.e. sender-specific effects) or across senders (receiver-specific effects). For example, a population-level intercept is one such parameter. To obtain a

“cleaner” partition of the variance and a more efficient MCMC sampling scheme, we decompose  $x_{i,j}$  into  $x_{i,j} = (x_{d,i,j}, x_{s,i}, x_{r,j})$ , i.e. into dyad specific regressors  $x_{d,i,j}$ , sender specific regressors  $x_{s,i}$  and receiver specific regressors  $x_{r,j}$ . The generalized bilinear model is then rewritten as

$$\theta_{i,j} = \beta'_d x_{d,i,j} + (\beta'_s x_{s,i} + a_i) + (\beta'_r x_{r,j} + b_j) + \gamma_{i,j} + z'_i z_j$$

or equivalently

$$\begin{aligned}\theta_{i,j} &= \beta'_d x_{d,i,j} + s_i + r_j + \gamma_{i,j} + z'_i z_j \\ s_i &= \beta'_s x_{s,i} + a_i \\ r_i &= \beta'_r x_{r,i} + b_i.\end{aligned}$$

This parameterization for the linear unit-level effects is similar to the “centered” parameterizations suggested by Gelfand et al. (1995, 1996). Note that an intercept can be thought of as both a sender or receiver specific effect. For symmetry, we include the constant 1/2 at the beginning of each  $x_{s,i}$  and  $x_{r,j}$  vector, and estimate the first components of  $\beta_s$  and  $\beta_r$  as being equal.

Using the above reparameterization for  $\theta_{i,j}$ , we estimate the parameters for the generalized bilinear regression model by constructing a Markov chain in  $\{\beta_d, \beta_s, \beta_r, \Sigma_{ab}, Z, \sigma_z^2, \Sigma_\gamma\}$  (where  $Z$  denotes the  $k \times n$  matrix of latent vectors), having  $p(\beta_d, \beta_s, \beta_r, \Sigma_{ab}, Z, \sigma_z^2, \Sigma_\gamma | Y)$  as the invariant distribution. This is obtained via an algorithm based on Gibbs sampling, which also samples  $s, r$  and the  $\theta$ 's. The basic algorithm is to iterate the following steps:

1. Sample linear effects:
  - (a) Sample  $\beta_d, s, r | \beta_s, \beta_r, \Sigma_{ab}, \Sigma_\gamma, \theta, Z$  (linear regression);
  - (b) Sample  $\beta_s, \beta_r | s, r, \Sigma_{ab}$  (linear regression);
  - (c) Sample  $\Sigma_{ab}$  and  $\Sigma_\gamma$  from their full conditionals.
2. Sample bilinear effects:
  - (a) For  $i = 1, \dots, n$ : sample  $z_i | \{z_j, j \neq i\}, \theta, \beta, s, r, \Sigma_z, \Sigma_\gamma$  (a linear regression);
  - (b) Sample  $\Sigma_z$  from its full conditional.
3. Sample dyad specific parameters: Update  $\{\theta_{i,j}, \theta_{j,i}\}$  using a Metropolis-Hastings step:
  - (a) Propose  $\begin{pmatrix} \theta_{i,j}^* \\ \theta_{j,i}^* \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \beta'_d x_{d,i,j} + a_i + b_j + z'_i z_j \\ \beta'_d x_{d,j,i} + a_j + b_i + z'_j z_i \end{pmatrix}, \Sigma_\gamma\right)$ ;
  - (b) Accept  $\begin{pmatrix} \theta_{i,j}^* \\ \theta_{j,i}^* \end{pmatrix}$  with probability  $\frac{p(y_{i,j} | \theta_{i,j}^*) p(y_{j,i} | \theta_{j,i}^*)}{p(y_{i,j} | \theta_{i,j}) p(y_{j,i} | \theta_{j,i})} \wedge 1$ .

Various combinations of the above steps can be used to estimate different models. The steps in 1 alone provide a Bayesian estimation procedure for the linear regression problem having an error covariance as in (2). Bayesian estimation of the normal bilinear model with the identity link could

proceed by replacing each  $\theta_{i,j}$  with  $y_{i,j}$  and only iterating steps 1 and 2. Estimation of a generalized linear mixed effects model with random effects structure given by (2) could proceed by iterating steps 1 and 3. The full conditional distributions required to perform steps 1 and 2 are given below.

Note that the  $\theta$ 's are essentially unrestricted in the above sampling scheme. At this level the fit is saturated and does not depend on the regressors, at least to the degree that the prior for  $\Sigma_\gamma$  is diffuse. What the MCMC algorithm above provides is essentially a saturated fit for the  $\theta$ 's (although somewhat smoothed by the common variance) and an ANOVA-like decomposition of the  $\theta$ 's into regressor, sender, receiver and inner-product effects.

#### 4.1 Conditional Distributions for the Linear Effects Components:

Noting that  $\theta_{i,j} - z_i' z_j = \beta_d' x_{i,j} + s_i + r_j + \gamma_{i,j}$ , we see that conditional on the  $\theta$ 's and  $z$ 's, the other parameters can be sampled using a standard Bayesian normal-theory regression approach, although with a complicated covariance structure.

