Standard errors for regression on relational data with exchangeable errors

Frank W. Marrs*  
Department of Statistics  
Colorado State University

Tyler H. McCormick  
Departments of Statistics and Sociology  
University of Washington

Bailey K. Fosdick  
Department of Statistics  
Colorado State University

Abstract

Relational arrays represent interactions or associations between pairs of actors, often in varied contexts or over time. Such data appear as, for example, trade flows between countries, financial transactions between individuals, contact frequencies between school children in classrooms, and dynamic protein-protein interactions. This paper proposes and evaluates a new class of parameter standard errors for models that represent elements of a relational array as a linear function of observable covariates. Uncertainty estimates for regression coefficients must account for both heterogeneity across actors and dependence arising from relations involving the same actor. Existing estimators of parameter standard errors that recognize such relational dependence rely on estimating extremely complex, heterogeneous structure across actors. Leveraging an exchangeability assumption, we derive parsimonious standard error estimators that pool information across actors and are substantially more accurate than existing estimators in a variety of settings. This exchangeability assumption is pervasive in network and array models in the statistics literature, but not previously considered when adjusting for dependence in a regression setting with relational data. We show that our estimator is consistent and demonstrate improvements in inference through simulation and a data set involving international trade.

Keywords: array data, confidence intervals, dependent data, estimating equations

*The authors contributed equally to this work and are listed in random order. Correspondence contacts: frank.marrs@colostate.edu and tylermc@uw.edu.
1 Introduction

Measurable relationships between pairs of actors are often represented as relational arrays. A relational array \( Y = \left\{ y_{ij}^{(r)} : i, j \in \{1, ..., n\}, i \neq j, r \in \{1, ..., R\} \right\} \) is composed of a series of \( R \) \((n \times n)\) matrices, each of which describes the directed pairwise relationships among \( n \) actors of type \( r \), e.g. time period \( r \) or relation context \( r \). The diagonal elements of each matrix \( \{y_{ii}^{(r)} : i \in \{1, ..., n\}\} \) are assumed to be undefined, as an actor cannot have a relationship with his/herself. Examples of data that can be represented as a relational array include flows of migrants between countries and daily interactions among school children over the course of a week. In economics, relational arrays can also describe monetary transfers between individuals as part of informal insurance markets (see, for example, Bardham (1984); Fafchamps (2006); Foster and Rosenzweig (2001); Attanasio et al. (2012); Banerjee et al. (2013)). Data sets represented as a single matrix of relations, i.e. \( R = 1 \), are also considered here. These data sets are extremely common in the social and biological sciences, and typically the array \( Y \) is then simply referred to as a (weighted) network.

We consider regression models that express the entries in a relational array as a linear function of observable covariates:

\[
y_{ij}^{(r)} = \beta^T x_{ij}^{(r)} + \xi_{ij}^{(r)}, \quad i, j \in \{1, ..., n\}, i \neq j, r \in \{1, ..., R\}.
\]  

(1)

Here \( y_{ij}^{(r)} \) is a (continuous) directed measure of the \( r \)th relation from actor \( i \) to actor \( j \) and \( x_{ij}^{(r)} \) is a \((p \times 1)\) vector of covariates, which are unrelated (i.e. exogenous) to the mean-zero error \( \xi_{ij}^{(r)} \). In a study on international trade, \( y_{ij}^{(r)} \) may denote the value of trade exported from country \( i \) to country \( j \) in year \( r \) and the covariates may include country-specific attributes such as GDP and population, as well as country pair characteristics such as geographic distance. We particularly consider applications where inference for the coefficient vector \( \beta \) is the primary goal in the analysis. Taking again informal insurance markets, Fafchamps and Gubert (2007) examine how covariates such as geographical proximity and kinship related
to risk sharing relations after economic shocks. Recent extensions of this work (e.g. Aker (2010); Blumenstock et al. (2011); Jack and Suri (2014)) explore the strength of the association between physical proximity and financial transactions among individuals with access to mobile phones. Throughout the paper, we assume that relations are directed such that the relationship from actor $i$ to actor $j$ may differ than that from $j$ to $i$. However, the methods we propose extend to the undirected/symmetric relation case in a straightforward manner. We discuss the extension to undirected arrays in the online supplement.

A core statistical challenge in modeling relational arrays arises from the innate dependencies among relations involving the same actor. For example, dependence often exists between trade relations involving the same country and between economic transfers originating from the same individual. We also expect there to be dependence in the errors that is not captured by observable covariates. This may arise, for example, from differences in gregariousness between individuals or, notably in the informal insurance markets, individual differences in risk aversion. Substantial dependence in the errors precludes the use of standard regression techniques for inference. While unbiased estimation for the coefficients $\beta$ in (1) is possible via ordinary least squares (OLS), accurate uncertainty quantification for $\beta$ (i.e. standard errors) requires consideration and estimation of any auxiliary dependence. Approaches for addressing this challenge have appeared in the statistics, biostatistics, and econometrics literatures and can be characterized into two broad classes.

The first set of approaches impose a parametric model on the errors. Specifically, they either use latent variables to model the array measurements as conditionally independent given the latent structure (e.g. see Holland et al. (1983); Wang and Wong (1987); Hoff et al. (2002); Hoff (2005)) or model the error covariance structure directly subject to a set of simplifying assumptions (e.g. see Hoff et al. (2011); Fosdick and Hoff (2014); Hoff et al. (2015)). While these methods provide parsimonious representations of the underlying error structure, the accuracy of inference on $\beta$ depends on the extent to which the true error structure is consistent with the specified parametric model.
The second approach to accounting for error dependence relies heavily on empirical estimates of the error structure based on the residuals in an estimating equation/moment condition framework. In contrast to the first approach, this framework typically makes few assumptions about the data generating process and utilizes a sandwich covariance estimator for the standard errors of the regression coefficients. Sandwich estimators employ the regression residuals to “adjust” the standard error estimate in case the moment conditions are misspecified or there is dependence structure within the errors. As a result, the sandwich estimator is commonly known as a robust estimator of the standard error. The quality of this correction depends on the accuracy of the error covariance estimate based on the residuals. In practice, current error covariance estimators for relational regression are hindered by the need to estimate a tremendous number of covariance parameters with minimal, noisy observations. These practical limitations have been recognized in other contexts (see King and Roberts (2014) for a discussion) and is the reason why Wakefield (2013) suggests such estimators be labeled “empirical” rather than “robust.”

In this paper, we extend the estimating equation/moment condition framework by incorporating an assumption implicit in many of the model-based approaches. Let $Y^{(r)}$ denote the $r^{th}$ $(n \times n)$ matrix slice in the array, containing all relations of type $r$. We propose leveraging an exchangeability assumption within, and potentially across, each matrix $Y^{(r)}$ to derive parsimonious estimates of the relational dependence. Our approach produces a dramatically simplified estimator that results in superior performance in inference, which we demonstrate through both theoretical and empirical studies.

This paper is organized as follows. The remainder of this section provides background on the estimating equation framework in the context of relational data. Section 2 describes current inference approaches arising from the network econometrics literature and literature on moment condition estimators with cross-sectional dependence focusing specifically on data sets where $R = 1$ (e.g. Conley (1999); Hansen (2015)). We discuss what it means for relational data to be exchangeable in Section 3 and present our proposed covariance matrix
estimator based on an exchangeability assumption. Section 4 describes the efficiency and bias of our proposed method compared to the current state of practice, as supported by extensive simulation evidence and theoretical results. We discuss extensions of our method for use with arrays with \( R > 1 \) in Section 5 and demonstrate our methodology using a data set of international trade flow over multiple decades in Section 6. We conclude with a discussion in Section 7.

1.1 Accounting for correlated errors in relational regression

A key statistical challenge in relational regression is accounting for the correlation structure present in the \( n \times n \times R \) array of error terms \( \Xi = \{ \xi_{ij}^{(r)} : i, j \in \{1, \ldots, n\}, i \neq j, r \in \{1, \ldots, R\} \} \).

First, consider a single matrix of relations \( Y^{(r)} = \{ y_{ij}^{(r)} : i, j \in \{1, \ldots, n\}, i \neq j \} \) and error matrix \( \Xi^{(r)} \) corresponding to relation type \( r \). There are two primary types of correlation we might expect among the errors. The first type is between relations within the same row or within the same column of the matrix \( \Xi^{(r)} \). Revisiting the international trade example, this dependence corresponds to correlation among a country’s exports (i.e. within a rows of \( \Xi^{(r)} \)) and correlation among a country’s imports (i.e. within a column of \( \Xi^{(r)} \)). These dependence patterns are often seen in array data in general, even when each dimension of the array is distinct (Hoff et al. (2011); Fosdick and Hoff (2014)). The second type of correlation we expect, which is specific to relational data, stems from the fact that the row and column index sets represent the same entities. Again, in the context of the trade data, we might expect France’s exports to Germany to depend upon the amount Spain exports to France. This corresponds to dependence between errors, say, \( \xi_{ij}^{(r)} \) and \( \xi_{ki}^{(r)} \).

We use an estimating equations/moment conditions framework (see, for example, Wakefield (2013); Hansen (2015)) to perform inference on \( \beta \). In relational regression, estimating equations \( g \) are defined such that for all \((i, j, r)\), \( \mathbb{E} \left[ g(y_{ij}^{(r)}, \beta) \right] = 0_p \), where \( 0_p \) is the \( p\)-
The estimator $\hat{\beta}$ is then defined as that which satisfies

$$G(Y, \hat{\beta}) := \sum_{i,j,r} g(y_{ij}^{(r)}, \hat{\beta}) = 0_p. \quad (2)$$

The estimating equations $g$ characterize specific features of the population distribution (e.g. the first moment), but critically, this approach does not fully specify the population distribution.

Consider the relational regression model as defined in (1). There are many $g$ functions one could specify which would provide reasonable $\beta$ estimates. One common specification is (see, for example, Chapter 11 in Hansen (2015) or Chapter 5 in Wakefield (2013))

$$g(y_{ij}^{(r)}, \beta) = x_{ij}^{(r)} \left( y_{ij}^{(r)} - \beta^T x_{ij}^{(r)} \right) \quad (3)$$

This corresponds to the score function of the multivariate normal likelihood assuming homoskedastic, independent errors and gives rise to the familiar ordinary least squares estimate $\hat{\beta}$: $\hat{\beta} = (X^T X)^{-1} X^T Y_v$, where $X$ is an $(n(n-1)R \times p)$ matrix of covariate vectors $\{x_{ij}^{(r)}\}$ and $Y_v$ is a vectorized representation of $Y$. Under regularity conditions (see Van der Vaart (2000) and Cameron et al. (2011), for example), the estimator satisfying (2) is consistent ($\hat{\beta} \to_p \beta$) and moreover asymptotically normal:

$$\sqrt{n} \left( \hat{\beta} - \beta \right) \to_d N(0_p, A^{-1} B (A^T)^{-1}) \quad (4)$$

where $A = \mathbb{E} \left[ \frac{\partial}{\partial \beta} G(Y, \beta) \right]$ and $B = \mathbb{E} \left[ G(Y, \beta) G(Y, \beta)^T \right]$, such that $G(Y, \beta)$ is as defined in (2). Estimating the asymptotic variance of $\hat{\beta}$ then amounts to estimating $A$ and $B$. Asymptotic covariance estimators of the form $\hat{A}^{-1} \hat{B} (A^T)^{-1}$ are commonly referred to as “sandwich” estimators (Huber (1967); White (1980)). Assuming independence across observations, the
elements of the covariance can be estimated as

\[ \hat{A} = \frac{1}{n(n-1)R} \sum_{i,j,r} \frac{\partial}{\partial \beta^T} g(y_{ij}^{(r)}, \hat{\beta}) \quad \text{and} \quad \hat{B} = \frac{1}{n(n-1)R} \sum_{i,j,r} g(y_{ij}^{(r)}, \hat{\beta}) g(y_{ij}^{(r)}, \hat{\beta})^T. \]

When \( g \) is defined as in (3) for relational data, \( A = X^T X \) and \( B = X^T \Omega X \), where \( \Omega = \text{V}[Y_v|X] \) is the covariance matrix of the relations, equivalently the errors. \( \Omega \) appears in the form of the variance for most \( g \) functions commonly used to estimate \( \beta \) in (1). Independence is violated in relational data as we expect relations involving the same actor(s) will be dependent. More complex covariance structures have been proposed that assume only subsets of the observations be independent. These independent subsets are often specified based on distance metrics derived from observable features of the data (see, e.g., White and Domowitz (1984); Liang and Zeger (1986); Conley (1999)). In the next section, we discuss in detail the extensions proposed for relational data.

2 Dyadic clustering estimator

To facilitate presentation, we first describe the current state-of-the-art sandwich covariance estimation framework with a single relation \( Y^{(1)} \), then move to arrays with \( R > 1 \). For notational simplicity, we presently drop the superscript (1) indexing the relation type and reintroduce it in Section 5 when needed. Thus, \( y_{ij} = y_{ij}^{(1)} \), \( x_{ij} = x_{ij}^{(1)} \), \( Y_v \) is an \( (n(n-1) \times 1) \) vector of relational observations in \( Y^{(1)} \), and \( X \) is the \( (n(n-1) \times p) \) matrix of covariates for these relations.

Consider a ordered pair \((i,j)\) and define \( \Theta_{ij} \) as the set consisting of all ordered pairs that contain an overlapping member with the pair \((i,j)\). In other words, \( \Theta_{ij} = \{ (k,l) : \{i,j\} \cap \{k,l\} \neq \emptyset \} \). Generalizing the standard estimating equation framework, Fafchamps and Gubert (2007), Cameron et al. (2011), and Aronow et al. (2015) propose and describe the properties of a flexible standard error estimator for relational regression which makes the sole assumption that two relations \((i,j)\) and \((k,l)\) are independent if \((i,j)\) and \((k,l)\) do
not share an actor (i.e. \((k, l) \not\in \Theta_{ij}\)). This implies that \(\text{Cov}(y_{ij}, y_{kl} | X) = \text{Cov}(\xi_{ij}, \xi_{kl}) = 0\) for non-overlapping pairs, but places no restrictions on the covariance elements for pairs that involve the same actor. Let \(\Omega_{DC}\) denote the covariance matrix \(V[\mathbf{Y}_v | X]\) subject to this non-overlapping pair independence assumption. [Fafchamps and Gubert (2007)] propose estimating each nonzero entry of \(\Omega_{DC}\) with a product of residuals, e.g. \(\hat{\text{Cov}}(\xi_{ij}, \xi_{ik}) = e_{ij}e_{ik}\). This may be expressed in matrix form as

\[
\hat{\Omega}_{DC} = \mathbf{e}\mathbf{e}^T \circ \mathbf{1}_{\{|i,j\cap\{k,l\}\neq \emptyset\}},
\]

(5)

where \(\mathbf{e}\) is the vector of residuals \(\{e_{ij} = y_{ij} - \hat{\beta}^T \mathbf{x}_{ij}\}\) for all relations, \(\mathbf{1}_{\{|i,j\cap\{k,l\}\neq \emptyset\}}\) is an \((n(n-1) \times n(n-1))\) matrix of indicators denoting which relation pairs share an actor, and ‘\(\circ\)’ denotes the matrix Hadamard (entry-wise) product. The estimator \(\hat{\Omega}_{DC}\) can be seen as that which takes the empirical covariance of the residuals defined by \(\mathbf{e}\mathbf{e}^T\) and systematically introduces zeros to enforce the non-overlapping pair independence assumption. We refer to the covariance estimator \(\hat{\Omega}_{DC}\) as the **dyadic clustering (DC) covariance estimator** as it owes its derivation to the extensive literature on “cluster-robust” standard error estimates. Restricting the covariances in \(\hat{\Omega}_{DC}\) between non-overlapping relations to be zero makes this estimator similar to that resulting from a two-way clustering approach which clusters on each relation sender (i.e. the rows of \(\Xi\)) and also clusters on each relation receiver (i.e. the columns of \(\Xi\)).

