Dyadic Regression

August 23, 2019

Bryan S. Graham[1]

Initial Draft: September 2018, This Draft: August 2019

Abstract

Dyadic data, where outcomes reflecting pairwise interaction among sampled units are of primary interest, arise frequently in social science research. Regression analyses with such data feature prominently in many research literatures (e.g., gravity models of trade). The dependence structure associated with dyadic data raises special estimation and, especially, inference issues. This chapter reviews currently available methods for (parametric) dyadic regression analysis and presents guidelines for empirical researchers.

JEL Classification: C14, C55, C57

Keywords: Networks, Dyadic Regression, U-Statistics

[1]Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880 and National Bureau of Economic Research, e-mail: bgraham@econ.berkeley.edu, web: `http://bryangraham.github.io/econometrics/`.
This is a draft chapter for the edited volume *The Econometric Analysis of Network Data* being prepared by Aureo de Paula and myself. It is based upon lecture notes prepared for a series of short courses on the econometrics of networks. I thank Michael Jansson for several useful discussions as well as participants in short courses in Olso, Norway (September, 2017), Hejnice, Czechia (February 2018), St. Gallen, Switzerland (October 2018), Annweiler, Germany (July 2019) and Prague, Czechia (August 2019) for useful feedback. Portions of this material were also presented at an invited session of the 2018 LACEA/LAMES meetings in Guayaquil, Ecuador. All the usual disclaimers apply. Financial support from NSF grant SES #1851647 is gratefully acknowledged.

Let Y_{ij} equal total exports from country i to country j as in Tinbergen (1962); here i and j are two of N independent random draws from a common population. Let W_i be a vector of country attributes and $R_{ij} = r(W_i, W_j)$ a vector of constructed dyad-specific attributes; R_{ij} typically includes the logarithm of both exporter and importer gross domestic product (GDP), the physical distance between i and j, as well as other variables (e.g., indicators for sharing a land border or belonging to a common customs union). The analyst, seeking to relate Y_{ij} and R_{ij}, posits the relationship

$$Y_{ij} = \exp\left(R_{ij}'\theta_0\right) A_i B_j V_{ij}, \tag{1}$$

with A_i, B_i and V_{ij} mean one random variables and $\{(V_{ij}, V_{ji})\}_{1 \leq i \leq N-1, j>i}$ independent of $\{W_i, A_i, B_i\}_{i=1}^{N}$ and independently and identically distributed across the $\binom{N}{2} = \frac{1}{2}N(N-1)$ dyads. Here the $\{A_i\}_{i=1}^{N}$ and $\{B_i\}_{i=1}^{N}$ sequences correspond, respectively, to (unobserved) exporter and importer heterogeneity terms. These terms are sometimes referred to as "multilaterial resistance" terms by empirical trade economists. For example, a high A_i might reflect an unmodeled export orientation of an economy or an undervalued currency. Similarly, a high B_i might capture unmodeled tastes for consumption. Head & Mayer (2014) survey the gravity model of trade, including its theoretical foundations.

Conditional on the exporter and importer effects we have

$$\mathbb{E}\left[Y_{ij} \middle| W_i, W_j, A_i, B_j\right] = \exp\left(R_{ij}'\theta_0\right) A_i B_j.$$

If, additionally, $\mathbb{E}\left[(A_i, B_i)' \middle| W_i\right] = (1,1)'$, such that W_i does not covary with the exporter and importer "multilaterial resistance" terms[2], then unconditional on A_i and B_j we have the *dyadic regression* function

$$\mathbb{E}\left[Y_{ij} \middle| W_i, W_j\right] = \exp\left(R_{ij}'\theta_0\right). \tag{2}$$

Interpret (2) as follows: draw countries i and j independently at random and record their values of W_i and W_j. Given this information set what is the mean square error (MSE) minimizing predictor of Y_{ij}? Equation (2) gives a parametric form for this prediction/regression function. This chapter surveys methods of estimation of, and inference on, θ_0.

Santos Silva & Tenreyro (2006) recommended estimating θ_0 by maximizing a Poisson pseudo log-likelihood with a conditional mean function given by (2) (cf. Gourieroux et al., 1984). For inference they constructed standard errors using the sandwich formula of Hu-

[2]If, for example, a subset of $\mathbb{W}$ is associated with membership in the World Trade Organization (WTO), then reasoning about this condition involves asking whether countries belonging to the WTO have a greater latent propensity to export or import? In what follows I entirely defer consideration of these questions and focus solely on the inferential issues raised by the network structure.

ber (1967); implicitly assuming that the $\{Y_{ij}\}_{1\leq i,j\leq N, i\neq j}$ are conditionally independent of one another given $\mathbf{W} = (W_1, \ldots, W_N)'$. In practice this conditional independence assumption, although routinely made in the empirical trade literature (e.g., Rose, 2004; Baldwin & Taglioni, 2007), is very unlikely to hold. Exports from, say, Japan to Korea likely covary with those from Japan to the United States. This follows because A_i – the Japan exporter effect – drives Japanese exports to both Korea and the United States. It is also possible that exports from Japan to Korea may covary with those from Korea to Thailand; perhaps because A_j and B_j – the Korean exporter and importer effects – covary (as would be true if there exist common unobserved drivers of Korean exporting and importing behavior).[3]

Loosely following Fafchamps & Gubert (2007) I call the above patterns of dependence "dyadic dependence" or "dyadic clustering". Consider two pairs of dyads, say $\{i_1, i_2\}$ and $\{j_1, j_2\}$, if these dyads share an agent in common – for example $i_1 = j_1$ – then $Y_{i_1 i_2}$ and $Y_{j_1 j_2}$ will covary. Failing to account for dependence of this type will, typically, result in standard errors which are too small and consequently more Type I errors in inference than is desired (e.g., Cameron & Miller, 2014; Aronow et al., 2017).

In this chapter I describe how to estimate and conduct inference on θ_0 in a way that appropriately accounts for dependence across dyads sharing a unit in common. Section 1 outlines the population and sampling framework. Section 2 introduces a composite maximum likelihood estimator. Section 3 develops the asymptotic properties of this estimator and discusses variance estimation. Sections 4 presents a small empirical illustration.