**Full conditional of  $(\beta_d, s, r)$ :** Similar to Wong's (1982) approach to the invariant normal model, we let  $u_{i,j} = \theta_{i,j} + \theta_{j,i} - 2z_i' z_j$  and  $v_{i,j} = \theta_{i,j} - \theta_{j,i}$  for  $i < j$ . We then have

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} X_u \\ X_v \end{pmatrix} \begin{pmatrix} \beta_d \\ s \\ r \end{pmatrix} + \begin{pmatrix} \delta_u \\ \delta_v \end{pmatrix}, \quad (5)$$

where  $X_u$  and  $X_v$  are the appropriate design matrices and  $\delta_u$  and  $\delta_v$  are vectors of independent error terms with variances  $\sigma_u^2 = 2\sigma_\gamma^2(1 + \rho)$  and  $\sigma_v^2 = 2\sigma_\gamma^2(1 - \rho)$  respectively. The full conditional distribution of  $(\beta_d, s, r)$  is then proportional to  $p(u, v | \beta_d, s, r, \Sigma_\gamma) \times p(s, r | \beta_s, \beta_r, \Sigma_{ab}) \times p(\beta_d)$ . For a multivariate normal  $(\mu_{\beta_d}, \Sigma_{\beta_d})$  prior distribution on  $\beta_d$ , the term in the exponent of the full conditional is

$$\phi' \left[ \begin{pmatrix} \Sigma_{\beta_d}^{-1} \mu_{\beta_d} \\ \Sigma_{sr}^{-1} X_{sr} \beta_{sr} \end{pmatrix} + X_u' u / \sigma_u^2 + X_v' v / \sigma_v^2 \right] - \frac{1}{2} \phi' \left[ \begin{pmatrix} \Sigma_{\beta_d}^{-1} & 0 \\ 0 & \Sigma_{sr}^{-1} \end{pmatrix} + X_u' X_u / \sigma_u^2 + X_v' X_v / \sigma_v^2 \right] \phi$$

where  $\phi' = (\beta_d' \ s' \ r')$ ,  $X_{sr}$  and  $\beta_{sr}$  are the combined design matrix and regression parameters for  $s$  and  $r$ , and  $\Sigma_{sr}$  is the covariance matrix of  $(s' \ r)'$ , which is easily derived from  $\Sigma_{ab}$ . The conditional distribution is thus multivariate normal  $(\mu, \Sigma)$  where

$$\begin{aligned} \mu &= \Sigma \left[ \begin{pmatrix} \Sigma_{\beta_d}^{-1} \beta_{d0} \\ \Sigma_{sr}^{-1} X_{sr} \beta_{sr} \end{pmatrix} + X_u' u / \sigma_u^2 + X_v' v / \sigma_v^2 \right] \\ \Sigma &= \left[ \begin{pmatrix} \Sigma_{\beta_d}^{-1} & 0 \\ 0 & \Sigma_{sr}^{-1} \end{pmatrix} + X_u' X_u / \sigma_u^2 + X_v' X_v / \sigma_v^2 \right]^{-1}. \end{aligned}$$

Note that the inverse of  $\Sigma_{sr}$  is given by

$$\Sigma_{sr}^{-1} = \begin{pmatrix} (\sigma_b^2 / \Delta) I_{n \times n} & -(\sigma_{ab} / \Delta) I_{n \times n} \\ -(\sigma_{ab} / \Delta) I_{n \times n} & (\sigma_a^2 / \Delta) I_{n \times n} \end{pmatrix}, \quad \Delta = \sigma_a^2 \sigma_b^2 - \sigma_{ab}^2.$$

**Full conditional of  $(\beta_s, \beta_r)$ :** The full conditional of  $(\beta_s, \beta_r)$  is proportional to  $p(s, r | \beta_s, \beta_r, \Sigma_{ab}) \times p(\beta_s, \beta_r)$ . Assuming a multivariate normal  $(\mu_{\beta_{sr}}, \Sigma_{\beta_{sr}})$  prior distribution for the combined regression parameters, the full conditional is a multivariate normal distribution with mean and variance  $(\mu, \Sigma)$  given by

$$\begin{aligned}\mu &= \Sigma \left[ \Sigma_{\beta_{sr}}^{-1} \mu_{\beta_{sr}} + X_{sr} \Sigma_{sr}^{-1} \begin{pmatrix} s \\ r \end{pmatrix} \right] \\ \Sigma &= (\Sigma_{\beta_{s,r}}^{-1} + X'_{sr} \Sigma_{sr}^{-1} X_{sr})^{-1}\end{aligned}$$

**Full conditional of  $\Sigma_{ab}$ :** The full conditional of  $\Sigma_{ab}$  is proportional to  $p(s, r | \beta_s, \beta_r, \Sigma_{ab}) p(\Sigma_{ab})$ . Using a prior distribution of  $\Sigma_{ab} \sim \text{inverse Wishart}(\Sigma_{ab0}, \nu)$  (parameterized so that  $E(\Sigma_{ab}) = \Sigma_{ab0}/(\nu - 3)$ ), the full conditional of  $\Sigma_{ab}$  is  $\Sigma_{ab} | a, b \sim \text{inverse Wishart}(\Sigma_{ab0} + (a \ b)'(a \ b), \nu + n)$ , where  $a = (s - X_s \beta_s)$  and  $b = (r - X_r \beta_r)$ .

**Full conditional of  $\Sigma_\gamma$ :** Using prior distributions of  $\sigma_u^2 \sim \text{inverse gamma}(\alpha_{u1}, \alpha_{u2})$  and  $\sigma_v^2 \sim \text{inverse gamma}(\alpha_{v1}, \alpha_{v2})$ , the full conditionals are given by  $\sigma_u^2 | u \sim \text{inverse gamma}(\alpha_{u1} + \frac{1}{2} \binom{n}{2}, \alpha_{u2} + \frac{1}{2} \sum [u_i - \hat{u}_{i,j}]^2)$  and  $\sigma_v^2 | v \sim \text{inverse gamma}(\alpha_{v1} + \frac{1}{2} \binom{n}{2}, \alpha_{v2} + \frac{1}{2} \sum [v_i - \hat{v}_{i,j}]^2)$ , where  $\hat{u}_{i,j} = E[u_{i,j} | \beta_d, x_{i,j}, s_i, r_j] = \beta'_d(x_{i,j} + x_{j,i}) + s_i + s_j + r_i + r_j$ , and  $\hat{v}_{i,j}$  is given similarly. The covariance matrix  $\Sigma_\gamma$  can be reconstructed from  $\sigma_u^2$  and  $\sigma_v^2$  via  $\sigma_\gamma^2 = (\sigma_u^2 + \sigma_v^2)/4$  and  $\rho = (\sigma_u^2 - \sigma_v^2)/(\sigma_u^2 + \sigma_v^2)$ .