When the \(\hat{\beta}\) estimator is that based on ordinary least squares (i.e. that associated with (3)), [Fafchamps and Gubert (2007)] propose a sandwich variance estimator for \(V[\hat{\beta}]\) based on the DC covariance estimator, which is equal to

\[
\hat{V}_{DC} = (X^TX)^{-1}X^T\hat{\Omega}_{DC}X(X^TX)^{-1}.
\]

(6)

We will refer to this as the DC estimator of \(V[\hat{\beta}]\). [Aronow et al. (2015)] show that \(\hat{V}_{DC}\) is consistent by showing that as the number of actors \(n\) grows, the number independent pairs
of actors grows with \( n^4 \) whereas the number of dependent pairs grows with \( n^3 \).

We contend that while the DC estimator of the variance in (6) is widely used, asymptotically consistent, and theoretically robust to a wide range of error dependence structures (making minimal assumptions), its utility is limited in practice for several reasons. First, the DC estimator estimates \( \mathcal{O}(n^3) \) covariance parameters from only \( \mathcal{O}(n^2) \) residuals. Even for relational matrices of moderate size, estimating this number of parameters is onerous computationally and statistically. Second, the DC approach estimates each nonzero covariance element independently with a single residual product: e.g. \( \hat{\text{Cov}}(\xi_{ij}, \xi_{ik}) = e_{ij}e_{ik} \). When there is substantial heterogeneity in the covariance structure, estimating each element individually in this way may be appropriate. However, the variability of these estimates is extreme since each is based on a single observation of the pair.

3 Standard errors with exchangeability

As discussed above, a major issue with the DC estimator defined in (5) and (6) is that it estimates the entries of a complex covariance structure with single products of residuals from the relational regression. In this section, we propose a novel estimator for \( \text{V}[\hat{\beta}] \) that leverages an exchangeability assumption in the estimation of \( \Omega \). In short, this assumption induces structure among portions of the covariance matrix \( \Omega \) corresponding to subsets of the relations with a similar arrangement, and pools information within these subsets. For this section, we continue discussion in terms of data sets containing a single relation \( Y = Y^{(1)} \), and discuss extensions of our proposed methodology to arrays in Section 5.

3.1 Exchangeability in relational models

A common modeling assumption for relational and array structured errors is exchangeability. Defined by de Finetti for a univariate sequence of random variables, exchangeability was generalized to array data and relational data by [Hoover (1979)] and [Aldous (1981)]. The errors
Figure 1: Five distinguishable configurations of relation pairs involving the bold orange relation in an exchangeable relational model: a) reciprocal relations; (b) relations share common receiver; (c) relations share common sender; (d) shared actor is the sender of one relation and receiver of the other; (e) no shared actors among the two relations.

in a relational data model are jointly exchangeable if the probability distribution of the array errors, Ξ, is invariant under any permutation of the rows and columns. Mathematically, this means

\[ P(Ξ) = P(Π(Ξ)), \]

where \( Π(Ξ) = \{ξ_{π(i)π(j)}\} \) is the error array with its rows and columns reordered according to permutation operator \( π \). Intuitively, exchangeability in the context of linear regression on an array simply means the observed covariates are sufficiently informative such the ordering of the row and column labeling in the error array is uninformative. Each of the conditionally independent parametric network models discussed in the introduction have this joint exchangeability property (see Hoff (2008) and Bickel and Chen (2009) for further discussion).

### 3.2 Impact of exchangeability on covariance structure

Under exchangeability, the covariance matrix Ω has at most six unique elements. To see this result intuitively, note that any relation has five distinguishable types of covariance configurations involving another relation, plus one variance term associated with the relation itself. Figure 1 shows the five distinguishable configurations of relation pairs that comprise the covariance structure. If a probability model for Ξ is jointly exchangeable, then all entries \( ξ_{ij} \) are marginally identically distributed under the model. This implies that each of the covariances corresponding to a particular configuration in Figure 1 (plus the variance term)
should have the same value across all possible actor labels. Stating this result formally, we have the following.

**Proposition 1.** *If a probability model for a directed relational matrix $\Xi$ is jointly exchangeable and has finite second moments, then the covariance matrix of $\Xi$ contains at most six unique values.*

**Proof:** Consider a probability model for a directed relational matrix $\Xi$ that satisfies the joint exchangeability and second moment criteria defined above. For any four, possibly non-unique, actors $\{i, j, k, l\}$, observe that the covariance between the errors $\xi_{ij}$ and $\xi_{kl}$ takes one of the following six values, depending on the relationships between the actor indices:

- $\text{Var}(\xi_{ij})$ if $i = k$ and $j = l$;
- $\text{Cov}(\xi_{ij}, \xi_{kj})$ if $i \neq k$ and $j = l$;
- $\text{Cov}(\xi_{ij}, \xi_{ji})$ if $i = l$ and $j = k$;
- $\text{Cov}(\xi_{ij}, \xi_{kl})$ if $i \neq k$ and $j \neq l$;
- $\text{Cov}(\xi_{ij}, \xi_{il})$ if $i = k$ and $j \neq l$;
- $\text{Cov}(\xi_{ij}, \xi_{kl})$ if $i \neq k$ and $j \neq l$.

Now consider an arbitrary permutation operation $\pi(\cdot)$ of the entire actor set $\{1, ..., n\}$. Note that exchangeability implies the bivariate distribution of the pair $(\xi_{ij}, \xi_{kl})$ must be the same as distribution of $(\xi_{\pi(i)\pi(j)}, \xi_{\pi(k)\pi(l)})$. Thus, the covariance of $\xi_{\pi(i)\pi(j)}$ and $\xi_{\pi(k)\pi(l)}$ must equal that of the original pair:

$$\text{Cov}(\xi_{ij}, \xi_{kl}) = \text{Cov}(\xi_{\pi(i)\pi(j)}, \xi_{\pi(k)\pi(l)}) \quad \text{for any } i, j, k, l.$$

By exchangeability this is true for all permutations $\pi(\cdot)$, establishing the result. $\square$

To illustrate the correspondence between joint exchangeability and the covariance entries, consider the bilinear mixed effects network regression model proposed in [Hoff (2005)]. This model uses an inner product measure to model the error structure in relations and can be
expressed as follows:

\begin{equation}
\begin{aligned}
y_{ij} &= \beta^T x_{ij} + \xi_{ij}; \\
\xi_{ij} &= a_i + b_j + z_i^T z_j + \gamma_{(ij)} + \epsilon_{ij}; \\
(a_i, b_i) &\sim N_2(0, \Sigma_{ab}); \\
\Sigma_{ab} = \begin{pmatrix}
\sigma_a^2 & \rho_{ab}\sigma_a \sigma_b \\
\rho_{ab}\sigma_a \sigma_b & \sigma_b^2
\end{pmatrix}; \\
(z_i, z_j) &\sim N_d(0, \sigma_z^2 I_d); \\
\gamma_{(ij)} &= \gamma_{(ji)} \sim N(0, \sigma_\gamma^2); \\
\epsilon_{ij} &\sim N(0, \sigma_\epsilon^2).
\end{aligned}
\end{equation}

where \(a_i, b_j, z_i, z_j,\) and \(\epsilon_{ij}\) are independent. Note that \(E[\xi_{ij}] = 0.\)

As presented in Hoff (2005), the elements of \(V[\Xi] = V[Y_\nu|X]\) are

- \(\text{Var}(\xi_{ij}) = \sigma_a^2 + \sigma_b^2 + d\sigma_z^4 + \sigma_\gamma^2 + \sigma_\epsilon^2,\)
- \(\text{Cov}(\xi_{ij}, \xi_{kj}) = \sigma_b^2,\)
- \(\text{Cov}(\xi_{ij}, \xi_{ji}) = 2\rho_{ab}\sigma_a \sigma_b + d\sigma_z^4 + \sigma_\gamma^2,\)
- \(\text{Cov}(\xi_{ij}, \xi_{ki}) = \text{Cov}(\xi_{ij}, \xi_{jk}) = \rho_{ab}\sigma_a \sigma_b.\)
- \(\text{Cov}(\xi_{ij}, \xi_{il}) = \sigma_a^2,\)
- \(\text{Cov}(\xi_{ij}, \xi_{kl}) = 0,\)
- \(\text{Cov}(\xi_{ij}, \xi_{kl}) = 0,\)
- \(\text{Cov}(\xi_{ij}, \xi_{kl}) = 0,\)

Note that there are six unique terms, corresponding to the five relation pair configurations shown in Figure [1] and a variance term. Moreover, these terms depend only on the population-level parameters of the data generating process and not on individual-level latent variables.

Like the results in Hoff (2005), our work draws on a much deeper, general literature on variance decompositions for structured and symmetric models. In regard to symmetry, a related notion to exchangeability, Dawid (1988) states that “the specification of the relevant symmetry represents a pre-modelling phase from which many important consequences flow.” Our work leverages these symmetries, assuming only, again quoting Dawid (1988), that there is “no reason to consider the observations in any one order rather than any other.” Work by Li and Loken (2002), Li et al. (2002), and Li (2006) generalize the social relations model (SRM) of Warner et al. (1979) to describe the family of symmetric probability distributions for dyadic data. Though these approaches confirm our findings on the gains of assuming exchangeability, their approach to modeling the covariance structure is quite different. These
approaches draw inspiration from the variance decomposition literature in statistics. This motivation leads to developing hypothesis tests that explore restrictions on the symmetries (i.e. invariance to transformations) as a null hypothesis, but impose a parametric form on the error terms (e.g. involving sender, receiver, and pairwise effects in the Warner et al. (1979) social relations model) and in some cases assume a Gaussian likelihood. In contrast, our motivation comes from econometric methods for nonparametric standard error estimation. As a result, we leverage the exchangeability assumption only to simplify the existing estimating equation uncertainty estimates, rather than attempt to fully specify a probability distribution for the data.

3.3 Covariance matrices of exchangeable relational arrays

Proposition 1 implies that at most six parameters are required to describe the dependence structure arising from jointly exchangeable relational models. Thus, we introduce a new class of covariance matrices $\Omega_E$ which contain five unique nonzero entries: one variance parameter $\sigma^2$ along the diagonal of $\Omega_E$ and four covariance parameters $\{\phi_a, \phi_b, \phi_c, \phi_d\}$ associated with (a-d) in Figure 1. Similar to the DC covariance model, we assume that the sixth parameter is zero. This assumption is equivalent to assuming non-overlapping directed pairs are independent, such that $\text{Cov}(\xi_{ij}, \xi_{kl}) = 0$, corresponding to (e) in Figure 1. Though there may be association between non-overlapping pairs, we expect this dependence to be small compared to dependence between pairs that share a member. Figure 2 shows the structure of $\Omega_E$ for a relational matrix with four actors $\{A, B, C, D\}$. We formally define the class $\Omega_E$ below.

**Definition 1.** An exchangeable covariance matrix is defined as $\Omega_E = \mathbb{E}[\Xi_v \Xi_v^T]$ arising from mean-zero random vector $\Xi_v = \text{vec}(\Xi)$, where $\Xi$ is a jointly exchangeable random matrix with $\xi_{ij}$ independent $\xi_{kl}$ whenever $\{i, j\} \cap \{k, l\} \neq \emptyset$. $\Omega_E$ has five unique terms, a variance and four covariances: $\{\sigma^2, \phi_a, \phi_b, \phi_c, \phi_d\}$.

We now show that for a linear model with error covariance matrix in the class of $\Omega_E$, the OLS estimate of the coefficients $\beta$ is asymptotically normal. Our asymptotic regime is the
Figure 2: Consider a matrix $Y$ containing the relations among four actors $\{A, B, C, D\}$ shown on the left. Since the relation between an actor and itself is undefined, the diagonal entries (blacked out in the picture) are not regarded as part of $Y$. Assuming joint exchangeability of the actors and that relations involving non-overlapping sets of actors are independent, the covariance matrix $\Omega_E$ contains five unique values.

The addition of actors to the relational data set, leading to asymptotics in $n$. For this result, we treat $X$ as a random variable.

**Theorem 1.** Consider the following assumptions regarding the data generating process:

(A1) The true data generating model is $Y_v = X\beta + \Xi_v$, where the errors $\Xi_v$ are mean-zero with exchangeable covariance matrix as defined in Definition 1.

(A2) At least one of $\{\phi_b, \phi_c, \phi_d\}$ is nonzero.

In addition, we impose the following regularity conditions:

(B1) The covariate vectors $x_{jk}$ form a random sample (i.e. are independent and identically distributed).

(B2) The fourth moments of the covariates and the errors are bounded: $E[(x_{jk}x_{jk}^T)^2] < C < \infty$, where the square is taken element-wise on $x_{jk}x_{jk}^T$, and $E[\xi_{jk}^4] < C' < \infty$. 

\[ Y = \begin{array}{cccc}
A & B & C & D \\
A & y_{AB} & y_{AC} & y_{AD} \\
B & y_{BA} & y_{BC} & y_{BD} \\
C & y_{CA} & y_{CB} & y_{CD} \\
D & y_{DA} & y_{DB} & y_{DC} \\
\end{array} \]

\[ \Omega_E = \begin{array}{cccccccccccc}
& y_{BA} & y_{BC} & y_{BD} & y_{DA} & y_{DB} & y_{DC} & y_{EA} & y_{EC} & y_{ED} & y_{FA} & y_{FB} & y_{FD} \\
y_{BA} & \sigma^2 & \phi_b & \phi_c & \phi_d & \phi_b & \phi_c & \phi_d & \phi_b & \phi_c & \phi_d & \phi_b & \phi_c & \phi_d \\
y_{BC} & \phi_b & \sigma^2 & \phi_b & \phi_b & \phi_c & \phi_d & \phi_b & \phi_c & \phi_d & \phi_b & \phi_c & \phi_d & \phi_b \\
y_{BD} & \phi_c & \phi_b & \sigma^2 & \phi_b & \phi_c & \phi_d & \phi_b & \phi c & \phi d & \phi_b & \phi_c & \phi d & \phi c \\
y_{DA} & \phi_d & \phi_b & \phi_d & \phi_d & \phi_d & \phi_d & \phi_d & \phi_d & \phi d & \phi d & \phi d & \phi d & \phi d \\
y_{DB} & \phi_b & \phi c & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d \\
y_{DC} & \phi c & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d \\
y_{EA} & \phi_b & \phi c & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d \\
y_{EC} & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c \\
y_{ED} & \phi c & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d \\
y_{FA} & \phi b & \phi c & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d \\
y_{FB} & \phi c & \phi b & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c & \phi c \\
y_{FD} & \phi d & \phi c & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d \\
y_{FD} & \phi c & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d & \phi d \\
\end{array} \]
(B3) The errors $\Xi$ and covariates $X$ are independent.

(B4) $X$ is full rank.

Given (A1) – (A2) and (B1) – (B4), the ordinary least squares estimate $\hat{\beta}$ is asymptotically normal:

$$\sqrt{n}(\hat{\beta} - \beta) \to_d N(0, (\phi_b + \phi_c + 2\phi_d) E_{XX}^{-1}),$$

where we define $E_{XX} := \mathbb{E}[x_{jk}x_{jk}^T]$ and $\to_d$ denotes element-wise convergence in distribution.