Dyadic data, where outcomes reflecting pairwise interaction among sampled units are of primary interest, arise frequently in social science research. Such data play central roles in contemporary empirical trade and international relations research (see, respectively, Tinbergen (1962) and Oneal & Russett (1999)). They also feature in work on international financial flows (Portes & Rey, 2005), development economics (Fafchamps & Gubert, 2007), and anthropology (Apicella et al., 2012) among other fields. Despite their prominence in empirical work, the properties of extant methods of estimation and inference for dyadic regression models are not fully understood. Only recently have researchers begun to formally study these methods (e.g., Aronow et al., 2017; Menzel, 2017; Tabord-Meehan, 2018; Davezies et al., 2019). Some of the results presented in this chapter are novel, others, while having antecedents going back decades, are not widely known among empirical researchers. Section 5 ends the chapter with a discussion of further reading (including historically important

[3]Researchers also sometimes "cluster" on dyads (e.g., Santos Silva & Tenreyro, 2010); this assumes that the elements of $\{(Y_{ij}, Y_{ji})\}_{1\leq i\leq N-1, j>i}$ are conditionally independent given covariates. While this allows for dependence between, say, exports from Japan to the United States and from the United States to Japan, it does not allow for dependence between, say, exports from Japan to the United States and from Japan to Canada.

references).

1 Population and sampling framework

Let $i \in \mathbb{N}$ index agents in some (infinite) population of interest. In what follows I will refer to agents as, equivalently, nodes, vertices, units and/or individuals. Let $W_i \in \mathbb{W} = \{w_1, \ldots, w_L\}$ be an observable attribute which partitions this population into $L = |\mathbb{W}|$ subpopulations or "types"; $\mathbb{N}(w) = \{i : W_i = w\}$ equals the index set associated with the subpopulation where $W_i = w$. While L may be very large, the size of each subpopulation is assumed infinite. In practice $\mathbb{W}$ will typically enumerate different combinations of distinct agent-specific attributes (e.g., $W_i = w_1$ may correspond to former British colonies in the tropics with per capita GDP below \$3,000). Heuristically we can think of $\mathbb{W}$ as consisting of the support points of an multinomial approximation to a (possibly continuous) underlying covariate space as in Chamberlain (1987).

The indexing of agents within subpopulations homogenous in W_i is arbitrary; from the standpoint of the researcher all vertices of the same type are exchangeable. Similar exchangeability assumptions underlie most cross-sectional microeconometric procedures. For each (ordered) pair of agents – or *directed dyad* – there exists an outcome of interest $Y_{ij} \in \mathbb{Y} \subseteq \mathbb{R}$. The first subscript in Y_{ij} indexes the directed dyads *ego,* or "sending" agent, while the second its *alter,* or "receiving" agent. The *adjacency matrix* $[Y_{ij}]_{i,j\in\mathbb{N}}$ collects all such outcomes into an (infinite) random array. Within-type exchangeability of agents implies a particular form of joint exchangeability of the adjacency matrix.

To describe this exchangeability condition let $\sigma_w : \mathbb{N} \to \mathbb{N}$ be any permutation of indices satisfying the restriction

$$\left[W_{\sigma_w(i)}\right]_{i\in\mathbb{N}} = [W_i]_{i\in\mathbb{N}} \,. \tag{3}$$

Condition (3) restricts relabelings to occur among agents of the same type (i.e., *within* the index sets $\mathbb{N}(w)$, $w \in \mathbb{W}$). Following Crane & Towsner (2018) a network is *relatively exchangeable* with respect to W (or W-exchangeable) if, for all permutations σ_w,

$$\left[Y_{\sigma_w(i)\sigma_w(j)}\right]_{i,j\in\mathbb{N}} \stackrel{D}{=} [Y_{ij}]_{i,j\in\mathbb{N}} \tag{4}$$

where $\stackrel{D}{=}$ denotes equality of distribution.

If we regard $[Y_{ij}]_{i,j\in\mathbb{N}}$ as a (weighted) directed network and W_i as vertex i's "color", then (4) is equivalent to the statement that all colored graph isomorphisms are equally probable. Since there is nothing in the researcher's information set which justifies attaching different probabilities to graphs which are isomorphic (as vertex colored graphs) any probability model for the adjacency matrix should satisfy (4). If W_i encodes all the vertex information observed by the analyst, then W-exchangeability is a natural *a priori* modeling restriction.

Condition (4) allows for the invocation of very powerful de Finetti (1931) type representation results for random arrays. These results provide an "as if" (nonparametric) data generating process for the network adjacency matrix. This, in turn, facilitates various probabilistic calculations (e.g., computing expectations and variances) and gives (tractable) structure to the dependence across the elements of $[Y_{ij}]_{i,j\in\mathbb{N}}$.

Let α, $\{U_i\}_{i\geq 1}$ and $\{(V_{ij}, V_{ji})\}_{i\geq 1, j>i}$ be i.i.d. random variables. We may normalize α, U_{ij} and V_{ij} to be $\mathcal{U}[0,1]$ – uniform on the unit interval – without loss of generality. We do allow for within-dyad dependence across V_{ij} and V_{ji}; the role such dependence will become apparent below. Next consider the random array $\left[Y_{ij}^*\right]_{i,j\in\mathbb{N}}$ generated according to the rule

$$Y_{ij} \overset{def}{\equiv} \tilde{h}\left(\alpha, W_i, W_j, U_i, U_j, V_{ij}\right). \tag{5}$$

Data generating process (DGP) (5) has a number of useful features. First, any pair of outcomes, $Y_{i_1 i_2}$ and $Y_{j_1 j_2}$, sharing at least one index in common are dependent. This holds true even conditional on their types $W_{i_1}, W_{i_2}, W_{j_1}$ and W_{j_2}. Second, if $Y_{i_1 i_2}$ and $Y_{j_1 j_2}$ share exactly one index in common, say $i_1 = j_2$, then they are independent if $U_{i_1} = U_{j_2}, U_{i_2}$and U_{j_1} are additionally conditioned on. Third, if they share both indices in common, as in $i_1 = j_2$ and $i_2 = j_1$, then there may be dependence even conditional on $U_{i_1} = U_{j_2}$ and $U_{i_2} = U_{j_1}$ due to the within-dyad dependence across $V_{i_1 i_2}$ and $V_{i_2 i_1}$. These patterns of structured dependence and conditional independence will be exploited below to derive the limit distribution of parametric dyadic regression coefficient estimates. Shalizi (2016) helpful calls models like (5) conditionally independent dyad (CID) models (see also Chandrasekhar (2015)).