## 4.2 Conditional distributions for the Bilinear Effects Component:

Let  $e_{i,j} = (\theta_{i,j} + \theta_{j,i} - \hat{u}_{i,j})/2$ , the residual of the symmetric part of the matrix of  $\theta$ 's after fitting the linear effects, and let  $\delta_{u,i,j} = \gamma_{i,j} + \gamma_{j,i}$ . Considering the full conditional of  $z_i$ , we have

$$\begin{aligned}e_{i,1} &= z'_i z_1 + \delta_{u,i,1}/2 \\ e_{i,2} &= z'_i z_2 + \delta_{u,i,2}/2 \\ &\vdots \\ e_{i,n} &= z'_i z_n + \delta_{u,i,n}/2,\end{aligned}$$

and we see that sampling  $z_i$  from its full conditional is equivalent to a (Bayesian) linear regression problem. Modeling the  $z$ 's as a priori independent multivariate normal  $(0, \Sigma_z)$  variables, the full conditional of  $z_i$  is multivariate normal  $(\mu, \Sigma)$  with

$$\begin{aligned}\mu &= 4 \Sigma Z_{-i} e_{i,-i} / \sigma_u^2 \\ \Sigma &= (\Sigma_z^{-1} + 4 Z_{-i} Z'_{-i} / \sigma_u^2)^{-1}\end{aligned}$$

where  $Z_{-i}$  denotes the  $k \times (n - 1)$  matrix obtained by removing the  $i$ th column of  $Z$ , and  $e_{i,-i}$  denotes the vector of residuals  $\{e_{i,j} : j \neq i\}$ . Using an inverse-Wishart  $(\Sigma_{z0}, \nu)$  prior, the full

conditional of  $\Sigma_z$  is inverse-Wishart( $\Sigma_{z0} + ZZ', \nu + n$ ). Alternatively, if we restrict  $\Sigma_z$  to be  $\sigma_z^2 I_{k \times k}$  and use an inverse gamma( $\alpha_0, \alpha_1$ ) prior, then the full conditional is given by  $\sigma_z^2 | Z \sim$  inverse gamma( $\alpha_0 + (nk)/2, \alpha_1 + \text{trace}(Z'Z)/2$ ).

## 5 Data Analysis: International Relations in Central Asia

We analyze data on international relations in central Asia as recorded by the Kansas Event Data Project (<http://www.ku.edu/~keds/project.html>) and described by Schrodtt, Simpson, and Gerner (2001). News stories are downloaded from the Reuters Business Briefing Service on Afghanistan, Armenia, Azerbaijan, and the former Soviet Republics of Central Asia, and political interactions between countries are recorded and categorized. We take our response  $y_{i,j}$  to be the total number of “positive” actions reportedly initiated by country  $i$  with target  $j$  from 1992 to 1999 (i.e. after the breakup of the Soviet Union), as recorded by the KEDS project. Positive actions here include such events as approval, endorsement or praise of one government by another, military assistance, formation of alliances, promises of financial or policy support and others (essentially all events having Goldstein scale values greater than 2.5, except cease-fire or ceding of power. See the KEDS project webpage for more details). We include in our population the 99 countries closest in geographic distance to Afghanistan, plus the United States, giving a total of  $n = 100$  countries for analysis. We note that seventeen of the one-hundred countries had zero actions as either initiators or targets of actions over the seven year period.

### 5.1 Data Description

Some descriptive plots of the raw data are given in Figure 1. Panel (a) plots  $\log(1 + \sum_{j:j \neq i} y_{i,j})$  versus  $\log(1 + \sum_{j:j \neq i} y_{j,i})$  for each country  $i$ . The quantities  $\sum_{j:j \neq i} y_{i,j}$  and  $\sum_{j:j \neq i} y_{j,i}$  are typically called the outdegree and indegree of unit  $i$ , respectively. Note the strong correlation, which suggests a large value of  $\sigma_{ab}/(\sigma_a \sigma_b)$  in the random effects model being considered. In panel (b) we plot the log of each country’s outdegree plus one,  $\log(1 + \sum_{j:j \neq i} y_{i,j})$ , versus log population, which suggests a positive relationship between response and population (a plot of log-indegree versus population is similar). In panel (c) we plot the response on a log scale versus the geographic distance in thousands of miles between countries  $i$  and  $j$ . More precisely, this distance is the “minimum distance” between two countries, and is zero if  $i$  and  $j$  share a border. On average, the number of events between two countries decreases as geographic distance increases. This pattern is made more clear by separating out the measurements involving the United States (which are circled).

### 5.2 Model and Priors

Note that the data are from an observational study, and that the data are not randomly sampled. Rather, we have defined a population of units based on geographic distance and have measurements