The proof of Theorem 1 is given in the online supplement. Note that only the covariances $\{\phi_b, \phi_c, \phi_d\}$ appear in asymptotic variance of $\hat{\beta}$. This results from the fact that there are an order of magnitude more of terms $\{\phi_b, \phi_c, \phi_d\}$ in the covariance matrix $\Omega_E$ than there are of the terms $\{\phi_a, \sigma^2\}$. In particular, in $\Omega_E$ there are $n(n-1)(n-2)$ dyadic pairs $(\xi_{ij}, \xi_{kl})$ of each of type (b) and type (c), $2n(n-1)(n-2)$ pairs of type (d), and $n(n-1)$ pairs of each of type (a) and $\sigma^2$. We make the assumption that at least one of the covariances $\{\phi_b, \phi_c, \phi_d\}$ is nonzero. Should the assumption be violated, i.e. $\phi_b = \phi_c = \phi_d = 0$, then all $\binom{n}{2}$ dyadic pairs of the form $(\xi_{ij}, \xi_{ji})$ are independent of one another, and the asymptotic normality of $\hat{\beta}$ follows from the usual independent data arguments. The canonical case of independent and identically distributed errors is recovered when $\phi_a = \phi_b = \phi_c = \phi_d = 0$.

### 3.4 Exchangeable covariance estimator

As emphasized above, the DC estimators in (8) and (9) estimate each nonzero element in $\Omega_{DC}$ using a single product of residuals. Here we introduce novel estimators inspired by the covariance structure $\Omega$ associated with relations are jointly exchangeable. Specifically we consider estimates of $\Omega$ in the class of exchangeable covariance matrices, as in Definition 1. Our new exchangeable (EXCH) covariance estimator $\hat{\Omega}_E$, and corresponding estimator
of $V[\beta]$ can be written, respectively, as

$$\hat{\Omega}_E = \hat{\sigma}^2 I_{n(n-1)} + \sum_{s=a}^{d} \hat{\phi}_s S_s, \quad \text{and} \quad \hat{V}_E = (X^T X)^{-1} X^T \hat{\Omega}_E X (X^T X)^{-1}, \quad (8)$$

where $S_s$ denotes a $(n(n-1) \times n(n-1))$ binary matrix with 1s in the entries corresponding to relation pairs of type $s \in \{a, b, c, d\}$ as defined in Figure 1.

We propose estimating the five parameters in $\Omega_E$ by simply averaging the residual products across pairs having the same index configurations corresponding to (a)-(d) in Figure 1.

These empirical mean estimates can be expressed

- $\hat{\sigma}^2 = \text{Var}(\xi_{ij}) = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} e_{ij}^2$;
- $\hat{\phi}_b = \text{Cov}(\xi_{ij}, \xi_{kj}) = \frac{1}{n(n-1)(n-2)} \sum_i \sum_{j \neq i} e_{ij} \left( \sum_k e_{kj} - e_{ij} \right)$;
- $\hat{\phi}_a = \text{Cov}(\xi_{ij}, \xi_{ji}) = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} e_{ij} e_{ji}$;
- $\hat{\phi}_c = \text{Cov}(\xi_{ij}, \xi_{ik}) = \frac{1}{n(n-1)(n-2)} \sum_i \sum_{j \neq i} e_{ij} \left( \sum_k e_{ik} - e_{ij} \right)$;
- $\hat{\phi}_d = \text{Cov}(\xi_{ij}, \xi_{ki}) = \text{Cov}(\xi_{ij}, \xi_{jl}) = \frac{1}{2n(n-1)(n-2)} \sum_i \sum_{j \neq i} e_{ij} \left( \sum_{k \neq i} e_{ki} + \sum_{k \neq j} e_{jk} - 2e_{ji} \right)$.

Even if the underlying data generating model is not jointly exchangeable, the proposed estimator will work well if the variability in the covariances among relations of the same type (i.e. (a)-(d) in Figure 1) is small. In this case, the reduction in estimation variance that arises from pooling will outweigh the small bias introduced in the estimation of each covariance entry.

### 4 Evaluating the exchangeable estimator

In this section, we empirically and theoretically evaluate the properties of our estimators in (8) for data with a single matrix of relations (i.e. $R = 1$). We first present simulation evidence of improved inference for $\hat{\beta}$ by simulating data from both exchangeable and non-exchangeable generative models. We then examine the theoretical properties in the spirit of, for example, Kauermann and Carroll (2001) by comparing the bias of $\hat{V}_E$ and $\hat{V}_{DC}$. We
also prove our variance estimator $\hat{\text{V}}_E$ is consistent for the true variance of $\hat{\beta}$ when the data generating process is jointly exchangeable.

4.1 Simulation evidence

We performed simulations using three different data generating models for the errors $\Xi = \{\xi_{ij}\}$: (i) independent and identically distributed errors, (ii) errors generated from the (exchangeable) bilinear mixed effects model of [Hoff (2005)] shown in (7) and (iii) errors generated from a non-exchangeable model. The non-exchangeable model included systematic noise in the upper-left quadrant of the relational error matrix $\Xi$. Since noise was added to actor relations in the same position in $\Xi$ in each simulation run, the distribution of the relations was not exchangeable: the distribution of the errors would be different for a reordering of the rows and columns.

For each simulation setting, we employed the following three-covariate regression model:

$$y_{ij} = \beta_1 + \beta_2 1_{x_2i \in C} 1_{x_2j \in C} + \beta_3 |x_{3i} - x_{3j}| + \beta_4 x_{4ij} + \xi_{ij}.$$  \hfill (9)

In this model, $\beta_1$ is an intercept; $\beta_2$ is a coefficient on a binary indicator of whether individuals $i$ and $j$ both belong to a pre-specified class $C$; $\beta_3$ is a coefficient on the absolute difference of a continuous, actor-specific covariate $x_{3i}$; and $\beta_4$ is that for a pair-specific continuous covariate $x_{4ij}$. For the entirety of the study, we fixed $\beta$ at a single set of values. Since the variance of $\hat{\beta}$ explicitly depends on $X$, we generated 500 random design matrices $X$ for each sample size of actors, and for each design matrix simulated 1,000 error matrices under each of the three models to assess the variability of the standard error estimates and accuracy of the subsequent confidence intervals for $\beta$. For additional details on the simulation study procedure, see the online supplement.

Figures 3-5 display the coverage probabilities for 95% confidence intervals for each $\beta$ for the three error settings. Along with the dyadic clustering (DC) and exchangeable (EXCH)
estimators, we also include the standard heteroskedasticity consistent (HC) estimator as a baseline, as in Aronow et al. (2015).

Figure 3: (IID Errors) Probability $\beta$ is in 95% confidence interval across 500 random $X$ draws when the errors are independent and identically distributed. Lines in the boxplots denote the median coverage, the box denotes the middle 80% of coverages, and the whiskers denote the middle 95% of coverages across the set of design matrices.

Figure 4: (Exchangeable Errors) Probability $\beta$ is in 95% confidence interval across 500 random $X$ draws when the errors are generated according to the exchangeable bilinear effects model. Lines in the boxplots denote the median coverage, the box denotes the middle 80% of coverages, and the whiskers denote the middle 95% of coverages across the set of design matrices.

We draw two key conclusions from our simulations. First, our proposed approach performs extremely well compared to the DC and HC alternatives, even when exchangeability is violated. Specifically, we see that the EXCH estimator produces confidence intervals with...
nominal, or near nominal, coverage for a variety of data generating processes. In addition, we see the variability in coverage across different $X$ realizations for the EXCH estimator is substantially smaller than that for the other estimators. We hypothesize the observed reduction in variability is a result of the averaging inherent in the EXCH estimator. In particular, the EXCH estimator replaces DC’s $O(n^3)$ unique residual products with five averages over subsets of these products, where each subset consists of residual products that are of the same covariance type. Intuitively, this averaging should result in a reduction of the variance of the EXCH estimator relative to that of the DC estimator. In the online supplement, we plot the standard deviation of the EXCH and DC standard error estimates; in these plots, we clearly see the reduction in variability of the EXCH estimator relative to the DC estimator. We also plot the expected error (given $X$) of the DC and EXCH standard error estimates relative to the true standard errors. We find that both estimators generally underestimate the true standard errors (and thus confidence interval width), although the EXCH estimator underestimates to a significantly lesser degree. Returning to the coverage plots, it is interesting that even when the exchangeable assumption is incorrect, as in Figure 5, we see better performance from the EXCH estimator than the others. This suggests the
reduction in the variability of the covariance entry estimates in the exchangeable estimator can outweigh the covariance model misspecification.

The second key observation we glean from the study is that the type of covariate (e.g. continuous actor-level characteristic versus product of binary indicators) affects the performance of all standard error estimators. For example, Figures 4 and 5 show that when there is structure in the errors, the variability in the confidence interval coverage across design matrices is far greater for the binary covariate than for either of the continuous covariates. Focusing specifically on the boxes representing the middle 80% of coverage levels across the 500 simulations associated with the binary coefficient (left-most plots in Figures 3 through 5), we see the EXCH estimator coverage varies from about 93-98%, whereas the DC estimator varies between 50-95% with no improvement as the sample size $n$ increases.

4.2 Consistency of $\widehat{V}_E$

We now show that if the data generating model is jointly exchangeable, the exchangeable covariance estimator is consistent for the true variance of the coefficients. Again our asymptotic regime is the addition of actors to the relational data set, i.e. increasing $n$.

**Theorem 2.** Under the conditions of Theorem 1, the exchangeable covariance estimator is consistent in the sense that

$$n\widehat{V}_E - nV[\beta] \to_p 0 \quad \text{as} \quad n \to \infty,$$

where ‘$\to_p$’ denotes element-wise convergence in probability.

The proof of Theorem 2 is given in the online supplement.

4.3 Bias of $\widehat{V}_{DC}$ and $\widehat{V}_E$

If the underlying model is jointly exchangeable, we expect the exchangeable estimate of $\Omega$ to be closer to the population $\Omega$ than the dyadic clustering estimate. Furthermore, since the
EXCH and DC sandwich variance estimators of $V[\hat{\beta}]$ are linear combinations of the elements in their respective error covariance matrix estimates, we expect a reduction in bias of the EXCH estimator $\hat{V}_E$ relative to the DC estimator $\hat{V}_{DC}$. In general, we observed reduced bias in the coverage rates in Figures 3 - 5. To complement the simulation evidence, we provide the following theorem establishing $\hat{V}_E$ has lower bias than $\hat{V}_{DC}$ in the case of a simple linear regression.

**Theorem 3.** Consider the case of relational simple linear regression, where $y_{ij} = z_{ij}\beta + \xi_{ij}$. Assume the errors are mean-zero with an exchangeable covariance matrix as defined in Definition 1 and that the covariate vector $z_{ij}$ is centered (i.e. $\sum_{i \neq j} z_{ij} = 0$). Then, the bias of the exchangeable covariance estimator is always less than or equal to the bias of the dyadic clustering sandwich variance estimator:

$$\left| \frac{\text{Bias}(\hat{V}_{DC})}{\text{Bias}(\hat{V}_E)} \right| \geq 1.$$ 

The proof of Theorem 3 is given in the online supplement.

5 Regressions involving relational arrays

In this section we extend our discussion of exchangeable estimators to the case when $R > 1$. We introduce two notions of exchangeability for relational array data and discuss models consistent with these assumptions. We separately consider the cases when the underlying model for the error array is exchangeable along the third dimension and that when it is not. Figure 6 illustrates the former case and two variations of the latter. Before dissecting the spectrum of possible exchangeability assumptions, we first revisit the treatment of error arrays with $R > 1$ by Aronow et al. (2015).
5.1 Dyadic clustering

Aronow et al. (2015) examine relational regression standard errors when the third dimension, indexed by \( r \) in \( Y = \{y_{ij}^{(r)}\} \), denotes time. Data consistent with this structure is, for example, country trade over time. Aronow et al. (2015)'s treatment is a direct extension of dyadic clustering in two dimensions: two errors \( \xi_{ij}^{(r)} \) and \( \xi_{k\ell}^{(s)} \) are assumed to be independent if the associated dyads do not share a member (i.e. \( \{i, j\} \cap \{k, \ell\} = \emptyset \)), regardless of the third dimension indices \( r \) and \( s \). As in the \( R = 1 \) case, each nonzero covariance entry is estimated by the corresponding residual product, i.e. \( \text{Cov}(\xi_{ij}^{(r)}, \xi_{k\ell}^{(s)}) = e_{ij}^{(r)} e_{k\ell}^{(s)} \). Note that this specification makes no assumptions about the dependence structure along the third dimension.

5.2 Exchangeability in the third dimension

Here we consider relational data that are fully exchangeable in the third dimension. Intuitively, the numbering of the row, column, and depth indices of an array with this property are uninformative. For example, consider the case where the relational array \( Y \) represents the quantity of trade between pairs of countries, decomposed by various categories of goods traded (e.g. intangible vs. tangible). Without reason to believe some pairs of good types are more dependent than others, we might be willing to assume the dependence structure along the third dimension is exchangeable.

Define a permutation of the third dimension indices \( \nu(.) \) in addition to the row and column permutation \( \pi(.) \) defined previously. An array probability model that is jointly exchangeable, as well as exchangeable in the third dimension, has the property that

\[
\text{Cov}(\xi_{\pi(i)\pi(j)}^{(\nu(r))}, \xi_{\pi(k)\pi(\ell)}^{(\nu(s))}) = \text{Cov}(\xi_{ij}^{(r)}, \xi_{k\ell}^{(s)}).
\]

(11)

It follows that the covariance matrix, denoted \( \Omega_{Ea} = V[\Xi_v] \), consists of 10 distinct nonzero parameters, corresponding to two submatrices \( \Omega_1 \) and \( \Omega_2 \), which each have exchangeable
structure as in Definition 1. Along the diagonal of $\Omega_{Ea}$ there are $R$ instances of the $n(n-1) \times n(n-1)$ matrix $\Omega_1$ represents covariance between observations that share the same third index, i.e. $r = s$. The off-diagonal blocks of $\Omega_{Ea}$ are populated with a second $n(n-1) \times n(n-1)$ exchangeable error matrix $\Omega_2$ for errors that do not share the same third index, i.e. $r \neq s$. This structure is depicted in Figure 6(a). As previously, we propose estimating each of the 10 unique values with the average of the corresponding residual products.

Jointly exchangeable models that model the slices of the error array $\{\Xi^{(1)}, \Xi^{(2)},...,\Xi^{(R)}\}$ as independent constitute a subclass of the models with full exchangeability. Specifically, they make the additional assumption that $\text{Cov}\left(\xi^{(r)}_{ij}, \xi^{(s)}_{k\ell}\right) = 0$ for $r \neq s$. In Figure 6(a), an assumption of independence along the third dimension corresponds to $\Omega_2 = 0$.

Figure 6: Covariance matrices $\Omega = V[\Xi_v]$ for exchangeable arrays with depth $R = 4$ where $\Xi_v^T = \left((\Xi_v^{(1)})^T, (\Xi_v^{(2)})^T, (\Xi_v^{(3)})^T, (\Xi_v^{(4)})^T\right)$. All matrices are symmetric, where the $(i,j)$ block denotes $\text{Cov}(\Xi_v^{(i)}, \Xi_v^{(j)})$. Subfigure (a) corresponds to full exchangeability yielding 2 unique blocks, (b) corresponds to no exchangeability in the third dimension with stationarity assumption yielding $R = 4$ unique blocks, and (c) corresponds to no exchangeability in the third dimension yielding $(\frac{R}{2}) + R = 10$ unique blocks. Each block contains five unique nonzero terms as in $\Omega_E$ in Figure 2.