Crane & Towsner (2018), extending Aldous (1981) and Hoover (1979), show that, for any random array $[Y_{ij}]_{i,j\in\mathbb{N}}$ satisfying (4), there exists another array $\left[Y_{ij}^*\right]_{i,j\in\mathbb{N}}$, generated according to (5), such that

$$[Y_{ij}]_{i,j\in\mathbb{N}} \overset{D}{\equiv} \left[Y_{ij}^*\right]_{i,j\in\mathbb{N}}. \tag{6}$$

Rule (5) can therefore be regarded as a nonparametric data generating process for $[Y_{ij}]_{i,j\in\mathbb{N}}$. Equation (6) implies that we may proceed 'as if' our W-exchangeable network was generated according to (5). In the spirit of Diaconis & Janson (2008) and Bickel & Chen (2009) and others, call $\tilde{h} : [0,1] \times \mathbb{W}^2 \times [0,1]^3 \to \mathbb{R}$ a *graphon*. Here α is an unidentiable mixing parameter, analogous to the one appearing in de Finetti's (1931) classic representation result for exchangeable binary sequences. Since I will focus on inference which is conditional on the empirical distribution of the data, α can be safely ignored and I will write

$h\left(W_i, W_j, U_i, U_j, V_{ij}\right) \overset{def}{\equiv} \tilde{h}\left(\alpha, W_i, W_j, U_i, U_j, V_{ij}\right)$ in what follows (cf., Bickel & Chen, 2009; Menzel, 2017).

The Crane & Towsner (2018) representation result implies that a very particular type of dependence structure is associated with W-exchangeability. Namely, as discussed earlier, $Y_{i_1 i_2}$ and $Y_{j_1 j_2}$ are (conditionally) independent when $\{i_1, i_2\}$ and $\{j_1, j_2\}$ share no indices in common and dependent when they do. This type of dependence structure, which is very much analogous to that which arises in the theory U-Statistics, is tractable and allows for the formulation of Laws of Large Numbers and Central Limit Theorems. The next few sections will show how to use this insight to develop asymptotic distribution theory for dyadic regression.

Sampling assumption

I will regard $\left[Y_{ij}\right]_{i,j\in\mathbb{N}}$ as an infinite random (weighted) graph, G_∞, with nodes $\mathbb{N}$ and (weighted) edges given by the non-zero elements of $\left[Y_{ij}\right]_{i,j\in\mathbb{N}}$. Let $\mathcal{V} = \{1, \ldots, N\}$ be a random sample of size N from $\mathbb{N}$. Let $G_N = G_\infty\left[\mathcal{V}\right]$ be the subgraph indexed by $\mathcal{V}$. We assume that the observed network corresponds to the one induced by a random sample of agents from the larger (infinite) graph. The sampling distribution of any statistic of G_N is induced by this (perhaps hypothetical) random sampling of agents from G_∞.

If G_∞ is relatively exchangeable, then G_N will we be as well. We can thus proceed 'as if'

$$Y_{ij} = h\left(W_i, W_j, U_i, U_j, V_{ij}\right)$$

for $1 \leq i, j \leq N$. In what follows we assume that we observe W_i for each sampled agent, and for each pair of sampled agents, we observe both Y_{ij} and Y_{ji}. The presentation here rules out self loops (i.e.,$Y_{ii} \equiv 0$), however incorporating them is natural in some empirical settings and what follows can be adapted to handle them. Similarly the extension to undirected outcomes, where $Y_{ij} = Y_{ji}$, is straightforward.

2 Composite likelihood

Let $f_{Y_{12}|W_1,W_2}\left(Y_{12}|\, W_1, W_2; \theta\right)$ be a parametric family for the conditional density of Y_{12} given W_1 and W_2. This family is chosen by the researcher. Let $l_{12}\left(\theta\right)$ denote the corresponding log-likelihood. As an example to help fix ideas, return to the variant of the gravity model of trade introduced in the introduction. Following Santos Silva & Tenreyro (2006) we set

$$l_{12}\left(\theta\right) = Y_{12}R_{12}'\theta - \exp\left(R_{12}'\theta\right),$$

which equals (up to a term not varying with θ) the log likelihood of a Poisson random variable Y_{12} with mean $\exp\left(R_{12}'\theta\right)$, and choose $\hat{\theta}$ to maximize

$$L_N(\theta) = \frac{1}{N}\frac{1}{N-1}\sum_i \sum_{j\neq i} l_{ij}(\theta). \tag{7}$$

The maximizer of (7) coincides with a maximum likelihood estimate based upon the assumption that $[Y_{ij}]_{1\leq i,j\leq N, i\neq j}$ are independent Poisson random variables conditional on $\mathbf{W} = (W_1, \ldots, W_N)'$.

In practice, trade flows are unlikely to be well-described by a Poisson distribution and independence of the summands in (7) is even less likely. As discussed earlier any two summands in (7) will be dependent if they share an index in common. The likelihood contribution associated with exports from Vanuatu to Fiji is not independent of that associated with exports from Fiji to Bangladesh. Dependencies of this type mean that proceeding 'as if' (7) is a correctly specified log-likelihood (or even an M-estimation criterion function) will lead to incorrect inference.