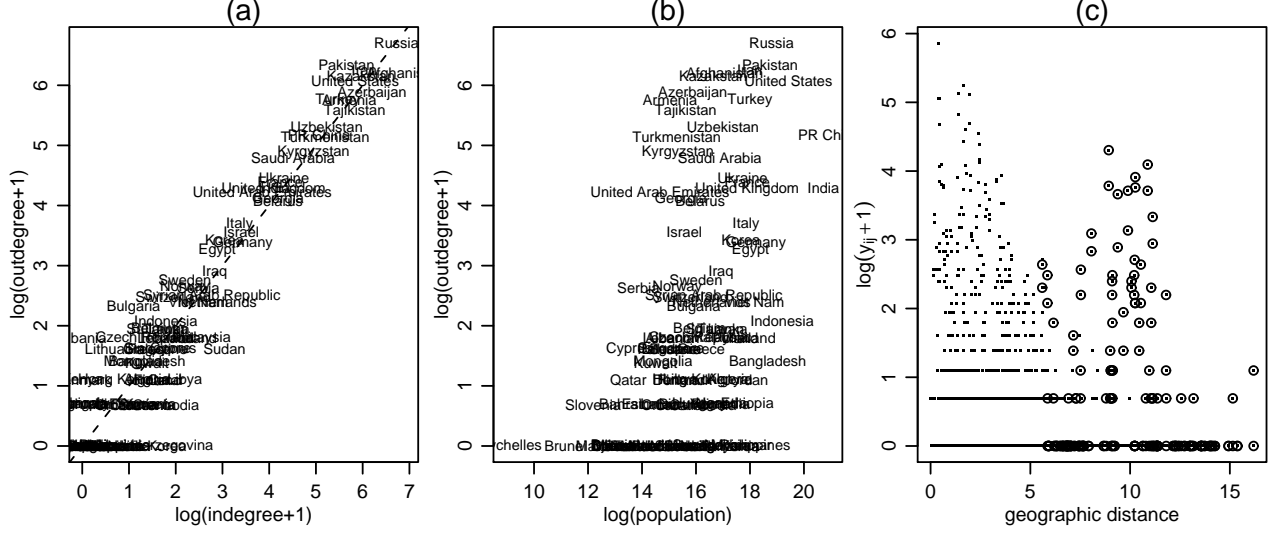


Figure 1: Relationships between (a) Outdegree and indegree; (b) Outdegree and population; (c) Response and geographic distance. Responses involving the United States are circled.

on all pairs. For this analysis, we primarily interpret a probability model as a tool for describing the variance in the dataset, and the regression coefficients as measures of the multiplicative, or log-linear, components of the relationship between response and regressors.

We fit the random effects model (4) to the data using a Poisson distribution and the log-link, so that each response  $y_{i,j}$  is assumed to have come from a Poisson distribution with mean  $e^{\theta_{i,j}}$ , and that the  $y$ 's are conditionally independent given the  $\theta$ 's. We decompose the variance in the  $\theta$ 's as follows:

$$\begin{aligned}\theta_{i,j} &= \beta_0 + \beta_d x_{i,j} + \beta_s x_i + \beta_r x_j + \epsilon_{i,j} \\ \epsilon_{i,j} &= a_i + b_j + \gamma_{i,j} + z'_i z_j,\end{aligned}$$

where  $x_{i,j}$  is the geographic distance between  $i$  and  $j$  and  $x_i$  is the log population of  $i$ . For estimation of variance components, we model the random effects as having the following multivariate normal distributions:  $(a_i, b_i)' \sim \text{MVN}(0, \Sigma_{ab})$ ,  $(\gamma_{i,j}, \gamma_{j,i})' \sim \text{MVN}(0, \Sigma_\gamma)$ ,  $z_i \sim \text{MVN}(0, \sigma_z^2 I_{k \times k})$ . Prior distributions of the parameters are taken to be

- $\beta \sim \text{multivariate normal}(0, 100 \times I_{4 \times 4})$ ;
- $\Sigma_{ab} \sim \text{inverse Wishart}(I_{2 \times 2}, 4)$ ;
- $\sigma_u^2, \sigma_v^2 \sim \text{i.i.d. inverse gamma}(1, 1)$ ,  $\sigma_\gamma^2 = (\sigma_u^2 + \sigma_v^2)/4$ ,  $\rho = (\sigma_u^2 - \sigma_v^2)/(\sigma_u^2 + \sigma_v^2)$ .

Posterior calculations proceed as described in Section 4.

$k$	LLP( $k$ )	$\log p(y \hat{\beta}, \hat{a}, \hat{b}, \hat{Z}, \hat{\Sigma}_\epsilon)$	AIC
0	-3558.78	-2432.54	-2638.54
1	-3351.76	-2316.56	-2622.56
2	-3078.79	-2214.68	-2620.68
3	-3076.73	-2123.49	-2629.49
4	-3077.30	-2038.05	-2644.05

Table 1: Selection of  $k$

### 5.3 Selecting the Latent Dimension:

One issue in model fitting is the selection of the dimension  $k$  of the latent variables  $z$ . Selection of  $k$  could depend on the goal of the analysis. For example, if the goal is descriptive, i.e. the desired end result is a decomposition of the variance into interpretable components, then a choice of  $k = 1, 2$  or  $3$  would allow for a simple graphical presentation of a multiplicative component of the variance. Alternatively, one could examine model fit as a function of  $k$  based on the log-likelihood, or use a cross-validation criterion if one is primarily concerned with predictive performance.

Considering likelihood-based measures of fit, the log-probability of the data given the values of the parameters gets evaluated for each update of the  $\theta$ 's, and so  $\log p(Y|\theta) = \sum_{i \neq j} \log p(y_{i,j}|\theta_{i,j})$  can be calculated with no extra effort. However, such a quantity is not appropriate for selecting between models. As described in Section 4, the model is essentially unrestricted in the  $\theta$ 's, giving a nearly saturated fit which does not depend much on the choice of  $k$  or the regressors (provided the prior for  $\Sigma_\gamma$  is sufficiently diffuse). A likelihood that is more appropriate is the marginal probability of data within a pair,  $\log p(Y|\beta, a, b, Z, \Sigma_\gamma) = \sum_{(i,j)} \log p(y_{i,j}, y_{j,i}|\beta, a_i, b_j, a_j, b_i, z_i, z_j, \Sigma_\gamma)$ , where the sum is over unordered pairs. This is essentially the log-likelihood treating the  $a, b$ , and  $z$ 's as fixed effects. Note that in general  $\log p(y_{i,j}, y_{j,i}|\beta, a_i, b_j, a_j, b_i, z_i, z_j, \Sigma_\gamma)$  is an integral over  $\gamma_{i,j}$  and  $\gamma_{j,i}$  that needs to be approximated, except in the case of the normal model with the identity link.