5.3 Partial exchangeability or no exchangeability in the third dimension

The assumption of exchangeability along the third dimension can be unnatural and inappropriate for certain data sets, so here we consider relaxing the fully exchangeable assumption
introduced Section 5.2. Consider again the quantity of trade between countries \(i\) and \(j\) as the relational response, except where trade decomposed by time period rather than by good type. We would expect the temporal index in the third dimension to be non-exchangeable, as we might expect errors associated with nearby time periods will be more dependent than those far apart.

In this section, we consider arrays which are jointly exchangeable along the rows and columns only, such that the ordering of the array in the third (depth) dimension must remain the same for the probability distribution to remain invariant. Intuitively, this property corresponds to one where the labeling of rows and columns is inconsequential, but the labeling of the third dimension is material. This exchangeability assumption implies

\[
\text{Cov} \left( \xi_{\pi(i)\pi(j)}, \xi_{\pi(k)\pi(l)} \right) = \text{Cov} \left( \xi_{ij}, \xi_{k\ell} \right).
\]

(12)

The full covariance matrix, denoted \(\Omega_{Ec} = V[\Xi_v]\), contains a separate \(n(n - 1) \times n(n - 1)\) exchangeable covariance matrix for each of the \(\binom{R}{2}\) unique third index pairings and each of the \(R\) diagonal variance matrices (see Figure 6(c)). Covariance matrices of this form contain \(5 \left( \binom{R}{2} + R \right)\) unique parameters.

This type of exchangeability assumption is extremely unrestrictive. Specifically, it places no constraints on the evolution of the dependence along the third dimension. However, a more restrictive assumption specifying the relationships among the covariances in the third dimension may be appropriate when we expect the behavior in this dimension vary in a particular manner. For example, if the third dimension corresponds to different time periods, it may be reasonable to assume stationarity along the third dimension. Specifically, this assumption implies the covariance across time periods only depends on the absolute difference in the time indices. In this case, there are five unique nonzero covariances for each difference in time \(|r - s|\), yielding \(5R\) unique nonzero values in the covariance matrix. We denote a covariance matrix with this structure by \(\Omega_{Eb}\) (see Figure 6(b)).
6 Patterns in international trade

In this section we demonstrate our exchangeable standard error estimator using data on international trade flow over multiple decades. We fit the model using Generalized Estimating Equations (GEE), which weights the estimating equations, $g$, in (3) by an estimate of the inverse of the “working” covariance matrix of the observations (see, for example, Chapter 8 in Wakefield (2013) for a review of GEE). When the assumed covariance structure is correct, this approach yields an estimator $\hat{\beta}$ which has improved efficiency over that based on unweighted equations in (3). In the remainder of this section, we outline how the exchangeable estimator can be used in a method of moments (weighted least squares) approach to estimation and present results from the international trade data.

6.1 Inference via GEE

Inference via GEE proceeds by first specifying a “working” covariance matrix for the errors, which serves as a weight for estimating equations. The choice of the working covariance matrix represents a trade-off between robustness and efficiency. If the working matrix resembles the true underlying covariance structure, then the efficiency of $\hat{\beta}$ improves over that resulting from the estimating equations in (3). Even if the working covariance is misspecified, the standard error estimates for $\hat{\beta}$ can be ‘corrected’ using the sandwich standard error estimators with an appropriate estimator $\hat{\Omega}$. However, these standard error estimates can be unstable if the assumed working structure differs greatly from the truth, which, of course, is unknown in practice (see discussion in Chapter 8 of Wakefield (2013), for example).

The GEE algorithm proceeds as follows. Let $W^{-1}$ be the working covariance matrix, then the estimate of $\beta$ is the solution to the GEE estimating equation

$$XW(Y_v - \beta^T X) = 0,$$
and the corresponding variance estimator of the coefficients is

\[ V[\hat{\beta}] = (X^T W X)^{-1} X^T W \Omega W X (X^T W X)^{-1}. \]

Our estimation algorithm is composed of a two-step iteration procedure. Given initial estimates \( \hat{\beta}^{(0)} \) and corresponding residuals \( \hat{\Xi}^{(0)} \), we iterate between two steps, such that for iteration \( \tau + 1 \):

1. Solving \( X \hat{W}^{(\tau)} \left( Y_v - X \hat{\beta}^{(\tau+1)} \right) = 0 \), set \( \hat{\beta}^{(\tau+1)} = \left( X^T \hat{W}^{(\tau)} X \right)^{-1} X^T \hat{W}^{(\tau)} Y_v \).

2. Use \( \hat{\beta}^{(\tau+1)} \) to calculate \( \hat{\Xi}^{(\tau+1)} \), and obtain estimates \( \hat{\Omega}^{(\tau+1)} \) and \( \hat{W}^{(\tau+1)} \).

These steps are repeated until convergence.

### 6.2 International trade models

We demonstrate the implications of using our exchangeable standard error estimator in a study of international trade between 58 countries. These data were previously analyzed and made available by Westveld and Hoff (2011)\(^1\). For each pair of countries, we observe yearly total volume of trade between the two countries for a period from 1981-2000. Following Westveld and Hoff (2011) and Tinbergen (1962), we model (log) trade in a given year using a modified gravity mean model. The gravity model, proposed by Tinbergen (1962), posits that the total trade between countries is proportional to overall economic activity of the countries weighted by the inverse of the distance between them (raised to a power). Following Ward and Hoff (2007), we also add an indicator for whether the nations’ militaries cooperated in the given year and a measure of democracy, i.e. polity, which ranges from 0 (highly authoritarian) to 20 (highly democratic).

\(^1\)See [https://doi.org/10.1214/10-AOS403SUPP](https://doi.org/10.1214/10-AOS403SUPP) for data.
The complete model has the form:

\[ \ln \text{Trade}_{ijt} = \beta_0 t + \beta_1 t \ln \text{GDP}_{it} + \beta_2 t \ln \text{GDP}_{jt} + \beta_3 t \ln D_{ijt} \]

\[ + \beta_4 t \text{Pol}_{it} + \beta_5 t \text{Pol}_{jt} + \beta_6 t \text{CC}_{ijt} + \beta_7 t (\text{Pol}_{it} \times \text{Pol}_{jt}) + \epsilon_{ijt}, \]

where \( \ln \text{Trade}_{ijt} \) is the (log) volume of trade between countries \( i \) and \( j \) at time \( t \); \( \ln \text{GDP}_{it} \) and \( \ln \text{GDP}_{jt} \) are the (log) Gross Domestic Product of nations \( i \) and \( j \), respectively; \( \ln D_{ijt} \) is the (log) geographic distance between nations; \( \text{CC}_{ijt} \) is the measure of cooperation in conflict (coded as \( +1 \) if nations were on the same side of a dispute and \( -1 \) if they were on opposing sides); and \( \text{Pol}_{it} \) and \( \text{Pol}_{jt} \) are the polity measures for \( i \) and \( j \), respectively.

We fit the regression model above using the GEE approach, where both the working covariance matrix \( W^{-1} \) and the population covariance matrix \( \Omega \) to have the covariance structure \( \Omega_{Ea} \), as in panel (a) of Figure 6. The estimator of \( \beta \) is then based the assumption that the error covariance structure is fully exchangeable. We place no further restrictions on the covariance structure beyond this exchangeability.

The exchangeability assumption underlying our approach differs substantially from assumptions frequently made in analyses of temporal relational data. For example, Westveld and Hoff (2011) explicitly decompose the error term for each pair and time, \( \epsilon_{ijt} \), into time-dependent sender and receiver effects which represent relational structure and a temporally dependent error. Specifically, Westveld and Hoff (2011) assume the dependence between time periods is autoregressive order one. The structure in Figure 6(a), in contrast, imposes an exchangeability restriction, but does not imply a specific decomposition between relational and temporal effects. The temporal structure in Figure 6(a) implies that the covariance is the same across overlapping dyads in different time points. We chose this model as a baseline error exchangeability structure that contains effects for relational structure within each time period and a general covariance across time periods. We expect the typical trade-off between modeling assumptions and efficiency to apply: Westveld and Hoff (2011) estimates
will be more efficient, but only if the assumed model structure matches the population. We also compare our results to the DC estimator of Aronow et al. (2015) described in Section 5.1. Recall the DC estimator makes even fewer assumptions than our method, but it cannot be used for GEE because the covariance matrix estimator $\hat{\Omega}_{DC}$ is always singular. Thus, in the following comparison, we use ordinary least squares to estimate $\beta$ as in (3) and estimate confidence intervals using $\hat{V}_{DC}$. In the online supplement, we provide a proof that $\hat{\Omega}_{DC}$ is always singular and a method for efficiently inverting $\hat{\Omega}_E$.

### 6.3 International trade results

The estimated coefficients and corresponding 95% confidence intervals and posterior credible intervals are in Figure 7. Coefficient estimates for Westveld and Hoff (2011) are posterior medians and 95% credible intervals based on a Bayesian estimation procedure. Interpreting Figure 7 requires care as (i) there is no ground truth and (ii) we are comparing three different inference paradigms. Nonetheless, focusing on two aspects of Figure 7 reveals important insights for practitioners determining which paradigm to use. First, consider the overall trends in estimated $\hat{\beta}$ across time for the three methods. Fitting using GEE produces coefficients much closer to Westveld and Hoff (2011) than OLS. In particular, the intercept estimated via OLS is approximately three times larger in magnitude than either Westveld and Hoff (2011) or our GEE estimates. We see a similar result with the other coefficient that is relatively constant over time, log GDP of exporter. The GEE estimates informed by our covariance estimator also produce trends that generally match those in Westveld and Hoff (2011). OLS also roughly matches the temporal trend in Westveld and Hoff (2011), with the exception of the cooperation in conflict variable. For this case, the Westveld and Hoff (2011) and GEE estimators are both nearly zero from 1990 onward. The OLS estimator, however, demonstrates substantial fluctuations that are not present in the other methods.

The second aspect to take note of is confidence interval width. The widths of the confidence intervals for the exchangeable GEE approach are generally comparable to Westveld and Hoff (2011).
Figure 7: Estimated coefficients for the all time periods using three different estimation techniques.
and Hoff (2011), while the DC interval widths are noticeably larger. The exchangeable GEE approach incorporates information about the covariance structure of the errors when estimating the regression coefficients. The OLS estimate of $\beta$, however, is identical to a GEE estimate when the working covariance $W^{-1} = I$. If the working covariance estimate is close to the (unknown) true covariance we expect efficiency gains in the GEE estimate of $\beta$ over the OLS estimate. The widths of the Westveld and Hoff (2011) and GEE intervals also tend to be more consistent across time periods than those from OLS/DC. For the cooperation in conflict variable, for example, the OLS/DC confidence intervals become markedly wider during the upward spikes, one of which, in the late 1980s, is only present in OLS/DC estimate.

7 Discussion and conclusion

This paper develops a new set of uncertainty estimators for regression models on relational arrays. The proposed estimators strike a balance between making additional assumptions to decrease variability and remaining robust to dependence heterogeneity and model misspecification. We show that the proposed estimators achieve better coverage than currently available methods in simulation studies, even when the underlying generative model violates the assumptions, and can be used to weight coefficient estimates in GEE.

Our estimator is not appropriate when the dependence among relations, i.e. the covariance structure, is extremely heterogeneous. This can happen in two ways: (i) heterogeneity in the covariance structure is endogenous with an observed covariate that is not included in the regression (a variant of omitted variable bias) and (ii) there is heterogeneity in the error variances even after accounting for all observables in the “true” generating model. In both cases, we could consider an extension of our approach that would further compromise between the unstructured covariance structure of the DC estimator and our exchangeable covariance structure. One could, for example, use a two-stage approach that first estimates
actor clusters using the residuals, then assumes exchangeability within but not across clusters.

Many relational data sets contain binary or count measures, such as the presence or absence of relations between actors or number of interactions. Estimating equation and GEE procedures are often used with non-continuous data whereby the $g$ equations in (2) involve a link function connecting the observed relation to the covariates, mirroring generalized linear regression procedures. While it is possible to impose an exchangeability assumption on the covariance matrix of the observations with non-continuous data, it is unclear how the assumption translates to an assumption about the data generating process. For example, consider the logit and probit regression models for binary data. Both models possess latent variable constructions which involve thresholding a latent continuous outcome composed of the linear regression function plus a random error. Exchangeability of these errors does not imply the relations themselves are exchangeable (conditional on the covariates) as the covariates impact the dependence structure among the relational measures. For this reason, the methods proposed here cannot be trivially applied to non-continuous relational data. However, this is a current area of research for the authors.

References

Aker, J. C. (2010), ‘Information from markets near and far: Mobile phones and agricultural markets in Niger’, *American Economic Journal: Applied Economics* 2(3), 46–59.

Aldous, D. J. (1981), ‘Representations for partially exchangeable arrays of random variables’, *Journal of Multivariate Analysis* 11(4), 581–598.

Aronow, P. M., Samii, C. and Assenova, V. A. (2015), ‘Cluster-robust variance estimation for dyadic data’, *Political Analysis* 23(4), 564–577.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G. and Meghir, C. (2012), ‘Risk pooling,
risk preferences, and social networks’, *American Economic Journal: Applied Economics* 4(2), 134–167.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2013), ‘The diffusion of microfinance’, *Science* 341, 363–373.

Bardham, P. (1984), *Land, Labor and Rural Poverty*, Oxford University Press.

Bickel, P. J. and Chen, A. (2009), ‘A nonparametric view of network models and Newman–Girvan and other modularities’, *Proceedings of the National Academy of Sciences* 106(50), 21068–21073.

Blumenstock, J. E., Fafchamps, M. and Eagle, N. (2011), ‘Risk and reciprocity over the mobile phone network: Evidence from Rwanda’, *Available at SSRN 1958042*.

Bolthausen, E. (1982), ‘On the central limit theorem for stationary mixing random fields’, *The Annals of Probability* 10(4), 1047–1050.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2011), ‘Robust inference with multiway clustering’, *Journal of Business & Economic Statistics* 29(2).

Conley, T. G. (1999), ‘GMM estimation with cross sectional dependence’, *Journal of Econometrics* 92(1), 1–45.

Dawid, A. P. (1988), ‘Symmetry models and hypotheses for structured data layouts’, *Journal of the Royal Statistical Society. Series B* 50(1), 1–34.

Fafchamps, M. (2006), ‘Development and social capital’, *The Journal of Development Studies* 42(7), 1180–1198.

Fafchamps, M. and Gubert, F. (2007), ‘The formation of risk sharing networks’, *Journal of Development Economics* 83(2), 326–350.
Fosdick, B. K. and Hoff, P. D. (2014), ‘Separable factor analysis with applications to mortality data’, *The Annals of Applied Statistics* **8**(1), 120–147.

Foster, A. D. and Rosenzweig, M. R. (2001), ‘Imperfect commitment, altruism, and the family: Evidence from transfer behavior in low-income rural areas’, *Review of Economics and Statistics* **83**(3), 389–407.

Guyon, X. and Ludena, C. (1995), *Random Fields on a Network: Modeling, Statistics, and Applications*, Probability and Its Applications, Springer New York.

Hansen, B. E. (2015), *Econometrics*, Online at [http://www.ssc.wisc.edu/~bhansen/econometrics/](http://www.ssc.wisc.edu/~bhansen/econometrics/)

Hoff, P. (2008), Modeling homophily and stochastic equivalence in symmetric relational data, in ‘Advances in Neural Information Processing Systems’, pp. 657–664.