If there exists some θ_0 such that $f_{Y_{12}|W_1,W_2}\left(Y_{12}|\, W_1, W_2; \theta_0\right)$ is the true density, then (5) corresponds to what is called a *composite* likelihood (e.g., Lindsey, 1988; Cox & Reid, 2004; Bellio & Varin, 2005). Because it does not correctly reflect the dependence structure across dyads, (5) is not a correctly specified log-likelihood function in the usual sense. If, however, the marginal density of $Y_{ij}|\, W_i, W_j$ is correctly specified, then $\hat{\theta}$ will generally be consistent for θ_0. That is we may have that

$$f_{Y_{12}|W_1,W_2}\left(Y_{12}|\, W_1, W_2\right) = f_{Y_{12}|W_1,W_2}\left(Y_{12}|\, W_1, W_2; \theta_0\right)$$

for some $\theta_0 \in \Theta$ (i.e., the marginal likelihood is correctly specified), but it *is not* the case that, setting $\mathbf{Y} = [Y_{ij}]_{1\leq i,j\leq N, i\neq j}$,

$$f_{\mathbf{Y}|\mathbf{W}}\left(\mathbf{Y}|\, \mathbf{W}\right) = \prod_{1\leq i,j\leq N, i\neq j} f_{Y_{12}|W_1,W_2}\left(Y_{ij}|\, W_i, W_j; \theta_0\right),$$

due to dependence across dyads sharing agents in common (i.e., the joint likelihood is not correctly specified). A composite log-likelihood is constructed by summing together a collection of component log-likelihoods; each such component is a log-likelihood for a portion of the sample (in this case a single *directed* dyad) but, because the joint dependence structure may not be modeled appropriately, the summation of all these components may not be the correct log likelihood for the sample as a whole.

If the marginal likelihood is itself misspecified, then (5) corresponds to what might be called a pseudo-composite-log-likelihood; "pseudo" in the sense of Gourieroux et al. (1984) and "composite" in the sense of Lindsey (1988). In what follows I outline how to conduct inference on the probability limit of $\hat{\theta}$ (denoted by θ_0 in all cases); the interpretation of this limit will, of course, depend on whether the pairwise likelihood is misspecified or not. In the context of the Santos Silva & Tenreyro (2006) gravity model example, if the true conditional mean equals $\exp\left(R_{ij}'\theta_0\right)$ for some $\theta_0 \in \Theta$, then $\hat{\theta}$ will be consistent for it (under regularity conditions). The key challenge is to characterize this estimate's sampling precision.

3 Limit distribution

To characterize the limit properties of $\hat{\theta}$ begin with a mean value expansion of the first order condition associated with the maximizer of (7). This yields, after some re-arrangement,

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right)=\left[-H_N\left(\bar{\theta}\right)\right]^{+}\sqrt{N}S_N\left(\theta_0\right)$$

with $\bar{\theta}$ a mean value between $\hat{\theta}$ and θ_0 which may vary from row to row, the + superscript denoting a Moore-Penrose inverse, and a "score" vector of

$$S_N\left(\theta\right)=\frac{1}{N}\frac{1}{N-1}\sum_{i}\sum_{j\neq i}s_{ij}\left(Z_{ij},\theta\right) \tag{8}$$

with $s\left(Z_{ij},\theta\right)=\partial l_{ij}\left(\theta\right)/\partial\theta$ for $Z_{ij}=\left(Y_{ij},W_i',W_j'\right)'$ and $H_N\left(\theta\right)=\frac{1}{N}\frac{1}{N-1}\sum_i\sum_{j\neq i}\frac{\partial^2 l_{ij}(\theta)}{\partial\theta\partial\theta'}$. In what follows I will just assume that $H_N\left(\bar{\theta}\right)\xrightarrow{p}\Gamma_0$, with Γ_0 invertible (see Graham (2017) for a formal argument in a related setting and Eagleson & Weber (1978) and Davezies et al. (2019) for more general results).

If the Hessian matrix converges in probability to Γ_0, as assumed, then

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right)=\Gamma_0^{-1}\sqrt{N}S_N\left(\theta_0\right)+o_p\left(1\right)$$

so that the asymptotic sampling properties of $\sqrt{N}\left(\hat{\theta}-\theta_0\right)$ will be driven by the behavior of $\sqrt{N}S_N\left(\theta_0\right)$. As pointed out by Fafchamps & Gubert (2007) and others, (8) is not a sum of independent random variables, hence a basic central limit theorem (CLT) cannot be (directly) applied.

My analysis of $\sqrt{N}S_N\left(\theta_0\right)$ borrows from the theory of U-Statistics (e.g., Ferguson, 2005;

van der Vaart, 2000). To make these connections clear it is convenient to re-write $S_N(\theta_0)$ as

$$S_N(\theta) = \binom{N}{2}^{-1} \sum_{i<j} \left\{ \frac{s(Z_{ij}, \theta) + s(Z_{ji}, \theta)}{2} \right\}$$

where $\sum_{i<j} \stackrel{def}{\equiv} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N}$.

Let $s_{ij} \stackrel{def}{\equiv} s(Z_{ij}, \theta_0)$, $S_N = S_N(\theta_0)$ and $\bar{s}(w, u, w', u') = \mathbb{E}\left[s_{12} \middle| W_1 = w, U_1 = u, W_2 = w', U_2 = u' \right]$; next decompose S_N as follows

$$S_N = U_N + V_N,$$

where U_N equals the projection of S_N onto $\mathbf{W} = [W_i]_{1 \leq i \leq N}$ *and* $\mathbf{U} = [U_i]_{1 \leq i \leq N}$:

$$U_N = \mathbb{E}\left[S_N \middle| \mathbf{W}, \mathbf{U} \right] = \binom{N}{2}^{-1} \sum_{i<j} \frac{\bar{s}(W_i, U_i, W_j, U_j) + \bar{s}(W_j, U_j, W_i, U_i)}{2} \qquad (9)$$

and $V_N = S_N - U_N$ is the corresponding projection error:

$$V_N = \binom{N}{2}^{-1} \sum_{i<j} \frac{[s(Z_{ij}, \theta) - \bar{s}(W_i, U_i, W_j, U_j)] + [s(Z_{ji}, \theta) - \bar{s}(W_j, U_j, W_i, U_i)]}{2}. \qquad (10)$$