In some situations the purpose of the model is to make predictions of unobserved data. For example, suppose only a subset of the  $n(n - 1)$  responses were randomly chosen to be measured. As long as we have some measurements for each unit, we can estimate the effects  $a, b$  and  $z$  for each unit and make predictions for missing responses based on these estimates. Although prediction is not the goal for these data, for illustrative purposes we compare the marginal probability criterion discussed above to the following four-fold cross validation procedure:

1. Randomly split the set of ordered pairs  $\{i, j : i \neq j\}$  into four test sets  $A_1, A_2, A_3, A_4$ .
2. For  $k = 0, 1, 2, 3, 4$  :
  - (a) For  $l = 1, 2, 3, 4$  :

- i. perform the MCMC algorithm using only  $\{y_{i,j} : \{i,j\} \notin A_l\}$ , but sample values of  $\theta_{i,j}$  for all ordered pairs.
  - ii. Based on the sampled values of  $\theta_{i,j}$  compute the posterior mean  $\hat{\theta}_{i,j}$  for  $\{i,j\} \in A_l$  and the log predictive probability  $\text{lpp}(A_l) = \sum_{\{i,j\} \in A_l} \log p(y_{i,j} | \hat{\theta}_{i,j})$ .
- (b) Measure the predictive performance for  $k$  as  $\text{LPP}(k) = \sum_{l=1}^4 \text{lpp}(A_l)$ .
3. Select  $k$  based on  $\text{LPP}(k)$ .

For these data, the marginal likelihood and cross-validation criteria for selecting  $k$  are given in Table 2. The cross validation procedure suggests that models having a dimension of  $k = 2, 3$  or 4 have roughly the same predictive performance. In terms of the marginal likelihood criterion, the biggest improvements in fit are in going from  $k = 0$  to  $k = 1$  and from  $k = 1$  to  $k = 2$ . The improvements in fit in going from 2 to 3 and from 3 to 4 dimensions are smaller. Using an AIC-like criterion and penalizing the improvement in likelihood by the number of additional parameters (100 per additional dimension), we would choose  $k = 2$ . Based on these results (and our ability to plot results in two-dimensions) we choose to present the results for the  $k = 2$  model in more detail.

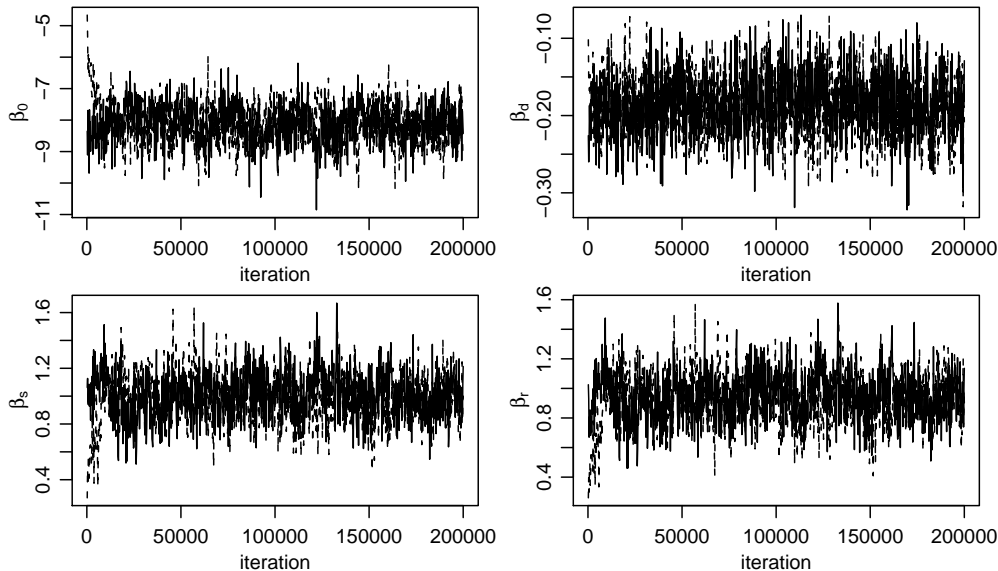


Figure 2: Marginal MCMC output for regression coefficients. Solid lines are from the Markov chain with data-informed starting values, dashed lines from the chain with uninformed starting values.

#### 5.4 Results for $k = 2$

Two Markov chains of length 200,000 each were constructed using the algorithm described above. The first chain used starting values of zero for all regression coefficients and country-specific intercepts, the identity matrix for  $\Sigma_{ab}$  and  $\Sigma_{\gamma}$ , a value of 0.1 for  $\sigma_z^2$ , and components of  $Z$  sampled

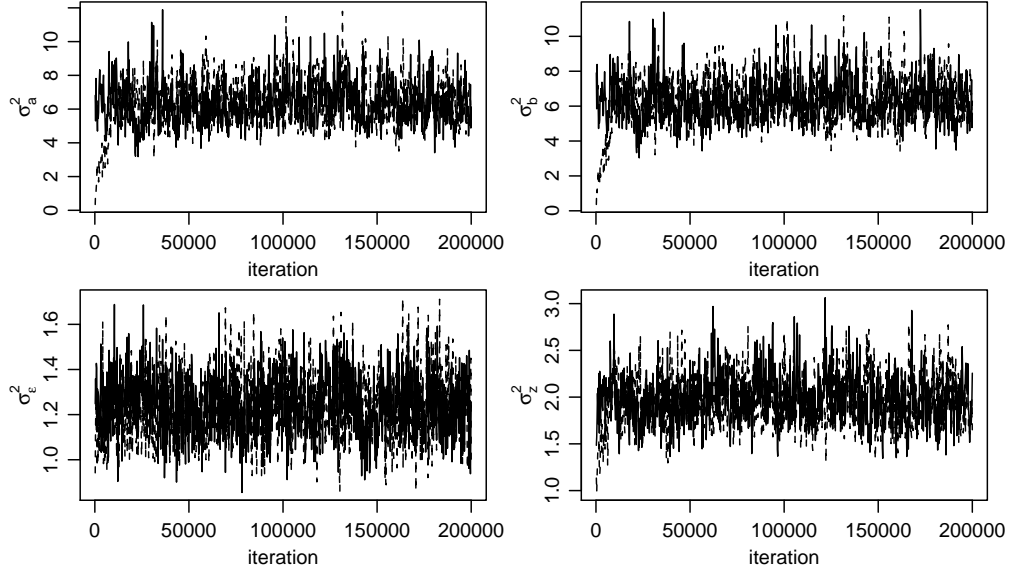


Figure 3: Marginal MCMC output for variance component parameters.