Hoff, P. D. (2005), ‘Bilinear mixed-effects models for dyadic data’, *Journal of the American Statistical Association* **100**(469), 286–295.

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**(460), 1090–1098.

Hoff, P. D. et al. (2011), ‘Separable covariance arrays via the Tucker product, with applications to multivariate relational data’, *Bayesian Analysis* **6**(2), 179–196.

Hoff, P. D. et al. (2015), ‘Multilinear tensor regression for longitudinal relational data’, *The Annals of Applied Statistics* **9**(3), 1169–1193.

Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983), ‘Stochastic blockmodels: First steps’, *Social Networks* **5**(2), 109–137.

Hoover, D. N. (1979), ‘Relations on probability spaces and arrays of random variables’, *Preprint, Institute for Advanced Study, Princeton, NJ* 2.
Huber, P. J. (1967), The behavior of maximum likelihood estimates under nonstandard conditions, in ‘Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability’, Vol. 1, pp. 221–233.

Jack, W. and Suri, T. (2014), ‘Risk sharing and transactions costs: Evidence from Kenya’s mobile money revolution’, *The American Economic Review* **104**(1), 183–223.

Kauermann, G. and Carroll, R. J. (2001), ‘A note on the efficiency of sandwich covariance matrix estimation’, *Journal of the American Statistical Association* **96**(456), 1387–1396.

King, G. and Roberts, M. E. (2014), ‘How robust standard errors expose methodological problems they do not fix, and what to do about it’, *Political Analysis* p. mpu015.

Li, H. (2006), ‘The covariance structure and likelihood function for multivariate dyadic data’, *Journal of Multivariate Analysis* **97**(6), 1263–1271.

Li, H. and Loken, E. (2002), ‘A unified theory of statistical analysis and inference for variance component models for dyadic data’, *Statistica Sinica* pp. 519–535.

Li, H. et al. (2002), ‘Modeling through group invariance: An interesting example with potential applications’, *The Annals of Statistics* **30**(4), 1069–1080.

Liang, K.-Y. and Zeger, S. L. (1986), ‘Longitudinal data analysis using generalized linear models’, *Biometrika* pp. 13–22.

Lumley, T. and Hamblett, N. M. (2003), ‘Asymptotics for marginal generalized linear models with sparse correlations’, *UW Biostatistics Working Paper Series* .

Tinbergen, J. (1962), *Shaping the world economy; suggestions for an international economic policy*, Twentieth Century Fund, New York.

Van der Vaart, A. W. (2000), *Asymptotic Statistics*, Vol. 3, Cambridge University Press.
Wakefield, J. (2013), *Bayesian and Frequentist Regression Methods*, Springer Science & Business Media.

Wang, Y. J. and Wong, G. Y. (1987), ‘Stochastic blockmodels for directed graphs’, *Journal of the American Statistical Association* 82(397), 8–19.

Ward, M. D. and Hoff, P. D. (2007), ‘Persistent patterns of international commerce’, *Journal of Peace Research* 44(2), 157–175.

Warner, R. M., 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(10), 1742.

Westveld, A. H. and Hoff, P. D. (2011), ‘A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict’, *The Annals of Applied Statistics* pp. 843–872.

White, H. (1980), ‘A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity’, *Econometrica: Journal of the Econometric Society* pp. 817–838.

White, H. and Domowitz, I. (1984), ‘Nonlinear regression with dependent observations’, *Econometrica: Journal of the Econometric Society* pp. 143–161.
A Simulation study details

As noted in Section 4, 500 random realizations of covariates were generated for each sample size of actors \( n \in \{20, 40, 80, 160, 320\} \). For each covariate realization, 1,000 random error realizations were generated for each of the three error settings: IID, exchangeable, and non-exchangeable. Using (9), a simulated data set was created from each covariate realization and error realization pair. The regression model was fit using ordinary least squares to each data set, and standard errors were estimated using the exchangeable, dyadic clustering, and heteroskedasticity-consistent sandwich variance estimators. Confidence interval coverage was estimated for each covariate realization by counting the fraction of confidence intervals that contain the true coefficient.

For all simulations, we fixed true coefficients \( \beta = [1, 1, 1, 1]^T \). We drew each \( x_{2i} \) from a Bernoulli(1/2) distribution independently. In the rare event that \( x_{2i} = x_{2j} \) for all \((i, j)\) pairs, one realization \( x_{2k} \) was randomly flipped to a 1 or 0. All \( x_{3i} \) and \( x_{4ij} \) were drawn independently from a standard normal distribution.

Each error setting was specified to have the same total variance: \( \sum_{ij} \text{Var}(\xi_{ij}) = 3n(n-1) \). This variance was chosen so that the variance of the error would be similar to that of the regression mean model \( \beta^T x_{ij} \). In the IID errors setting, \( \xi_{ij} \sim iid \ N(0, 3) \) for all \((i, j)\). To generate the non-exchangeable errors, a mean-zero random effect was added to the upper left quadrant of \( V[\Xi_v] \). The errors for the non-exchangeable error setting may be written

\[
\xi_{ij} = \tau 1_{i \leq [n/2]} 1_{j \leq [n/2]} + \epsilon_{ij}, \quad \tau \sim N \left( 0, \frac{9n}{4[n/2]} \right), \quad \epsilon_{ij} \sim iid N(0, 3/4).
\]

Finally, the distribution of the exchangeable (bilinear mixed effects model) error setting is defined in (7). We selected the dimension of the latent space to be \( d = 2 \), the correlation between sender and receiver effects as \( \rho_{ab} = 1/2 \), and the sender variance to be twice that
of receiver variance: $\sigma_a^2 = 2\sigma_b^2$. We further specified $\sigma_z = \sigma_\gamma = \sigma_b$. Finally, we selected $\sigma_\epsilon^2 = \frac{3}{4}$. With the aforementioned choices, the restriction $\sum_{ij} \text{Var}(\xi_{ij}) = 3n(n - 1)$ generated a quadratic equation in $\sigma_b$. The standard deviations that resulted from solving this quadratic equation are shown in Table 1.

Table 1: Approximate standard deviations for exchangeable error setting

| $\sigma_\epsilon$ | $\sigma_a$ | $\sigma_b$ | $\sigma_\gamma$ | $\sigma_z$ |
|-----------------|------------|------------|-----------------|------------|
| 0.866           | 0.957      | 0.677      | 0.677           | 0.677      |

Figure 8: (IID Errors) The top row of plots are the average differences in standard errors across random realizations of $X$, where the average is taken over 1,000 error realizations. The bottom row of plots show the standard deviations of the standard error estimates across random $X$. Lines in the boxplots denote the median, the box denotes the middle 80% of values, and the whiskers denote the middle 95% of values.
A.1 Confidence interval widths

To examine the relative confidence interval widths between the exchangeable and dyadic clustering sandwich variance estimators, it is sufficient to examine the values of the standard error estimates. In all simulations we generate 95% confidence intervals by using the typical normal approximation of plus or minus 1.96 times the standard error. We plot the empirical expected standard error given $X$ relative to the true standard error given $X$ in Figures 8-10. We estimate the expectation by averaging the standard error estimates across the 1,000 error realizations, for each $X$ realization. We also compute the standard deviation of the standard error estimates given $X$.

We observe that, for IID and exchangeable error structures in Figures 8 and 9, the standard errors resulting from the exchangeable estimator are much closer to the true standard errors than those resulting from the dyadic clustering estimator. This fact suggests that the dyadic clustering estimator fails to account for a portion of the dependency in the error structure. We note that both procedures generally produce underestimates of the true standard errors, however, the dyadic clustering estimator trades some efficiency for robustness. We observe that the standard deviation of the standard error estimates when using the exchangeable estimator are typically lower than those when using dyadic clustering under IID and exchangeable errors. Intuitively, the lower variability of the exchangeable estimator relative to the dyadic clustering estimator the result of the averaging present in the exchangeable estimator. Finally, the trends observed under IID and exchangeable error structures do not fully persist to non-exchangeable errors. However, the trends of larger expectation and smaller standard deviation are present more often than not under non-exchangeable errors (Figure 10).
Figure 9: (Exchangeable Errors) The top row of plots are the average differences in standard errors across random realizations of $X$, where the average is taken over 1,000 error realizations. The bottom row of plots show the standard deviations of the standard error estimates across random $X$. Lines in the boxplots denote the median, the box denotes the middle 80% of values, and the whiskers denote the middle 95% of values. The ordinate axis is truncated where appropriate to show the estimators of interest.

B DC covariance matrix invertibility

Ideally, for a covariance matrix estimate $\hat{\Omega}$ to be of utmost utility, it must be invertible. For example, if we wish to reweight the estimating equations, as in GEE, and solve iteratively for both the variance matrix and regression coefficients simultaneously, the estimate of the covariance matrix must be nonsingular. However, in many cases the DC estimator is singular and hence cannot be used as a reweighting matrix. In cases when the DC estimator is singular, it can still be used in the ‘meat’ ($B$ matrix) in the coefficient sandwich estimator covariance matrix.

Theorem 4. The dyadic clustering estimate of the error variance, $\hat{\Omega}_{DC}$, is singular for
\[1_{x_2i \in C}1_{x_2j \in C} \quad |x_3i - x_3j| \quad x_{ij}
\]

Figure 10: **(Non-exchangeable Errors)** The top row of plots are the average differences in standard errors across random realizations of \(X\), where the average is taken over 1,000 error realizations. The bottom row of plots show the standard deviations of the standard error estimates across random \(X\). Lines in the boxplots denote the median, the box denotes the middle 80% of values, and the whiskers denote the middle 95% of values. The ordinate axis is truncated where appropriate to show the estimators of interest.

directed data.

**Proof.** The DC estimator can be written as the Hadamard product between the outer product of the residuals and a matrix of indicators of whether the dyad indices share a member.

\[
\tilde{\Omega}_{DC} = ee^T \circ 1_{\{(i,j) \cap (k,l) \neq \emptyset\}}
\]

The rank of the outer product of the residuals is one: \(\text{rank}(ee^T) = 1\). The rank of the indicator matrix is at most \(n(n - 1)/2\), since the indices \((i, j)\) share a member with an arbitrary pair \((k, \ell)\) if and only if the indices \((j, i)\) do as well. Thus, the column of \(1_{\{(i,j) \cap (k,l) \neq \emptyset\}}\)
corresponding to \((i, j)\) is the same as that corresponding to \((j, i)\).

For any two square matrices of equal size \(A\) and \(B\), \(\text{rank}(A \circ B) \leq \text{rank}(A)\text{rank}(B)\). Thus,

\[
\text{rank}(\hat{\Omega}_{DC}) = \text{rank}(ee^T \circ 1_{\{(i,j) \cap (k,l) \neq \emptyset\}}) \\
\leq \text{rank}(ee^T)\text{rank}(1_{\{(i,j) \cap (k,l) \neq \emptyset\}}) \\
\leq \frac{n(n-1)}{2}
\]

\(\hat{\Omega}_{DC}\) is therefore not full rank.

\[\square\]

**Remark 1.** Theorem 4 does not hold for undirected data when \(R = 1\). If the data are undirected, then the bound does not guarantee singularity of \(\hat{\Omega}_{DC}\) since the dimension of \(\hat{\Omega}_{DC}\) is exactly \(n(n-1)/2\). In practice, we find that \(\hat{\Omega}_{DC}\) is full rank in this special case.

**Remark 2.** The result of Theorem 4 holds for both directed and undirected data when \(R > 1\). In this case, the column in the indicator matrix \(1_{\{(i,j) \cap (k,l) \neq \emptyset\}}\) corresponding to the indices \((i, j, s)\) is the same as that column corresponding to \((i, j, t)\) for all values of \(t \in \{1, \ldots, R\}\). Thus, again \(\hat{\Omega}_{DC}\) is not full rank.

## C Undirected arrays

This section specializes the results presented in the manuscript to an undirected network. Consider the case when \(R = 1\) and suppose the network contains the relations among \(n\) actors. The covariance of the errors \(\Omega\) contains three unique elements

\[
\text{Cov}(\xi_{ij}, \xi_{ij}) := \theta, \quad \text{Cov}(\xi_{ij}, \xi_{ki}) := \phi, \quad \text{Cov}(\xi_{ij}, \xi_{kl}) := 0.
\]

As in the directed case, we assume the last covariance, corresponding to two relations which share no common member, is zero. We again estimate the two remaining terms using the
residual matrix $E = \{e_{ij}\} \in \mathbb{R}^{n \times n}$. Note that the residual matrix we consider is for the entire $n \times n$ network and thus contains duplicate off-diagonal entries corresponding to pairs \{(i, j), (j, i)\}. We set the diagonal of $E$ to zero as the relation between an actor and itself is undefined.

The estimate of $\theta$ is the empirical mean of each squared residual and can be expressed

$$\hat{\theta} = \frac{tr(EE)}{n(n - 1)}$$

where $tr(\cdot)$ denotes the matrix trace operator.

Similarly, the estimate of $\phi$ is

$$\hat{\phi} = \frac{1}{2m} \sum_i \sum_{j \neq i} e_{ij} \left( \sum_{k \neq i} e_{ik} + \sum_{k \neq j} e_{kj} - 2e_{ij} \right)$$

$$= \frac{1}{m} 1^T (EE) 1 - tr(EE) \quad \text{where} \quad m = n(n - 1)(n - 2).$$

D Efficient inversion of $\Omega_E$

To perform the GEE procedure as described in Section 6, we must invert the exchangeable variance matrix $\Omega_E$ as defined in Figure 2. For now, we work in the case where $R = 1$. Since $\Omega_E$ is a real symmetric matrix, its inverse is real and symmetric as well. However, we can say more about the patterns in the inverse $\Omega^{-1}_E$. Recall that $\Omega_E$ has at most six unique terms; call these parameters $\phi$. We find that the inverse $\Omega^{-1}_E$ has at most six unique terms as well. If we define the parameters in $\Omega^{-1}_E$ as $p$, we can write

$$\Omega_E(\phi)\Omega^{-1}_E(p) = I \quad \text{for} \quad \phi, p \in \mathbb{R}^6$$

where $I$ is the $n(n - 1) \times n(n - 1)$ identity. Lastly, we make the conjecture that the parameter pattern in $\Omega^{-1}_E$ is exactly the same as that in $\Omega_E$; we find this conjecture to be true in practice. One caveat is that the locations in which we assume zeros in $\Omega_E$ are not zero in $\Omega^{-1}_E$ in general.
We can find the inverse parameters $p$ from $\phi$ without inverting the entire matrix $\Omega_E$ by instead solving the following linear system

$$C(\phi, n)p = [1, 0, 0, 0, 0, 0]^T \quad \text{for } C(\phi, n) \in \mathbb{R}^{6 \times 6},$$

(14)

where $C(\phi, n)$ is a set of six linear equations based on the parameters $\phi$ and the number of actors $n$ and is depicted in Figure 11. Thus, we replace the need to invert the $n(n-1) \times n(n-1)$ matrix $\Omega_E$ by the inversion of the $6 \times 6$ matrix $C(\phi, n)$. Using this procedure, the computational cost associated with finding the inverse of $\Omega_E$ is independent of the network size.

$$\begin{bmatrix}
\phi_1 & \phi_2 & (n-2)\phi_3 & (n-2)\phi_4 & 2(n-2)\phi_5 & (n-2)(n-3)\phi_6 \\
\phi_2 & \phi_1 & (n-2)\phi_5 & (n-2)\phi_6 & (n-2)(\phi_3 + \phi_4) & (n-2)(n-3)\phi_6 \\
\phi_3 & \phi_5 & \phi_1 + (n-3)\phi_3 & \phi_5 + (n-3)\phi_6 & \phi_2 + \phi_4 + (n-3)(\phi_5 + \phi_6) & (n-3)(\phi_4 + \phi_5 + (n-4)\phi_6) \\
\phi_4 & \phi_5 & \phi_5 + (n-3)\phi_4 & \phi_1 + (n-3)\phi_4 & \phi_2 + \phi_3 + (n-3)(\phi_5 + \phi_6) & (n-3)(\phi_3 + \phi_5 + (n-4)\phi_6) \\
\phi_5 & \phi_4 & \phi_2 + (n-3)\phi_5 & \phi_3 + (n-3)\phi_6 & \phi_1 + \phi_5 + (n-3)(\phi_4 + \phi_6) & (n-3)(\phi_3 + \phi_5 + (n-4)\phi_6) \\
\phi_6 & \phi_6 & \phi_4 + \phi_5 + (n-4)\phi_6 & \phi_3 + \phi_5 + (n-4)\phi_6 & \phi_3 + \phi_4 + 2\phi_5 + 2(n-4)\phi_6 & \phi_1 + \phi_2 + (n-4)(\phi_3 + \phi_4 + 2\phi_5 + (n-5)\phi_6)
\end{bmatrix}$$

Figure 11: Matrix $C(\phi, n)$.