Observe that U_N and V_N are uncorrelated by construction. Furthermore U_N is a U-statistic, albeit defined – partially – in terms of the latent variable U_i. Although we can not numerically evaluate U_N, we can characterize is sampling properties as $N \to \infty$. In order to do so we further decompose U_N into a Hájek projection and a second remainder term:

$$U_N = U_{1N} + U_{2N}$$

where, defining $\bar{s}_1^e(w, u) = \mathbb{E}[\bar{s}(w, u, W_1, U_1)]$ and $\bar{s}_1^a(w, u) = \mathbb{E}[\bar{s}(W_1, U_1, w, u)]$,

$$\begin{aligned} U_{1N} =& \frac{2}{N} \sum_{i=1}^{N} \frac{\bar{s}_1^e(W_i, U_i) + \bar{s}_1^a(W_i, U_i)}{2} \\ U_{2N} =& \binom{N}{2}^{-1} \sum_{i<j} \left\{ \frac{\bar{s}(W_i, U_i, W_j, U_j) + \bar{s}(W_j, U_j, W_i, U_i)}{2} \right. \\ & \left. - \frac{\bar{s}_1^e(W_i, U_i) + \bar{s}_1^a(W_i, U_i)}{2} - \frac{\bar{s}_1^e(W_j, U_j) + \bar{s}_1^a(W_j, U_j)}{2} \right\} \end{aligned}$$

The superscript in $\bar{s}_1^e(W_i, U_i)$ stands for 'ego' since $\bar{s}_1^e(W_1, U_1) = \mathbb{E}\left[\bar{s}(W_1, U_1, W_2, U_2) \middle| W_1, U_1 \right]$ corresponds to the expected value of a (generic) dyad's contribution to the composite likelihood's score vector holding its ego's attributes fixed. Similarly the superscript in $\bar{s}_1^a(W_i, U_i)$ stands for 'alter', since it is her attributes being held fixed in that average.

Putting things together yields the score decomposition

$$S_N = \overbrace{\underbrace{U_{1N}}_{\text{(Second) Hájek Projection}} + \underbrace{U_{2N}}_{\text{(Second) Projection Error}}}^{\text{(First) Projection onto } \mathbf{W} \text{ and } \mathbf{U}} + \overbrace{V_N}^{\text{(First) Projection Error}}.$$

The limit distribution of $\sqrt{N}\left(\hat{\theta} - \theta_0\right)$ depends on the joint behavior of U_{1N}, U_{2N} and V_N as $N \to \infty$. A similar type of double projection argument was utilized by Graham (2017) to characterize the limit distribution of the Tetrad Logit estimator.[4] The analyses of Menzel (2017) and Graham et al. (2019) both utilize a similar decomposition.

Variance calculation

In this section I first derive the sampling variance of $\sqrt{N}\left(\hat{\theta} - \theta_0\right)$ and then provide an interpretation of it. I begin by calculating the variance of S_N :

$$\mathbb{V}(S_N) = \mathbb{V}(U_{1N}) + \mathbb{V}(U_{2N}) + \mathbb{V}(V_N).$$

Let

$$\Sigma_q = \mathbb{C}\left(\bar{s}\left(W_{i_1}, U_{i_1}, W_{i_2}, U_{i_2}\right) + \bar{s}\left(W_{i_2}, U_{i_2}, W_{i_1}, U_{i_1}\right), \bar{s}\left(W_{j_1}, U_{j_1}, W_{j_2}, U_{j_2}\right) + \bar{s}\left(W_{j_2}, U_{j_2}, W_{j_1}, U_{j_1}\right)\right)$$

when the dyads $\{i_1, i_2\}$ and $\{j_1, j_2\}$ share $q = 0, 1, 2$ indices in common. A Hoeffding (1948) variance decomposition gives

$$\begin{aligned} \mathbb{V}(U_N) &= \mathbb{V}(U_{1N}) + \mathbb{V}(U_{2N}) \\ &\frac{4}{N}\Sigma_1 + \frac{2}{N(N-1)}(\Sigma_2 - \Sigma_1). \end{aligned}$$

Direct calculation yields (see Appendix A)

$$\begin{aligned} \Sigma_1 &\overset{def}{\equiv} \mathbb{V}\left(\frac{\bar{s}_1^e(W_1, U_1) + \bar{s}_1^a(W_1, U_1)}{2}\right) \\ &= \frac{\Omega_{12,13} + 2\Omega_{12,31} + \Omega_{21,31}}{4} \end{aligned} \tag{11}$$

[4]It is also implicit in the analysis of Bickel et al. (2011).

with

$$\Omega_{i_1 i_2, j_1 j_2} = C\left(\bar{s}\left(W_{i_1}, U_{i_1}, W_{i_2}, U_{i_2}\right), \bar{s}\left(W_{j_1}, U_{j_1}, W_{j_2}, U_{j_2}\right)\right).$$

Similarly we have

$$\begin{aligned} \Sigma_2 &= \mathbb{V}\left(\frac{\bar{s}\left(W_1, U_1, W_2, U_2\right) + \bar{s}\left(W_2, U_2, W_1, U_1\right)}{2}\right) \\ &= \frac{\Omega_{12,12} + \Omega_{12,21}}{2} \end{aligned} \tag{12}$$

and, in an abuse of notation, letting $\Sigma_3 \overset{def}{\equiv} \mathbb{V}\left(\sqrt{\binom{N}{2}} V_N\right)$,

$$\begin{aligned} \Sigma_3 &= \mathbb{E}\left[\frac{\Delta_{12,12}\left(W_1, U_1, W_2, U_2\right) + \Delta_{12,21}\left(W_1, U_1, W_2, U_2\right)}{2}\right] \\ &= \frac{\bar{\Delta}_{12,12} + \bar{\Delta}_{12,21}}{2} \end{aligned} \tag{13}$$

where

$$\begin{aligned} \Delta_{12,12}\left(W_1, U_1, W_2, U_2\right) &= \mathbb{V}\left(\left. s\left(Z_{12}, \theta\right) \right| W_1, U_1, W_2, U_2\right) \\ \Delta_{12,21}\left(W_1, U_1, W_2, U_2\right) &= \mathbb{E}\left[\left. s\left(Z_{12}, \theta\right) s\left(Z_{21}, \theta\right)' \right| W_1, U_1, W_2, U_2\right]. \end{aligned}$$