	$\beta_d$	$\beta_s$	$\beta_r$	$\sigma_a^2$	$\sigma_b^2$	$\sigma_{ab}$	$\sigma_\gamma^2$	$\rho$	$\sigma_z^2$
mean	-0.18	1.00	0.94	6.46	6.37	6.4	1.23	0.95	1.99
sd	0.04	0.17	0.17	1.23	1.2	1.21	0.14	0.01	0.27

Table 2: Posterior means and standard deviations for  $k = 2$

independently from a normal  $(0, \sigma_z^2)$  distribution. The second chain used starting values obtained from the following procedure: Maximum likelihood estimates of  $\beta_d$ ,  $s$  and  $r$  were obtained by fitting an ordinary generalized linear model using geographic distance as a regressor and sender and receiver labels as factor variables. Estimates of  $\beta_0$ ,  $\beta_s$ ,  $\beta_r$ , and  $\Sigma_{ab}$  were obtained from the estimates of  $s$  and  $r$ . The iteratively reweighted least-squares fitting procedure produces a matrix  $R$  of working residuals, with the off diagonal elements undefined. An estimate  $\hat{Z}$  of  $Z$  was then obtained by approximating  $R$  with a matrix product of the form  $Z'Z$ . This can be done with an iterative least-squares procedure, similar to the Gibbs sampling procedure outlined in Section 4.2: see ten Berge and Kiers (1989) for more details on this problem. An estimate of  $\Sigma_\gamma$  is then obtained from  $R - \hat{Z}'\hat{Z}$ .

Samples of parameter values were saved from the Markov chains every 100 iterations, and are plotted in Figures 2 and 3. Both chains appear to have achieved stationarity after about 50,000 iterations, and so we base our inference on the saved samples after this point. Posterior means and standard deviations of the model parameters, based on the 3000 saved MCMC samples (1500 from each chain), are given in Table 2. As in the raw data, we see a negative relation between response and geographic distance ( $E[\beta_d|y] = -0.18$ ), and a positive relation between response and country

populations ( $E[\beta_s|y] = 1.00, E[\beta_r|y] = 0.94$ ). We also estimate a strong positive correlation of within-dyad responses as well as the within-country random effects  $a$  and  $b$ .

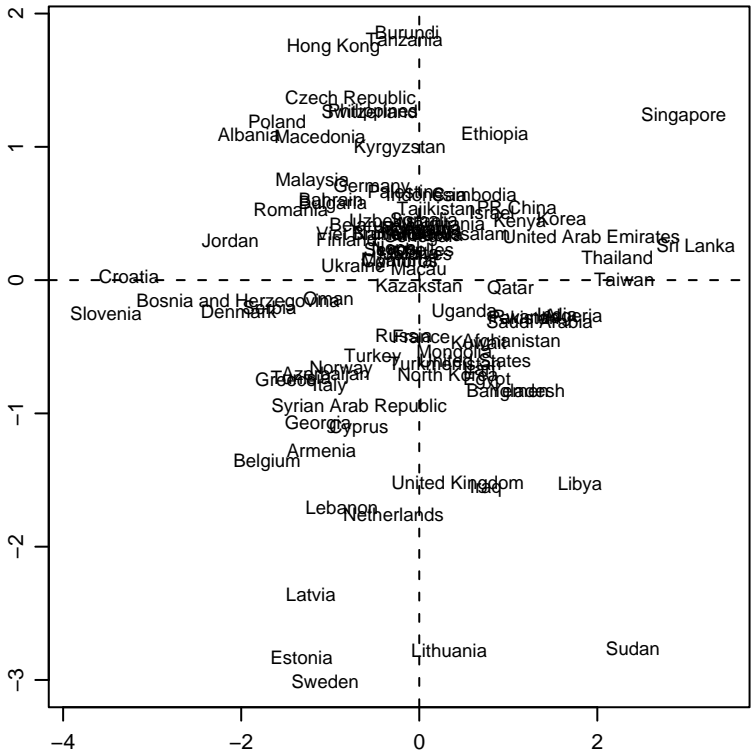


Figure 4: Posterior mean of  $Z$

Next, we analyze the posterior distribution of the the  $k \times n$  matrix of latent vectors  $Z$ . Note that the probability model depends on  $Z$  only through the matrix of inner products  $Z'Z$ , which is invariant under rotations and reflections of  $Z$ . Therefore,  $\log \Pr(Y|Z, \beta, X) = \log \Pr(Y|Z^*, \beta, X)$  for any  $Z^*$  which is equivalent to  $Z$  under the operations of rotation or reflection. Values of  $Z$  sampled from the posterior distribution may seem at first to be highly variable, but perhaps are nearly rotations of each other and are thus not highly variable in terms of the resulting inner product matrices. To appropriately compare sample values of  $Z$ , we must first rotate them to a common orientation. For these data this is done using a “Procrustean” transformation (Sibson 1978), in which for each sample  $Z$  we find the rotation  $Z^*$  of  $Z$  that has the smallest sum of squared deviations from an arbitrary fixed reference matrix  $Z_0$ . The rotated matrix  $Z^*$  which minimizes the sum of squares is given by  $Z^* = Z_0 Z' (Z Z_0' Z_0 Z')^{-1/2} Z$ . See Hoff et al. (2002) for further discussion. The resulting mean of  $Z^*$  is given in Figure 4. Marginal uncertainty in the  $z$ ’s could be displayed by plotting sample  $z^*$ ’s over the plot of the means, using colors to distinguish between countries.