Now consider the case of array data with $R > 1$. Inversion of the exchangeable covariance matrices $\Omega = V[\Xi_v]$ in Figure 6 requires consideration of the patterns in the block matrices. Focusing on Figure 6(a), note that $\Omega_E$ is parametrized by twelve terms. We denote the first six parameters as $\phi^{(1)}$ and the second six $\phi^{(2)}$, corresponding to $\Omega_1$ and $\Omega_2$ respectively. Again the inverse $\Omega^{-1}$ has the exact same block matrix pattern as $\Omega$. Thus, the inverse may be parametrized by $p^{(1)}$ and $p^{(2)}$, each with length six, defined by the following linear equations.

$$C(\phi^{(1)}, n)p^{(1)} + (R - 1)C(\phi^{(2)}, n)p^{(2)} = [1, 0, 0, 0, 0, 0]^T$$

(15)

$$C(\phi^{(2)}, n)p^{(1)} + C(\phi^{(1)}, n)p^{(2)} + (R - 2)C(\phi^{(2)}, n)p^{(2)} = 0_{6 \times 1}$$
This is twelve linear equations in $p^{(1)}$ and $p^{(2)}$. In this formulation we reduce a $Rn(n - 1) \times Rn(n - 1)$ inversion to a $12 \times 12$ inversion for calculation of $\Omega_{E}^{-1}$. Again, note that there is no dependence of the complexity of the inversion on the array dimensions $n$ and $R$. The inverses of the other possible exchangeable covariance matrices in Figure 6 while more complex, can be calculated using a similar procedure that again omits dependence on array dimension $n$.

E Proof of Theorem [1]

For this proof, we adopt slightly different notation to simplify the representation of the exchangeable covariance estimator. Recall that the exchangeable covariance estimator for the OLS estimating equations is

$$
\hat{V}_{E} = (X^{T}X)^{-1}X^{T}\hat{\Omega}_{E}X(X^{T}X)^{-1},
$$

where $\hat{\Omega}_{E}$ is the exchangeable estimate of the error covariance matrix, consisting of five averages of residual products. Here we express $\hat{\Omega}_{E}$ as

$$
\hat{\Omega}_{E} = \sum_{i=1}^{5} \hat{\phi}_{i}S_{i}, \quad \text{where} \quad \hat{\phi}_{i} = \frac{\sum_{(jk,\ell m)\in \Theta_{i}} e_{jk}e_{\ell m}}{|\Theta_{i}|} \quad \text{for} \quad i \in \{1, 2, 3, 4, 5\}. \quad (16)
$$

This amounts to mapping $\sigma^{2} \mapsto \phi_{1}$, $\phi_{a} \mapsto \phi_{2}$, ..., $\phi_{d} \mapsto \phi_{5}$, and re-indexing the $S$ matrices accordingly. Additionally, when we consider sequences of jointly exchangeable random variables $\{W_{ij}\}_{i,j=1}^{n}$, it is understood that the sequence arises from a relational array such that entries with $i = j$ are undefined. Thus, sums over the sequence are of $n(n - 1)$ terms and we define $\sum_{ij} W_{ij} = \sum_{i\neq j} W_{ij}$.

We work in the asymptotic regime of number of actors $n$, where actors are added incrementally to the relational data set. To establish asymptotic normality of $\hat{\beta}$, we wish to
\[
\sqrt{n}(\beta - \beta) \to_d N(0, (\phi_3 + \phi_4 + 2\phi_5)E^{-1}_{XX}), \quad (17)
\]

where \(\to_d\) denotes element-wise convergence in distribution and \(E_{XX} := \mathbb{E}[x_{jk}x_{jk}^T]\) (which is the same for all ordered pairs \((j, k)\) by condition (B1)). The motivation for the proof argument follows from the expression

\[
\sqrt{n}(\beta - \beta) = \left(\frac{\sum_{jk} x_{jk}x_{jk}^T}{n(n-1)}\right)^{-1} \frac{\sqrt{n} \sum_{jk} x_{jk}\xi_{jk}}{n(n-1)}. \quad (18)
\]

We show that \(\left(\frac{\sum_{jk} x_{jk}x_{jk}}{n(n-1)}\right)^{-1}\) converges in probability to \(E^{-1}_{XX}\). Then, it is sufficient to show asymptotic normality of the second multiplicative term in (18).

By condition (B1), the joint exchangeability and independence of non-overlapping pairs of the sequence \(\{\xi_{ij}\}_{i,j=1}^n\) extends to the component sequences in the vectors \(\{x_{ij}\xi_{ij}\}_{i,j=1}^n\). Thus, to show asymptotic normality of \(\beta\), we first prove a theorem stating that the average of jointly exchangeable random variables is asymptotically normal. Specifically, for \(\{W_{ij}\}_{i,j=1}^n\) jointly exchangeable as in Definition 1 we show

\[
k_n \frac{\sum_{ij} W_{ij}}{\sigma} \to_d N(0, 1) \quad (19)
\]

for some normalizing constant \(\sigma\) and fixed sequence \(k_n \to 0\) as \(n \to \infty\).

To prove (19), we provide two supporting lemmas:

- **Lemma 1**: Provides a sufficient condition for asymptotic normality of a sequence of measures based on the standard normal characteristic function.

- **Lemma 2**: Provides a bound for a variance that surfaces in the proof of asymptotic normality in (19).

From (19), we immediately have the marginal asymptotic normality of the mean of the vector
component sequences in \( \{x_{ij} \xi_{ij}\}_{i,j=1}^n \). To establish joint asymptotic normality, we premultiply each \( x_{ij} \xi_{ij} \) by the symmetric square root of its covariance matrix. Joint asymptotic normality of the mean of the vector sequence \( \{x_{ij} \xi_{ij}\}_{i,j=1}^n \) establishes joint asymptotic normality of \( \hat{\beta} \) via (18).

### E.1 Lemmas and theorem in support of Theorem 1

The following is Lemma 2 in [Bolthausen (1982)](1982) and provides a sufficient condition for asymptotic normality. We abuse notation slightly, letting \( i \) be the imaginary unit where appropriate.

**Lemma 1 (Bolthausen (1982)).** Let \( \nu_n \) be a sequence of probability distributions over \( \mathbb{R} \) which satisfies

1. \( \sup_n \int x^2 d\nu_n(x) < \infty \), and
2. for all \( \lambda \in \mathbb{R} \), \( \lim_n \int (i \lambda - x)e^{i \lambda x} d\nu_n(x) = 0 \).

Then, \( \nu_n \to_d N(0,1) \).

To provide intuition for Lemma 1, the integral in condition (2) is identically zero when \( \nu_n \) is standard normal.

The next lemma provides a sufficient condition on the dependence structure in \( \{W_{ij}\}_{i,j=1}^n \) necessary for the proof Theorem 5. Again we emphasize that terms in \( \{W_{ij}\}_{i,j=1}^n \) with \( i = j \) are undefined. The type of counting argument we use in the proof of Lemma 2 resurfaces in the proof of Theorem 2.

**Lemma 2.** Let \( \{W_{ij}\}_{i,j=1}^n \) be a sequence of jointly exchangeable random variables as in Definition 1 with \( ||W_{ij}||_4 < L < \infty \), where \( ||W_{ij}||_p := \mathbb{E}[||W_{ij}||^p]^{1/p} \) for \( p > 0 \). Then,

\[
\frac{1}{n^6} V \left[ \sum_{ij} \sum_{k \in \Theta_{ij}} W_{ij} W_{kl} \right] < \frac{CL^4}{n} \to 0 \quad \text{as} \quad n \to \infty,
\]

(20)
for some $C < \infty$, where $\Theta_{ij}$ is the set of ordered pairs $(k, l)$ that share an index with $(i, j)$.

**Proof.** We start by counting the number of nonzero terms in the variance; a similar argument resurfaces in the proof of Theorem 2. By definition we write

$$
\frac{1}{n^6} V \left[ \sum_{ij} \sum_{kl \in \Theta_{ij}} W_{ij} W_{kl} \right] = \frac{1}{n^6} \sum_{ij} \sum_{kl \in \Theta_{ij}} \sum_{rs \in \Theta_{rs}} \sum_{tu \in \Theta_{tu}} \text{Cov}(W_{ij} W_{kl}, W_{rs} W_{tu}).
$$

Each covariance of (21) is bounded by $L^4$. To bound the variance, we will show the number of nonzero entries in the sum is $O(n^5)$. For Cov$(W_{ij} W_{kl}, W_{rs} W_{tu}) \neq 0$, there must be overlap between the index sets $\{i, j, k, l\}$ and $\{r, s, t, u\}$. Further, the sum in (21) is taken over index sets that themselves contain overlap, i.e. $\{i, j\} \cap \{k, l\} \neq \emptyset$ and $\{r, s\} \cap \{t, u\} \neq \emptyset$. For example, the index sets $\{i, j, i, l\}$ and $\{i, s, i, u\}$ have nonzero covariance in (21). Since there are 5 unique indices in the union of the sets $\{i, j, i, l\}$ and $\{i, s, i, u\}$, there are $O(n^5)$ such index set pairs of this form. There are 96 index set pairs that result in nonzero covariance Cov$(W_{ij} W_{kl}, W_{rs} W_{tu})$. For example, another such pair is $\{i, j, i, l\}$ and $\{i, j, i, j\}$. Each of these 96 index set pairs is $O(n^5)$. Thus, the sum of covariances in (21) is over $O(n^5)$ bounded elements.

**Theorem 5.** Let $\{W_{ij}\}_{i,j=1}^n$ be a sequence of jointly exchangeable random variables as in Definition 1 with at least one of $\{\phi_3, \phi_4, \phi_5\}$ nonzero. If $||W_{ij}||_4 < L < \infty$, then

$$
\frac{\sqrt{n} \sum_{ij} W_{ij}}{n(n-1)} \rightarrow_d N(0, \phi_3 + \phi_4 + 2\phi_5) \quad \text{as} \quad n \rightarrow \infty.
$$

**Proof.** We first show that $\phi_3 + \phi_4 + 2\phi_5$ is the correct limiting variance. Writing the variance of the expression on the left hand side of (22) explicitly and recalling that entries such that
\(i = j\) are undefined, we see

\[
V \left[ \frac{\sqrt{n}}{n(n-1)} \sum_{ij} W_{ij} \right] = \frac{n}{n^2(n-1)^2} \sum_{ij} \sum_{kl} \text{Cov}(W_{ij}, W_{kl})
\]

\[
= \frac{n}{n^2(n-1)^2} \left( n(n-1)(\phi_1 + \phi_2) + n(n-1)(n-2)(\phi_3 + \phi_4 + 2\phi_5) \right)
\]

\[
\to \phi_3 + \phi_4 + 2\phi_5 \text{ as } n \to \infty,
\]

by the properties of joint exchangeability of \(\{W_{ij}\}_{i,j=1}^n\) as described in 3.3. This variance is finite and nonzero by assumption. To prove (22), it is sufficient to show

\[
\bar{S}_n := \frac{\sum_{ij} W_{ij}}{n^{3/2} \sqrt{\phi_3 + \phi_4 + 2\phi_5}} \to_d N(0,1).
\]

Define the limiting variance as \(\sigma_n^2 = n^3(\phi_3 + \phi_4 + 2\phi_5)\) and the sum \(S_n = \sum_{ij} W_{ij}\).

To establish (24), we employ Lemma [1] where \(\nu_n\) is the probability measure corresponding to \(\bar{S}_n\) for \(n \in \{1, 2, \ldots\}\). Condition 1 of Lemma [1] is satisfied since

\[
\mathbb{E}[(\bar{S}_n)^2] = \frac{V \left[ \sum_{ij} W_{ij} \right]}{n^{3}(\phi_3 + \phi_4 + 2\phi_5)} < CL^2
\]

for \(C < \infty\) and all \(n \in \{1, 2, \ldots\}\). Thus, to prove (24), it is sufficient to show Condition 2 of Lemma [1] for all \(\lambda \in \mathbb{R}\),

\[
\mathbb{E} \left[ (i\lambda - \bar{S}_n) e^{i\lambda \bar{S}_n} \right] \to 0 \text{ as } n \to \infty.
\]

We decompose the term in the expectation as in [Guyon and Ludena (1995) and Lumley and Hamblett (2003)]:

\[
(i\lambda - \bar{S}_n) e^{i\lambda \bar{S}_n} = A_1 - A_2 - A_3,
\]
where \( A_1 = i\lambda e^{i\lambda S_n} \left( 1 - \frac{1}{\sigma_n^2} \sum_{ij} W_{ij} S_{ij,n} \right), \quad A_3 = \frac{1}{\sigma_n} \sum_{ij} W_{ij} e^{i\lambda(S_n-S_{ij,n})}, \)
\[ A_2 = \frac{e^{i\lambda \bar{S}_n}}{\sigma_n} \sum_{ij} W_{ij} \left( 1 - i\lambda \bar{S}_{ij,n} - e^{-i\lambda S_{ij,n}} \right), \quad S_{ij,n} = \sum_{kl \in \Theta_{ij}} W_{kl}, \text{ and } \bar{S}_{ij,n} = S_{ij,n}/\sigma_n. \]

To satisfy (26) it remains to be shown that \( \lim_{n \to \infty} E[A_m] = 0 \) for each \( m \in \{1, 2, 3\} \).

\( A_1 \): First notice that \( |e^{i\lambda \bar{S}_n}| \leq 1 \). Using this fact and Lemma 2,
\[ 0 \leq E[|A_1|^2] \leq E[|A_1|^2] \leq \lambda^2 \mathbb{E} \left[ \left| 1 - \frac{1}{\sigma_n^2} \sum_{ij} W_{ij} S_{ij,n} \right|^2 \right] \]
\[ = \frac{\lambda^2}{\sigma_n^4} V \left[ \sum_{ij} \sum_{kl \in \Theta_{ij}} W_{ij} W_{kl} \right] + \lambda^2 \left( 1 - \frac{V \left[ \sum_{ij} W_{ij} \right]}{\sigma_n^2} \right)^2 \]
\[ \leq \lambda^2 CL^4 \frac{1}{n} + \lambda^2 \left( 1 - \frac{\sigma_n^2 + O(n^{-1})}{\sigma_n^2} \right)^2 \]
\[ = \lambda^2 \left( \frac{CL^4}{n} + \frac{O(n^{-2})}{\sigma_n^2} \right) \to 0 \]

for all real \( \lambda \). \( E[|A_1|^2] \) limiting to zero implies \( E[|A_1|] \) limits to zero, and hence \( E[A_1] \) limits to zero.