From (11), (12) and (13) we have, collecting terms, a variance of S_N equal to

$$\begin{aligned} \mathbb{V}\left(S_N\right) =& \mathbb{V}\left(U_{1N}\right) + \mathbb{V}\left(U_{2N}\right) + \mathbb{V}\left(V_N\right) \\ & \frac{4}{N}\Sigma_1 + \frac{2}{N\left(N-1\right)}\left(\Sigma_2 - 2\Sigma_1\right) + \frac{2}{N\left(N-1\right)}\Sigma_3 \\ =& \left(\Omega_{12,13} + 2\Omega_{12,31} + \Omega_{21,31}\right)\left(\frac{N-2}{N-1}\right) \\ & + \frac{1}{N-1}\left(\Omega_{12,12} + \bar{\Delta}_{12,12} + \Omega_{12,21} + \bar{\Delta}_{12,21}\right). \end{aligned} \tag{14}$$

To understand (14) note that there are exactly $\binom{N}{2}\binom{2}{1}\binom{N-2}{1} = N\left(N-1\right)\left(N-2\right)$ pairs of dyads sharing one agent in common. Consequently, applying the variance operator to S_N yields a total of $N\left(N-1\right)\left(N-2\right)$ non-zero covariance terms across the $\binom{N}{2}$ summands in S_N. It is these covariance terms which account for the leading term in (14). The second and third terms in (14) arise from the $\binom{N}{2}$ variances of the summands in S_N. Indeed, it is helpful to note that

$$\Sigma_2 = \mathbb{V}\left(\mathbb{E}\left[\left.\frac{s(Z_{12},\theta)+s(Z_{21},\theta)}{2}\right| W_1,U_1,W_2,U_2\right]\right)$$
$$\Sigma_3 = \mathbb{E}\left[\mathbb{V}\left(\left.\frac{s(Z_{12},\theta)+s(Z_{21},\theta)}{2}\right| W_1,U_1,W_2,U_2\right)\right]$$

and hence that

$$\mathbb{V}\left(\frac{s(Z_{12},\theta)+s(Z_{21},\theta)}{2}\right) = \Sigma_2+\Sigma_3. \tag{15}$$

Although it may be that $\Sigma_2+\Sigma_3 \geq \Sigma_1$ (in a positive definite sense), the larger number of non-zero covariance terms generated by applying the variance operator to S_N contributes more to its variability, than the smaller number of own variance terms. Inspecting (14) it is clear that the multiplying by $\sqrt{N}$ stabilizes the variance such that

$$\mathbb{V}\left(\sqrt{N}S_N\right) = 4\Sigma_1 + O\left(N^{-1}\right)$$

and hence

$$\mathbb{V}\left(\sqrt{N}\left(\hat{\theta}-\theta\right)\right) \rightarrow 4\left(\Gamma'\Sigma_1^{-1}\Gamma\right)^{-1}$$

as $N \rightarrow \infty$.

If a researcher uses standard software, for example a Poisson regression program, to maximize the composite log-likelihood (7) and then chooses to report robust Huber (1967) type standard errors, this corresponds to assuming that

$$\Omega_{12,13} = \Omega_{12,31} = \Omega_{21,31} = \Omega_{12,21} = \bar{\Delta}_{12,21} = 0.$$

This approach would ignore the dominant variance term and part of the higher order term as well. If, instead, the researcher clustered her standard errors on dyads, as in, for example, Santos Silva & Tenreyro (2010), then this corresponds to assuming that

$$\Omega_{12,13} = \Omega_{12,31} = \Omega_{21,31} = 0$$

but allowing $\Omega_{12,21}$ and/or $\bar{\Delta}_{12,21}$ to differ from zero. This approach would still erroneously ignore the dominant variance term. In both cases reported confidence intervals are likely to undercover the true parameter; perhaps by a substantial margin. This is shown, by example, via Monte Carlo simulation below.

Variance estimation

Graham (TBD) provides a comprehensive discussion of variance estimation for dyadic regression. One approach to variance estimation he reviews shows that Σ_1 can be estimated by the analog covariance estimate

$$\hat{\Sigma}_1 = \frac{1}{4}\frac{2}{N(N-1)(N-1)}\sum_{i=1}^{N-2}\sum_{j=i+1}^{N-1}\sum_{k=j+1}^{N}\left\{(\hat{s}_{ij}+\hat{s}_{ji})(\hat{s}_{ik}+\hat{s}_{ki})' \right.$$
$$\left. (\hat{s}_{ij}+\hat{s}_{ji})(\hat{s}_{jk}+\hat{s}_{kj})' + (\hat{s}_{ik}+\hat{s}_{ki})(\hat{s}_{jk}+\hat{s}_{kj})'\right\},$$

where the summation is over all triads in the sampled network. Each triad can itself be partitioned into three different pairs of dyads, each sharing an agent in common.

It turns out, as inspection of (15) suggests, it is easiest to estimate the sum of Σ_2 and Σ_3 jointly by

$$\widehat{\Sigma_2+\Sigma_3} = \frac{1}{4}\frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(\hat{s}_{ij}+\hat{s}_{ji})(\hat{s}_{ij}+\hat{s}_{ji})'.$$

The Jacobian matrix, Γ_0, may be estimated by $-H_N\left(\hat{\theta}\right)$, which is typically available as a by-product of estimation in most commercial software. Putting things together gives a variance estimate of

$$\hat{\mathbb{V}}\left(\sqrt{N}\left(\hat{\theta}-\theta_0\right)\right) = \hat{\Gamma}^{-1}\left(4\hat{\Sigma}_1 + \frac{2}{N-1}\left(\widehat{\Sigma_2+\Sigma_3}-2\hat{\Sigma}_1\right)\right)\left(\hat{\Gamma}^{-1}\right)'. \tag{16}$$

Graham (TBD) shows that (16) is numerically equivalent, up to a finite sample correction, to the variance estimator proposed by Fafchamps & Gubert (2007). This variance estimator includes estimates of asymptotically negligible terms. Although these terms are negligible when the sample is large enough, in practice they may be sizable in real world settings.