Generally, two countries will be modeled as having  $z$ ’s in the same direction if they have large responses to one another relative to their total number of actions and covariate values, and/or if their responses involving other countries are similar (a model which can distinguish between these

two phenomena is proposed in the discussion). For example, Croatia and Slovenia are each recorded as the initiator of an action with the other as a target, and each initiates an action with Serbia as well. With the exception of one action from Slovenia to Italy, these are the only events recorded for Croatia and Slovenia, and so these countries are “similar” in that they have actions involving each other and to Serbia, and only one other action involving another country. Bosnia-Herzegovina and Denmark have no actions with Croatia or Slovenia, but like Croatia and Slovenia they each have one action with Serbia and very few actions otherwise (each has one action with Azerbaijan, and no other actions), and are thus located in a similar direction. Serbia, although active with this group of countries (on the scale of their response rates), has actions with 10 other countries, and is therefore placed more towards the center. Of course, the posterior variances of the  $z$ ’s for Croatia, Slovenia, Bosnia-Herzegovina, and Denmark are quite high, as our information about them is coming primarily from the few nonzero responses among them.

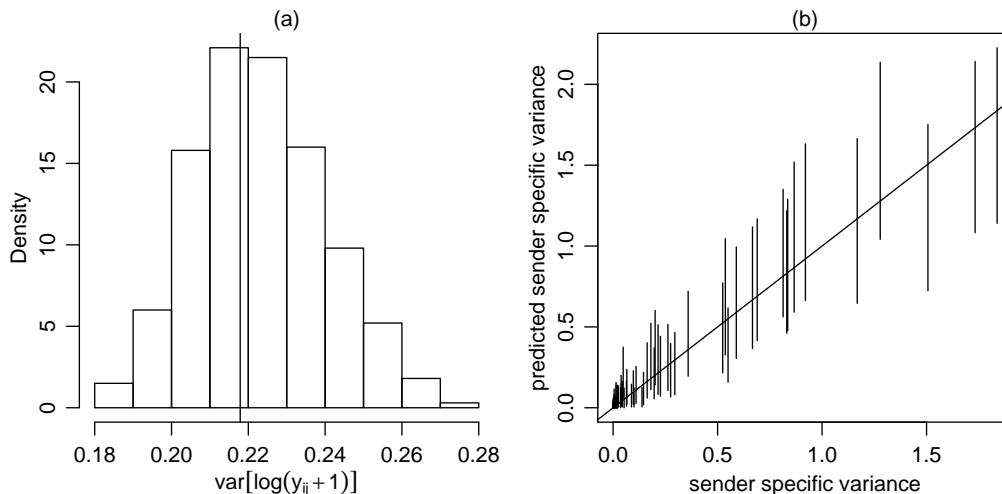


Figure 5: Goodness of fit tests: (a) Posterior predictive distribution of population variance. (b) Posterior predictive confidence regions for country-specific variance in action initiation.

Finally, we evaluate some aspects of model adequacy via goodness of fit statistics. This is done by comparing the observed value of a statistic of interest  $T(Y)$  to its posterior predictive distribution  $p(T(Y_{\text{pred}})|Y)$ . Samples from the posterior predictive distribution are obtained by simulating datasets using the parameters sampled by the Markov chain (see, for example Gelman, Carlin, Stern and Rubin 1995 chapter 6).

In the present case we might be interested in any over or under dispersion of the data relative to the Poisson model. We evaluate any such lack of fit by considering as test statistics the overall sample variance of  $\log(y_{i,j} + 1)$ , as well as the sample variance of  $\{\log(y_{i,j} + 1) : j \neq i\}$  for each  $i$ , that is, the variance of responses from each sender, on a log scale. The posterior predictive distributions of these quantities were estimated by sub-sampling 1000 values of  $(\beta_d, s, r, Z, \Sigma_\gamma)$  from the two

Markov chains, generating a dataset from each sub-sampled set of parameter values, and then computing the statistics from each generated dataset. The results are plotted in Figure 5. The first panel gives a histogram of 1000 samples from the posterior predictive distribution of the overall variance. The posterior predictive distribution is centered around the observed overall variance, given by the vertical line, and no lack of fit is indicated by this statistic. The second panel of Figure 5 plots the observed sender-specific variances for each country versus a 95% posterior predictive interval for that quantity. The confidence intervals contain the observed values for 97 of the 100 countries, and thus do not indicate much lack-of-fit. The Poisson model seems to fit the variance in response reasonably well, at least in terms of these statistics.

## 6 Discussion

This article has presented an approach to modeling third order dependence patterns often seen in dyadic datasets, such as social networks. The models are based on generalized linear mixed effects models with the addition of a reduced-rank interaction term composed of inner products of latent characteristic vectors. Such an approach allows for the analysis of dyadic data using familiar regression tools, but also allows one to capture patterns such as transitivity, balance, and clusterability which are often of interest to social science researchers. Other approaches to capturing such dependence patterns have used metric distances (Hoff et al. 2002) and ultrametric distances (Schweinberger and Snijders, 2003), although not in the presence of the covariance structure (2). While such latent distance models may be easy to understand, the inner-product approach has some conceptual appeal, as the term  $z_i'z_j$  can be viewed as a mean-zero random effect.

Another dependence pattern often of interest to researchers is that of stochastic equivalence, in which two units  $i$  and  $j$  are said to be stochastically equivalent if their responses have the same probability distribution, i.e.  $p(y_{i,1}, \dots, y_{i,n}) = p(y_{j,1}, \dots, y_{j,n})$ . The model considered in this paper, as well as the latent distance approaches mentioned above, potentially confound stochastic equivalence patterns with those of clusterability and balance: two units will generally be estimated to have similar latent characteristic vectors if they have strong relations to each other, or have similar relations to others unit units. However, in some datasets there may be clusters of units that relate similarly to others, but not strongly to each other. Nowicki and Snijders (2001) considered a latent class model which identified clusters of such stochastically equivalent units, but did not separately consider clustering based on strength of relations. A possible approach to modeling both types of patterns is to extend the bilinear effect discussed in this paper to a more general asymmetric bilinear effect such as  $z_i'Rz_j$ , where  $R$  is a  $k \times k$  matrix. Estimation of similar types of effects has been considered by Gabriel (1998), and least squares representations of an asymmetric matrix  $Y$  by  $Z'RZ$  has been considered by ten Berge and Kiers (1989), Kiers (1989) and Trendafilov (2002), among others. In the present application, the vector  $z_i$  could be interpreted as giving grades of membership for unit  $i$  to each of  $k$  classes, and  $R_{lm}$  as the response rate from class  $l$  to  $m$ .