\( A_2 \): By Taylor expansion of \( e^{-i\lambda \bar{S}_{ij,n}} \), we can write
\[ \left| 1 - i\lambda \bar{S}_{ij,n} - e^{-i\lambda S_{ij,n}} \right| \leq c\lambda^2 \left( \bar{S}_{ij,n} \right)^2, \]
for some \( 0 < c < \infty \) and all \( n, \lambda \). Using this bound and the fact that \( |\Theta_{ij}| = 4n - 6 \), we evaluate \( E[|A_2|] \) directly below:
\[ \mathbb{E}[|A_2|] \leq \frac{1}{\sigma_n} \mathbb{E} \left[ \sum_{ij} |W_{ij}| \left| 1 - i\lambda \bar{S}_{ij,n} - e^{-i\lambda \bar{S}_{ij,n}} \right| \right], \quad (33) \]
\[ \leq \frac{c\lambda^2}{\sigma_n^3} \sum_{ij} \mathbb{E} \left[ |W_{ij}| (S_{ij,n})^2 \right], \quad (34) \]
\[ \leq \frac{c\lambda^2 n(n-1)(4n-6)2^3 L^3}{\sigma_n^3} \to 0, \quad (35) \]

for all real \( \lambda \). As \( \mathbb{E}[|A_2|] \) limits to zero, so does \( \mathbb{E}[A_2] \).

**A_3**: Note that \( S_{ij,n} \) sums all terms in the sequence \( \{W_{ij}\}_{i,j=1}^n \) that depend upon \( W_{ij} \), including \( W_{ij} \) itself. Thus, \( W_{ij} \) and \( \bar{S}_n - \bar{S}_{ij,n} \) are independent. It follows immediately that
\[
\mathbb{E} \left[ \frac{1}{\sigma_n} \sum_{ij} W_{ij} e^{i\lambda(\bar{S}_n - \bar{S}_{ij,n})} \right] = \frac{1}{\sigma_n} \sum_{ij} \mathbb{E}[W_{ij}] \mathbb{E}[e^{i\lambda(S_{ij,n} - \bar{S}_{ij,n})}] = 0, \quad (36)
\]
since \( \mathbb{E}[W_{ij}] = 0 \) for all ordered pairs \((i, j)\).

Hence, \( \lim_{n \to \infty} \mathbb{E}[A_m] = 0 \) for each \( m \in \{1, 2, 3\} \) and we have the convergence in \( (26) \), implying \( \bar{S}_n \to_d N(0,1) \) by Lemma \([1]\) which gives the desired result in \( (22) \).

**E.2 Proof of asymptotic normality of \( \hat{\beta} \)**

*Proof.* We begin by writing
\[
\sqrt{n}(\hat{\beta} - \beta) = \left( \frac{\sum_{jk} x_{jk} x_{jk}^T}{n(n-1)} \right)^{-1} \frac{\sqrt{n} \sum_{jk} x_{jk} \xi_{jk}}{n(n-1)}, \quad (37)
\]
again emphasizing that entries in the sum with \( j = k \) are undefined and omitted. The sum \( \sum_{jk} x_{jk} x_{jk}^T \) is of independent and identically distributed random variables with \( \mathbb{E}[(x_{jk} x_{jk}^T)^2] < \infty \) by conditions (B1) and (B2). Further, the inverse map is continuous.
Thus, by the weak law of large numbers and the continuous mapping theorem, we have

\[
\left( \frac{\sum_{jk} x_{jk} x_{jk}^T}{n(n-1)} \right)^{-1} \to_p E^{-1}_{XX}. \tag{38}
\]

We now analyze the second multiplicative term in (37). Showing asymptotic normality of this term is sufficient to show asymptotic normality of the expression on the left hand side of (37). Recall \( x_{jk}^T = [x_{jk}^{(1)}, x_{jk}^{(2)}, \ldots, x_{jk}^{(p)}] \). We wish to show that the sum of vectors

\[
\frac{\sqrt{n}}{n(n-1)} \sum_{jk} E^{-1/2}_{XX} x_{jk} \xi_{jk} \to_d N(0, I_p \sigma_0^2), \tag{39}
\]

for some limiting variance \( \sigma_0^2 \). Note that each component of the vector \( \tilde{x}_{jk} \xi_{jk} \), where \( \tilde{x}_{jk} := E^{-1/2}_{XX} x_{jk} \), is independent every other. Thus, it is sufficient to show marginal asymptotic normality of each vector component \( \tilde{x}_{jk}^{(l)} \xi_{jk} \) for \( l \in \{1, 2, \ldots, p\} \) to establish joint asymptotic normality of the vector on the left hand side of (39).

Without loss of generality, consider the first component \( \tilde{x}_{jk}^{(1)} \xi_{jk} \) and the corresponding sequence of scalars \( \{\tilde{x}_{jk}^{(1)} \xi_{jk}\}_{j,k=1}^n \). By the independence of \( X \) and \( \Xi \) in (B3), this is a mean-zero exchangeable sequence of scalar random variables. Additionally, the condition of finite moments in (B2) implies that \( \sup_{jk} ||\tilde{x}_{jk}^{(1)} \xi_{jk}||_4 < L \). Thus, we apply Theorem 5 with \( \sigma_n^2 = n^3 V[\sum_{jk} \tilde{x}_{jk}^{(1)} \xi_{jk}] = n^3(\phi_3 + \phi_4 + 2\phi_5) \), which gives that

\[
\frac{\sqrt{n} \sum_{jk} \tilde{x}_{jk}^{(1)} \xi_{jk}}{n(n-1)} \to_d N(0, \phi_3 + \phi_4 + 2\phi_5). \tag{40}
\]

This holds for all components \( \tilde{x}_{jk}^{(l)} \xi_{jk} \) for \( l \in \{1, 2, \ldots, p\} \). By independence of the vector components of \( \tilde{x}_{jk} \xi_{jk} \), the marginal asymptotic normality of each component of \( \sum_{jk} \tilde{x}_{jk} \xi_{jk} \) implies joint normality of the vector, and we have

\[
\frac{\sqrt{n} \sum_{jk} x_{jk} \xi_{jk}}{n(n-1)} \to_d N(0, (\phi_3 + \phi_4 + 2\phi_5) E_{XX}). \tag{41}
\]
Combining the convergence in probability in (38) and the asymptotic normality of (41), we obtain the desired result.

F Proof of Theorem 2

For this proof, we adopt the same change in notation as in Appendix E defined in (16). We deviate slightly in that we denote Θ_i to denote dyadic pairs (j, k) and (l, m) that share a member in the i^{th} manner. For example, for i = 3 we must have j = l and m ≠ k. We use the same assumptions as in Theorem 1.

This proof is outlined as follows. We initially prove that the exchangeable estimator \( \hat{V}_E \) is consistent if the exchangeable parameter estimates \{\hat{\phi}_i : i = 1, \ldots, 5\} are consistent for the true parameters. We then prove consistency of \( \hat{\phi}_i \) in two steps: (a) we show parameter estimates \( \tilde{\phi}_i \) based on the unobserved true errors \( \Xi \) are consistent and then (b) we show that the parameter estimates \( \hat{\phi}_i \) are asymptotically equivalent to \( \tilde{\phi}_i \). We require the consistency of \( \hat{\beta} \) result (implied by Theorem 1) for this last step.

F.1 Proof of consistency of \( \hat{V}_E \)

Proof of Theorem 2 We first note that from Theorem 1 the order of convergence of \( \hat{\beta} \) is \( \sqrt{n} \). Thus, we choose the rate \( n \) as our asymptotic regime for consistency of \( \hat{V}_E \). We wish to show that

\[
n\hat{V}_E - nV[\hat{\beta}] \to_p 0.
\]

(42)

1. Sufficient to show consistency of \( \{\hat{\phi}_i\} \)

Here we show that to prove consistency of \( \hat{V}_E \), it is sufficient to prove the consistency of the parameter estimates \( \{\hat{\phi}_i\} \) for the true parameters. We begin by writing the difference of
variances in (42) as

\[ n\hat{v}_E - n\hat{v}[\hat{\beta}] = n(XX^T)^{-1}X^T(\Omega_E - \Omega_E)XX^T(XX^T)^{-1} \]

\[ = \frac{n}{n^2(n-1)^2} \left( \frac{XX^T}{n(n-1)} \right)^{-1} \left( \frac{XX^T}{n(n-1)} \right) \left( \frac{XX^T}{n(n-1)} \right)^{-1} \]

\[ = \sum_{i=1}^{5} \frac{\vert\Theta_i\vert}{n(n-1)^2} \left( \hat{\phi}_i - \phi_i \right) \left( \frac{XX^T}{n(n-1)} \right) \left( \frac{XX^T}{n(n-1)} \right)^{-1} \left( \sum_{(j,k,\ell,m) \in \Theta_i} x_{jk}x_{\ell m}^T \right) \left( \frac{XX^T}{n(n-1)} \right)^{-1} \]

\[ : = \sum_{i=1}^{5} c_i \left( \hat{\phi}_i - \phi_i \right) h_i(X), \quad (43) \]

where \( c_i = \vert\Theta_i\vert/n(n-1) \) and \( h_i(X) \) contains the remaining terms which are functions of \( X \).

By the counting argument used to show Lemma 2, each \( \vert\Theta_i\vert \) is at most \( O(n^3) \), so each \( c_i \rightarrow d_i \) for some constant \( d_i \). To obtain the result in (42), it is sufficient then to show \( \hat{\phi}_i - \phi_i \rightarrow_p 0 \) for each \( i \) and each \( h_i(X) \) converges in probability to some constant. We focus on the latter first.

We will show that the expectation of \( h_i(X) \) is a constant matrix for all \( i \) and \( n \) and show that the variance of \( h_i(X) \) tends to zero for all \( i \). Recall that we previously established convergence of the inverted terms in \( h_i(X) \) to constant matrix \( E_{XX}^{-1} \) in (38). When \( i = 1 \), the middle term in \( h_1(X) \) has \( (j,k) = (\ell,m) \) and the argument used to prove consistency in (38) applies, yielding \( h_1(X) \rightarrow_p E_{XX}^{-1} \). For \( i > 1 \), each entry in the middle sum has expectation

\[ \mathbb{E}[x_{jk}x_{\ell m}^T] = \mathbb{E}[x_{jk}] \mathbb{E}[x_{jk}]^T \]

for all \( n \) by the assumption of a random sample (B1). We now examine the variance of each \( h_i(X) \) for \( i > 1 \):

\[ V \left[ \frac{\sum_{(j,k,\ell,m) \in \Theta_i} x_{jk}x_{\ell m}^T}{\vert\Theta_i\vert} \right] = \frac{1}{\vert\Theta_i\vert^2} \sum_{(j,k,\ell,m) \in \Theta_i} \sum_{(r,s,t,u) \in \Theta_i} \text{Cov} (x_{jk}x_{\ell m}^T, x_{rs}x_{tu}^T). \quad (44) \]

We utilize a counting argument similar to that in Lemma 2 noting each covariance in (44) is bounded by condition (B2). Each sum is over \( \vert\Theta_i\vert \) terms, so there are \( \vert\Theta_i\vert^2 \) summed terms for each \( i > 1 \). The covariance between \( x_{jk}x_{\ell m}^T \) and \( x_{rs}x_{tu}^T \) is nonzero only if one pair of
indices from the second sum equals a pair from the first, e.g. \((r, s) = (j, k)\). This reduces the number of nonzero covariances from the maximum possible \(|\Theta_i|^2\) by at least a factor of \(n\) for each \(i\). For example, consider the case of \(i = 3\) corresponding to dyads with a common sender (see Figure 1(c)), where \(|\Theta_3| = O(n^3)\). Each set of indices in \(\Theta_3\) must be of the form \((j, k, j, m)\), and thus, for there to be nonzero covariance, the second set of indices must be of the form \((j, k, j, u)\), for example. The set of indices \(\{j, k, j, m, j, k, j, u\}\) is of order \(O(n^4) = |\Theta_3|^2n^{-2}\). Note that we can consider other forms in the indices of the second sum that give rise to nonzero covariance, such as \((j, s, j, m)\) and so on. However, there are four such combinations, all of which are \(O(n^4)\). Thus, the number of nonzero covariances is \(O(n^4)\). This same argument holds for all \(i \in \{2, \ldots, 5\}\), and thus we have

\[
V \left[ \frac{\sum_{(jk, \ell m) \in e_i} x_{jk} x_{\ell m}^T}{|\Theta_i|} \right] = \frac{|\Theta_i|^2 O(n^{-1})}{|\Theta_i|^2} \to 0. \tag{45}
\]

Since the expectation tends to a constant and variance to zero, we have that

\[
\frac{\sum_{(jk, \ell m) \in e_i} x_{jk} x_{\ell m}^T}{|\Theta_i|} \to_p \begin{cases} E_{XX} & \text{for } i = 1 \\ E[X_{jk}]E[X_{jk}]^T & \text{for } i > 1. \end{cases} \tag{46}
\]

which implies \(h_i(X) \to_p \begin{cases} E_{XX}^{-1} & \text{for } i = 1 \\ E_{XX}^{-1}E[X_{jk}]E[X_{jk}]^T E_{XX}^{-1} & \text{for } i > 1. \end{cases} \tag{47}
\]

We have shown \(c_i\) and \(h_i(X)\) in (43) both converge in probability to constants. Thus, to show consistency of \(\hat{V}_E\) in (42), it is sufficient to show convergence in probability of \(\hat{\phi}_i - \phi_i\) to zero for each \(i \in \{1, \ldots, 5\}\).

We now consider consistency of the parameter estimates \(\hat{\phi}_i\). First, define error averages
\{\tilde{\phi}_i : i = 1, 2, 3, 4, 5\} analogous to the parameter estimates, such that for each \(i\)

\[
\tilde{\phi}_i = \frac{1}{|\Theta_i|} \sum_{(jk,\ell m) \in \Theta_i} \xi_{jk} \xi_{\ell m}.
\]

(48)

We will show \(\tilde{\phi}_i - \phi_i\) converges in probability to zero, and then do the same for \(\hat{\phi}_i - \tilde{\phi}_i\). This is sufficient for showing \(\hat{\phi}_i - \phi_i \to_p 0\) as \(\hat{\phi}_i - \phi_i = (\hat{\phi}_i - \tilde{\phi}_i) + (\tilde{\phi}_i - \phi_i)\).

2. Consistency of \(\tilde{\phi}_i\) for \(\phi_i\)

To show convergence in probability of \(\tilde{\phi}_i - \phi_i\) to zero, we use the argument that the bias and variance both tend to zero. By assumption (A1), \(E[\xi_{jk} \xi_{\ell m}] = \phi_i\) for every index pair \((jk, \ell m) \in \Theta_i\). Thus, \(E[\tilde{\phi}_i - \phi_i] = 0\) for all \(n\) and \(i \in \{1, \ldots, 5\}\). We now turn to the variance:

\[
V[\tilde{\phi}_i] = \frac{1}{|\Theta_i|^2} \sum_{(jk,\ell m) \in \Theta_i} \sum_{(rs,\ell u) \in \Theta_i} \text{Cov}(\xi_{jk} \xi_{\ell m}, \xi_{rs} \xi_{\ell u}).
\]

(49)

We again make a counting argument similar to that in Lemma 2. By condition (B2), each of the \(|\Theta_i|^2\) covariances in the sum above are bounded. The covariance between \(\xi_{jk} \xi_{\ell m}\) and \(\xi_{rs} \xi_{\ell u}\) is nonzero only if there is overlap between their two index sets. This reduces the number of nonzero covariances from the maximum possible \(|\Theta_i|^2\) by a factor of at least \(n\).