Limit distribution

The variance calculations outlined above imply that $\sqrt{N}S_N = \sqrt{N}U_{1N} + o_p(1)$ and hence that

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right) = \Gamma_0^{-1}\sqrt{N}U_{1N} + o_p(1).$$

Since U_{1N} is the sum of i.i.d. random variables a CLT gives

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right)\xrightarrow{D}\mathcal{N}\left(0,4\left(\Gamma_0'\Sigma_1^{-1}\Gamma_0\right)^{-1}\right), \tag{17}$$

The variance expression, equation (14), indicates that inference based upon the limit distribution (17) would ignore higher order variance terms included in (16). In practice, as has been shown in other contexts, an approach to inference which incorporates estimates of these higher order variance terms may result in inference with better size properties (e.g., Graham et al., 2014; Cattaneo et al., 2014; Graham et al., 2019). In practice I suggest using the normal reference distribution, but with a variance estimated by (16), which includes asymptotically negligible terms which may nevertheless be large in real world samples.

4 Empirical illustration

This section provides an example of a dyadic regression analysis using the dataset constructed by João Santos Silva and Silvana Tenreyro (2006) in their widely-cited paper "The Log of Gravity". This dataset, which as of the Fall of 2019 was available for download at `http://personal.lse.ac.uk/tenreyro/LGW.html`, includes information on $N = 136$ countries, corresponding to 18,360 directed trading relationships. Here I present a simple specification which includes only the log of exporter and importer GDP, respectively `lyex` and `lyim`, as well as the log distance (`ldist`) between the two trading countries. Maximizing (7) yields a fitted regression function of

$$\hat{\mathbb{E}}\left[Y_{ij}\middle|\,W_i,W_j\right]=\exp\left(\begin{array}{c}-5.688\\(1.9382)\end{array}+\begin{array}{c}0.9047\quad\texttt{lyex}\\(0.0750)\end{array}+\begin{array}{c}0.8941\quad\texttt{lyim}\\(0.0668)\end{array}+\begin{array}{c}-0.5676\quad\texttt{ldist}\\(0.0982)\end{array}\right).$$

Standard errors which cluster on dyads, but ignore dependence across dyads sharing a single agent in common, are reported in parentheses below the coefficient estimates. Specifically these standard errors coincide with square roots of the diagonal elements of

$$\frac{2}{N\left(N-1\right)}\hat{\Gamma}^{-1}\left(\widehat{\Sigma_2+\Sigma_3}\right)\left(\hat{\Gamma}^{-1}\right)'. \tag{18}$$

The coefficient estimates and reported standard errors are unremarkable in the context of the empirical trade literature. I refer the reader to Santos Silva & Tenreyro (2006) or Head & Mayer (2014) for additional context.

If, instead, the Fafchamps & Gubert (2007) dyadic robust variance-covariance estimator

Table 1: Coverage of different confidence intervals with dyadic data

	i.i.d.	dyadic clustered
θ_1	0.789	0.950
θ_2	0.520	0.942
θ_3	0.556	0.941

Notes: Actual coverage of nominal 0.95 confidence intervals. The data generating process is as described in the text. Coverage estimates are based upon 1,000 simulations. Intervals are Wald-type; constructed by taking the coefficient point estimate and adding and subtracting 1.96 times a standard error estimate. For the the "i.i.d." column this standard error is based upon the assumption of independence across dyads (see equation (18)). In the "dyadic clustered" column standard errors which account for dependence across pairs of dyads sharing an agent in common are used (see equation (16)).

is used to construct standard errors (see (16) earlier), I get

$$\hat{\mathbb{E}}\left[Y_{ij}\right| W_i, W_j] = \exp\left(\begin{array}{c} -5.688 \\ (3.6781) \end{array} + \begin{array}{c} 0.9047 \\ (0.1319) \end{array} \texttt{lyex} + \begin{array}{c} 0.8941 \\ (0.1345) \end{array} \texttt{lyim} + \begin{array}{c} -0.5676 \\ (0.2191) \end{array} \texttt{ldist}\right).$$

Standard errors which account for dependence across dyads sharing an agent in common are approximately twice those which ignore such dependence.

Monte Carlo experiment

Next I report on a small Monte Carlo experiment to illustrate the properties of inference methods based on the different variance-covariance estimates described above. I set $N = 200$ and generate outcome data for all $N(N-1)$ ordered pairs of agents according to the outcome model:

$$Y_{ij} = \exp\left(\theta_1 R_{ij} + \theta_2 W_{2i} + \theta_2 W_{2j}\right) A_i A_j U_{ij}$$

Here A_i, for $i = 1, ..., N$, is a sequence of i.i.d. log normal random variables, each with mean 1 and scale parameter σ_A; U_{ij} for $i = 1, ..., n$ with $n = N(N-1)$ is also sequence of i.i.d. log normal random variables, each with mean 1 and scale parameter σ.

Each agent is uniformly at random assigned a location on the unit square, (W_{1i}, W_{2i}), $R_{ij} = \sqrt{(W_{1i} - W_{1j})^2 + (W_{2i} - W_{2j})^2}$ equals the distance between agents i and j on that square; W_{3i} is a standard uniform random variable. I set $\theta_1 = -1$, $\theta_1 = -1/2$ and $\theta_3 = 1/2$. I set $\sigma = 1$ and $\sigma_A = 1/4$. This generates moderate, but meaningful, dependence across any two dyads sharing at least one agent in common.

Table 1 reports Monte Carlo estimates of confidence interval coverage (the nominal coverage of the intervals should be 0.95). These estimates are based upon 1,000 simulated datasets.