Interestingly, the restriction of each  $z_i$  to be unity at one component and zero at the others gives a representation of the latent class model of Nowicki and Snijders (2000). Unrestricted estimation of  $z'_i R z_j$ , in the presence of the error structure (2), is a topic of current research by the author.

The data analyzed in Section 5, along with R-functions for implementing the proposed methods, are available at the author's website [www.stat.washington.edu/hoff](http://www.stat.washington.edu/hoff).

## References

- Aldous, D. J. (1985), "Exchangeability and related topics," in *École d'été de probabilités de Saint-Flour, XIII—1983*, vol. 1117 of *Lecture Notes in Math.*, pp. 1–198, Springer, Berlin.
- Andersson, S. and Madsen, J. (1998), "Symmetry and lattice conditional independence in a multivariate normal distribution," *The Annals of Statistics*, 26, 525–572.
- Booth, J. G. and Hobert, J. P. (1998), "Standard errors of prediction in generalized linear mixed models," *Journal of the American Statistical Association*, 93, 262–272.
- Breslow, N. E. and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9–25.
- Cockerham, C. C. and Weir, B. S. (1977), "Quadratic analyses of reciprocal crosses," *Biometrics*, 33, 187–204.
- Gabriel, K. R. (1978), "Least squares approximation of matrices by additive and multiplicative models," *Journal of the Royal Statistical Society. Series B. Methodological*, 40, 186–196.
- Gabriel, K. R. (1998), "Generalised bilinear regression," *Biometrika*, 85, 689–700.
- Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995), "Efficient parameterisations for normal linear mixed models," *Biometrika*, 82, 479–488.
- Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1996), "Efficient parametrizations for generalized linear mixed models," in *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pp. 165–180, Oxford Univ. Press, New York.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian data analysis*, Chapman & Hall Ltd, London.
- Gill, P. S. and Swartz, T. B. (2001), "Statistical analyses for round robin interaction data," *The Canadian Journal of Statistics*, 29, 321–331.
- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002), "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, 97, 1090–1098.

- Kiers, H. A. L. (1989), “An alternating least squares algorithm for fitting the two- and three-way DEDICOM model and the IDIOSCAL model,” *Psychometrika*, 54, 515–521.
- Li, H. (2002), “Modeling through group invariance: an interesting example with potential applications,” *The Annals of Statistics*, 30, 1069–1080.
- Li, H. and Loken, E. (2002), “A unified theory of statistical analysis and inference for variance component models for dyadic data,” *Statistica Sinica*, 12, 519–535.
- Marasinghe, M. G. and Johnson, D. E. (1982), “A test of incomplete additivity in the multiplicative interaction model,” *Journal of the American Statistical Association*, 77, 869–877.
- McGilchrist, C. A. (1994), “Estimation in generalized mixed models,” *Journal of the Royal Statistical Society. Series B. Methodological*, 56, 61–69.
- Natarajan, R. and Kass, R. E. (2000), “Reference Bayesian methods for generalized linear mixed models,” *Journal of the American Statistical Association*, 95, 227–237.
- Nowicki, K. and Snijders, T. A. B. (2001), “Estimation and prediction for stochastic blockstructures,” *Journal of the American Statistical Association*, 96, 1077–1087.
- Oman, S. D. (1991), “Multiplicative effects in mixed model analysis of variance,” *Biometrika*, 78, 729–739.
- Schall, R. (1991), “Estimation in generalized linear models with random effects,” *Biometrika*, 78, 719–727.
- Schrodt, P. A., Simpson, E. M., and Gerner, D. J. (2001), “Monitoring conflict using automated coding of newswire sources,” Paper presented at the High-Level Scientific Conference on Identifying Wars, Uppsala University, Uppsala, Sweden. [http://www.pcr.uu.se/Schrodt\\_Uppsala.pdf](http://www.pcr.uu.se/Schrodt_Uppsala.pdf).
- Schweinberger, M. and Snijders, T. (2003), “Settings in social networks: A measurement model.” *Submitted*.
- Sibson, R. (1978), “Studies in the robustness of multidimensional scaling,” *Journal of the Royal Statistical Society, Series B, Methodological*, 40, 234–238.
- ten Berge, J. M. F. and Kiers, H. A. L. (1989), “Fitting the off-diagonal DEDICOM model in the least-squares sense by a generalization of the Harman and Jones MINRES procedure of factor analysis,” *Psychometrika*, 54, 333–337.
- Trendafilov, N. T. (2002), “GIPSCAL revisited. A projected gradient approach,” *Statistics and Computing*, 12, 135–145.

- Warner, R., Kenny, D. A., and Stoto, M. (1979), “A new round robin analysis of variance for social interaction data,” *Journal of Personality and Social Psychology*, 37, 1742–1757.
- Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge.
- Wasserman, S. and Pattison, P. (1996), “Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and  $p^*$ ,” *Psychometrika*, 61, 401–425.
- Wolfinger, R. and O’Connell, M. (1993), “Generalized linear mixed models: A pseudo-likelihood approach,” *Journal of Statistical Computation and Simulation*, 48, 233–243.
- Wong, G. Y. (1982), “Round robin analysis of variance via maximum likelihood,” *Journal of the American Statistical Association*, 77, 714–724.
- Zeger, S. L. and Karim, M. R. (1991), “Generalized linear models with random effects: A Gibbs sampling approach,” *Journal of the American Statistical Association*, 86, 79–86.