Again, consider the case of \(i = 3\) where \(|\Theta_3| = \mathcal{O}(n^3)\). Each set of indices in \(\Theta_3\) must be of the form \((j, k, j, m)\), and thus the second set of indices must be of the form \((j, s, j, u)\), for example, for the covariance to be nonzero. The set of indices \\{\(j, k, j, m, j, s, j, u\)\} is of order \(\mathcal{O}(n^5) = |\Theta_3|^2 n^{-1}\). There are other forms of indices in the second sum that give rise to nonzero covariance, such as \((k, s, k, u)\) and so on. However, there are nine such forms, each of which is \(\mathcal{O}(n^5)\). Thus, the number of nonzero covariances is \(\mathcal{O}(n^5)\), and hence, we have

\[
V[\tilde{\phi}_i] = \frac{|\Theta_i|^2 \mathcal{O}(n^{-1})}{|\Theta_i|^2} \to 0.
\]

(50)

This same argument holds for all \(i\), and thus, we have the desired consistency: \(\tilde{\phi}_i - \phi_i \to_p 0\)
for \( i = 1, \ldots, 5 \).

3. Asymptotic equivalence of \( \hat{\phi}_i \) and \( \tilde{\phi}_i \)

We now show that \( \hat{\phi}_i - \tilde{\phi}_i \) converges in probability to zero. We first write the expression in terms of the estimated coefficients \( \hat{\beta} \):

\[
\hat{\phi}_i - \tilde{\phi}_i = \sum_{(jk, \ell m) \in \Theta_i} e_{jk} e_{\ell m} - \xi_{jk} \xi_{\ell m} = \frac{1}{|\Theta_i|} \sum_{(jk, \ell m) \in \Theta_i} \left( (\beta - \hat{\beta}_n)^T (x_{jk} x_{\ell m}^T) (\beta - \hat{\beta}_n) - (\beta - \hat{\beta}_n)^T (\xi_{jk} x_{\ell m} + \xi_{\ell m} x_{jk}) \right). \tag{51}
\]

By Theorem 1, \( \hat{\beta} - \beta \) converges to zero in probability. By Slutsky’s theorem, if the terms in (51) involving elements of \( X \) and \( \Xi \) converge in probability to any constant, then \( \hat{\phi}_i - \tilde{\phi}_i \) converges in probability to zero. In (46) we showed convergence of the term involving \( x_{jk} x_{\ell m}^T \).

Furthermore, by condition (B3), we have that \( \mathbb{E}[\xi_{jk} x_{\ell m}] = \mathbb{E}[\xi_{\ell m} x_{jk}] = 0 \). It remains to be shown that the variance of the error-covariate averages tend to zero. Consider the variance of the first error-covariate averages:

\[
V \left[ \frac{1}{|\Theta_i|} \sum_{(jk, \ell m) \in \Theta_i} \xi_{jk} x_{\ell m} \right] = \frac{1}{|\Theta_i|^2} \sum_{(jk, \ell m) \in \Theta_i} \sum_{(rs, tu) \in \Theta_i} \text{Cov} \left( \xi_{jk} x_{\ell m}, \xi_{rs} x_{tu} \right), \tag{52}
\]

\[
= \frac{1}{|\Theta_i|^2} \sum_{(jk, \ell m) \in \Theta_i} \sum_{(rs, tu) \in \Theta_i} \mathbb{E} \left[ x_{\ell m} x_{tu}^T \right] \text{Cov} \left( \xi_{jk}, \xi_{rs} \right). \tag{53}
\]

In writing (53), we use condition (B3) and simplify by conditioning on \( X \) and using the law of total variance. By the same counting arguments used to establish (45) and (50), there are \( |\Theta_i|^2 \mathcal{O}(n^{-1}) \) nonzero bounded covariances in (53). Thus, we have

\[
V \left[ \frac{1}{|\Theta_i|} \sum_{(jk, \ell m) \in \Theta_i} \xi_{jk} x_{\ell m} \right] = \frac{|\Theta_i|^2 \mathcal{O}(n^{-1})}{|\Theta_i|^2} \rightarrow 0. \tag{54}
\]
Since the expectation and variance both tend to zero, we have

\[
\frac{1}{|\Theta_i|} \sum_{(jk, \ell m) \in \Theta_i} \xi_{jk} x_{\ell m} \rightarrow_p 0. \tag{55}
\]

The same argument applies to the second error-covariate term in \((51)\). Thus, we have shown that consistency of \(\widehat{\beta}\) implies

\[
\widehat{\phi}_i - \tilde{\phi}_i \rightarrow_p 0. \tag{56}
\]

\(\Box\)

### G Proof of Theorem 3

For the purposes of simplicity, we adopt the slight notation modifications introduced in equation \((16)\) at the beginning of Appendix E here as well. In addition, we use a single index, say \(k\), to denote an ordered dyad \((i, j)\). Thus, rather than indexing a network element \(y_{ij}\), we instead denote it \(y_k\).

**Proof.** The dyadic clustering and exchangeable covariance estimators in the simple linear regression setting may be expressed

\[
\hat{V}_{DC} = (Z^T Z)^{-2} \sum_{i=1}^{5} \sum_{(j,k) \in \Theta_i} z_j z_k e_j e_k;
\]

\[
\hat{V}_{E} = (Z^T Z)^{-2} \sum_{i=1}^{5} \hat{\phi}_i \sum_{(j,k) \in \Theta_i} z_j z_k \quad \text{where} \quad \hat{\phi}_i = \frac{\sum_{(j,k) \in \Theta_i} e_j e_k}{|\Theta_i|},
\]

and \(Z^T = [z_1, ..., z_{n(n-1)}]\) is the vector of covariates for all relations. We emphasize that here the covariate \(Z\) is treated as fixed, rather than coming from some population distribution.

To calculate the bias of each estimators, we need the expectation of the residual products \(\{e_j e_k\}\) in the above expressions. We derive this below and denote the true variance of
the regression coefficient for a given covariate vector $Z$ as $V^*$. (Recall under the stated assumptions in the theorem, $V^* = (Z^T Z)^{-2} \sum_{i=1}^{5} \phi_i \sum_{(j,k) \in \Theta_i} z_j z_k$.)

Since each residual is defined $e_j = y_j - z_j \hat{\beta} = \xi_j - z_j (\hat{\beta} - \beta)$, the product of the $j^{th}$ and $k^{th}$ residuals is

$$
e_j e_k = \xi_j \xi_k - \xi_j z_j (\hat{\beta} - \beta) - \xi_j z_k (\hat{\beta} - \beta) + z_j z_k (\hat{\beta} - \beta)^2 = \xi_j \xi_k - e_k z_j (\hat{\beta} - \beta) - e_j z_k (\hat{\beta} - \beta) - z_j z_k (\hat{\beta} - \beta)^2,$$

where the last equation results from using the identity $\xi_i = e_i + z_i (\hat{\beta} - \beta)$ in the second and third terms of the first equation. Substituting $\hat{y}_k = z_k \hat{\beta}$, we find

$$\mathbb{E}[e_j e_k] = \mathbb{E}[\xi_j \xi_k] - z_j z_k \mathbb{E}[(\hat{\beta} - \beta)^2]$$

since the residuals and fitted values resulting from OLS are orthogonal (i.e. $\mathbb{E}[e_j \hat{y}_k] = 0$). In addition, $\mathbb{E}[(\hat{\beta} - \beta)^2] = V[\hat{\beta}] = V^*$ since $\hat{\beta}$ is unbiased for a fixed $Z$. The pair of relations $j$ and $k$ must be one of six relation types shown in Figure 1. Suppose the pair is of type $i$; then,

$$\mathbb{E}[e_j e_k] = \phi_i - z_j z_k V^*. \quad (57)$$
It follows that

\[
\text{Bias}(\hat{V}_{\text{DC}}) = -V^* + (Z^T Z)^{-2} \sum_{i=1}^5 \sum_{(j,k) \in \Theta_i} z_j z_k \mathbb{E}[e_j e_k]
\]

\[
= -V^* + (Z^T Z)^{-2} \sum_{i=1}^5 \sum_{(j,k) \in \Theta_i} z_j z_k (\phi_i - z_j z_k V^*) \quad \text{by (57)}
\]

\[
= -V^* + (Z^T Z)^{-2} \sum_{i=1}^5 \left( \phi_i \sum_{(j,k) \in \Theta_i} z_j z_k - V^* \sum_{(j,k) \in \Theta_i} z_j^2 z_k^2 \right)
\]

\[
= -(Z^T Z)^{-2} \sum_{i=1}^5 V^* \sum_{(j,k) \in \Theta_i} z_j^2 z_k^2.
\]

A similar argument establishes the result for the exchangeable estimator.

\[
\text{Bias}(\hat{V}_E) = -V^* + (Z^T Z)^{-2} \sum_{i=1}^5 \mathbb{E}[\hat{\phi}_i] \sum_{(j,k) \in \Theta_i} z_j z_k
\]

\[
= -V^* + (Z^T Z)^{-2} \sum_{i=1}^5 \left( \frac{\sum_{(s,t) \in \Theta_i} \mathbb{E}[\epsilon_s \epsilon_t]}{|\Theta_i|} \right) \sum_{(j,k) \in \Theta_i} z_j z_k
\]

\[
= -V^* + (Z^T Z)^{-2} \sum_{i=1}^5 \left( \frac{\sum_{(s,t) \in \Theta_i} (\phi_i - z_s z_t V^*)}{|\Theta_i|} \right) \sum_{(j,k) \in \Theta_i} z_j z_k \quad \text{by (57)}
\]

\[
= -V^* + (Z^T Z)^{-2} \sum_{i=1}^5 \left( \phi_i \sum_{(j,k) \in \Theta_i} z_j z_k - V^* \frac{\sum_{(s,t) \in \Theta_i} z_s z_t}{|\Theta_i|} \right)^2
\]

\[
= -(Z^T Z)^{-2} \sum_{i=1}^5 V^* \frac{\left( \sum_{(j,k) \in \Theta_i} z_j z_k \right)^2}{|\Theta_i|}
\]

The result in the theorem follows directly from the Cauchy-Schwarz inequality. Define a new set of vectors \( \{W_i : i = 1, 2, \ldots, 5\} \), such that \( W_i \) contains elements \( w_{im} = z_j z_k \) for \( m \in \{1, 2, \ldots, |\Theta_i|\} \).
We now argue the numerator is always larger than the denominator in the above expression. Consider the difference of the $i^{th}$ terms in the numerator and denominator:

\[
\text{num.} - \text{denom.} = \sum_{m=1}^{|\Theta_i|} w_{im}^2 - |\Theta_i|^{-1} \left( \sum_{m=1}^{|\Theta_i|} w_{im} \right)^2 \propto \sum_{m=1}^{|\Theta_i|} w_{im}^2 - \left( \sum_{m=1}^{|\Theta_i|} w_{im} \right)^2 \geq 0.
\]

This last expression is proportional to the sample variance of the entries in $W_i$ and hence is positive. This implies that for each $i$, \( \sum_{m=1}^{|\Theta_i|} w_{im}^2 \geq |\Theta_i|^{-1} \left( \sum_{m=1}^{|\Theta_i|} w_{im} \right)^2 \) and it follows that the ratio of the absolute values of the biases in (58) is greater than or equal to one.

\[\square\]

**H Eigenvalues of $\Omega_E$**

Since the entries in the exchangeable covariance matrix estimator $\hat{\Omega}_E$ are empirical averages, it is possible the estimate is not positive definite. Here we briefly investigate the constraints on the parameters that guarantee the resulting covariance matrix is positive definite for $R = 1$. Note that for computing the sandwich estimator variance of $\hat{\beta}$ and making inference on $\hat{\beta}$, positive definiteness of $\hat{\Omega}_E$ is not necessary. However, if a GEE procedure is employed, the inverse of the covariance matrix estimator is required, and hence positive definiteness of $\hat{\Omega}_E$ is desired.
H.1 Undirected networks

We focus first on the undirected case, where the exchangeable covariance matrix contains two distinct nonzero entries: a variance $\sigma^2$ and a parameter $\phi$ in the off-diagonal representing the correlation between any dyad pairs that share an actor. Below we consider the correlation matrix, rather than the covariance matrix, which contains only nonzero correlation value. We denote this value by $a$, and note that $a = \phi/\sigma^2$.

Based on a thorough empirical investigation, we conjecture that the exchangeable correlation matrix corresponding to a undirected network of $n$ individuals, which has nonzero value $a$ in select off-diagonal entries, has exactly three eigenvalues as given below.

| Eigenvalue          | Multiplicity |
|---------------------|--------------|
| $1 + 2(n - 2)a$     | 1            |
| $1 - 2a$            | $\frac{1}{2}n(n - 3)$ |
| $1 + (n - 4)a$      | $n - 1$      |

The correlation matrix is positive definite if and only if all eigenvalues are positive. Thus, if $a \in \left(\frac{-1}{2(n-2)}, \frac{1}{2}\right)$, the correlation matrix is positive definite. Notice that the upper bound on $a$ does not vary with $n$. Using the relation between $a$ and $\{\sigma^2, \phi\}$, this constraint can be re-expressed as a constraint on the covariance parameters.

H.2 Directed networks

We find empirically that the directed covariance matrix $\Omega_E$ has five unique eigenvalues. Further, each of the eigenvalues are contained within the set of six eigenvalues of the matrix $C$, introduced in Appendix D and used in computation of the inverse of the exchangeable covariance matrix. As $C$ is a bilinear function of $\Omega_E$, this observation does not appear implausible. We may construct $C = A^T \Omega_E B$ for $A, B \in \mathbb{R}^{n(n-1\times 6)}$ and $A^T B = I$. One such pair is $B$ taken to be the first column of $S_1$ thought $S_6$ and $A$ taken to be all zeros except for a single 1 in each column occupying rows $\{1, n, 2n, 2, n + 1, n(n - 1)\}$, respectively.
In analyzing the eigenvalues of the directed covariance matrix $\Omega_E$, we again focus on the exchangeable correlation matrix which contains four nonzero off-diagonal elements $\{a, b, c, d\}$ corresponding, respectively, in placement to $\{\phi_a, \phi_b, \phi_c, \phi_d\}$ in the exchange covariance matrix $\Omega_E$. Note $a = \phi_a/\sigma^2$, $b = \phi_b/\sigma^2$, and so on. Based on an eigenvalue analysis of $C$ and various empirical studies, we conjecture the eigenvalues for the exchangeable correlation matrix associated with a directed network of $n$ individuals has exactly five eigenvalues as given below.

| Eigenvalue(s) | Multiplicity |
|---------------|--------------|
| $1 + a + (n - 2)(b + c) + 2(n - 2)d$ | 1 |
| $1 + a - (b + c + 2d)$ | $(n - 1)(n - 2)/2 - 1$ |
| $1 - (a + b + c) + 2d$ | $(n - 1)(n - 2)/2$ |
| $((n - 3)(b + c) - 2d + 2)/2 \pm \sqrt{(\alpha + \beta)/2}$ | $n - 1$ |

where $\alpha = (c^2 + b^2)(n^2 - 2n + 1) + 4d^2(n^2 - 6n + 9) + 2bc(1 - n^2 + 2n)$ and $\beta = ad(8n - 24) + (b + c)d(12 - 4n) + 4a(a - (b + c))$. As in the undirected case, these constraints can be re-expressed as constraints on the original five covariance parameters.