The coverage properties of two intervals are evaluated. The first is a Wald-based interval which uses standard errors constructed from (18). This corresponds to assuming independence across dyads or "clustering on dyads". Confidence intervals constructed in this way are routinely reported in, for example, the trade literature. The coverage of these intervals is presented in first column of Table 1. The second interval is based on the Fafchamps-Gubert variance estimate (see (16) above). The coverage of these intervals, which do take into account dependence across pairs of dyads sharing an agent in common, are reported in column two of the table.

In the experiment, the intervals which do not appropriately account fo dyadic clustering, drastically undercover the truth, whereas those based on the variance estimator outline above have actual coverage very close to 0.95. While there is no doubt additional work to be done on variance estimation and inference in the dyadic context, a preliminary suggestion is to report standard errors and confidence intervals based upon equation (16) of the previous section. These intervals perform well in the simulation experiment, while those which ignore dyadic dependence, are not recommended.

5 Further reading

Although the use of gravity models by economists dates back to Tinbergen (1962), discussions of how to account for cross dyad dependence when conducting inference have been rare. Kolaczyk (2009, Chapter 7), in his widely cited monograph on network statistics, discusses logistic regression with dyadic data. He notes that standard inference procedures are inappropriate due to the presence of dyadic dependence, but is unable to offer a solution due to the lack of formal results in the literature (available at that time).

Fafchamps & Gubert (2007) proposed a variance-covariance estimator which allows for dyadic-dependence. Their estimator coincides with the bias-corrected one discussed in Graham (TBD) and is the one recommended here. Additional versions (and analyses) of this estimator are provided by Cameron & Miller (2014) and Aronow et al. (2017). A special case of the Fafchamps & Gubert (2007) variance estimator actually appears in Holland & Leinhardt (1976) in the context of an analysis of subgraph estimation. Snijders & Borgatti (1999) suggested using the Jackknife for variance estimation of network statistics. Results in, for example, Callaert & Veraverbeke (1981) and the references therein, suggest that this estimate is (almost) numerically equivalent to $\hat{\Sigma}_1$ defined above.

Aldous' (1981) representation result evidently inspired some work on LLNs and CLTs for so called dissociated random variables and exchangeable random arrays (e.g., Eagleson & Weber, 1978). The influence of this work on empirical practice appears to have been minimal. Bickel et al. (2011), evidently inspired by the variance calculations of Picard et al. (2008), but perhaps more accurately picking up where Holland & Leinhardt (1976) stopped (albeit inadvertently), present asymptotic normality results for subgraph counts. Network density, which corresponds to the mean $[N(N-1)]^{-1}\sum_{i\neq j}Y_{ij}$ when Y_{ij} is binary, is the simplest example they consider and also prototypical for understanding regression. The limit theory sketched hear was novel at the time of drafting, but substantially related results – independently derived – appear in Menzel (2017) and Davezies et al. (2019). Both of these papers also present bootstrap procedures appropriate for network data. The Menzel (2017) paper focuses on the important problem of graphon degeneracy. This occurs when the graphon only weakly varies in U_i and U_j; degeneracy effects rates of convergence and limit distributions. Graham et al. (2019) present results on kernel density estimation with dyadic data. Tabord-Meehan (2018) showed asymptotic normality of dyadic linear regression coefficients using a rather different approach.

A Derivations

Expression (11) of the main text is an implication of calculations like

$$\begin{aligned}\mathbb{V}\left(\bar{s}_1^e\left(W_1,U_1\right)\right) &= \mathbb{E}\left[\mathbb{E}\left[\left.\bar{s}\left(W_1,U_1,W_2,U_2\right)\right|W_1,U_1\right]\mathbb{E}\left[\left.\bar{s}\left(W_1,U_1,W_2,U_2\right)\right|W_1,U_1\right]'\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\left.\bar{s}\left(W_1,U_1,W_2,U_2\right)\right|W_1,U_1\right]\mathbb{E}\left[\left.\bar{s}\left(W_1,U_1,W_3,U_3\right)\right|W_1,U_1\right]'\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\left.\bar{s}\left(W_1,U_1,W_2,U_2\right)\bar{s}\left(W_1,U_1,W_3,U_3\right)'\right|W_1,U_1\right]\right]\\ &= \mathbb{E}\left[\bar{s}\left(W_1,U_1,W_2,U_2\right)\bar{s}\left(W_1,U_1,W_3,U_3\right)'\right]\\ &= \Omega_{12,13}.\end{aligned}$$

The second equality immediately above follows because $W_2,U_2|\,W_1,U_1 \overset{D}{=} W_3,U_3|\,W_1,U_1 \overset{D}{=} W_2,U_2$, the third by independence of $\bar{s}\left(W_1,U_1,W_2,U_2\right)$ and $\bar{s}\left(W_1,U_1,W_3,U_3\right)$ *conditional* on W_1,U_1, and the fourth by iterated expectations.

References

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

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature*, 481(7382), 497 – 501.

Aronow, P. M., Samii, C., & Assenova, V. A. (2017). Cluster-robust variance estimation for dyadic data. *Political Analysis*, 23(4), 564 – 577.

Baldwin, R. & Taglioni, D. (2007). Trade effects of the euro: a comparison of estimators. *Journal of Economic Integration*, 22(4), 780 – 818.

Bellio, R. & Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling*, 5(3), 217 – 227.

Bickel, P. J. & 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.

Bickel, P. J., Chen, A., & Levina, E. (2011). The method of moments and degree distributions for network models. *Annals of Statistics*, 39(5), 2280 – 2301.

Callaert, H. & Veraverbeke, N. (1981). The order of the normal approximation for a studentized u-statistic. *Annals of Statistics*, 9(1), 194 – 200.

Cameron, A. C. & Miller, D. L. (2014). *Robust inference for dyadic data*. Technical report, University of California - Davis.

Cattaneo, M., Crump, R., & Jansson, M. (2014). Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory*, 30(1), 176 – 200.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3), 305 – 334.

Chandrasekhar, A. (2015). Econometrics of network formation. In Y. Bramoullé, A. Galeotti, & B. Rogers (Eds.), *Oxford Handbook on the Economics of Networks*. Oxford University Press.

Cox, D. R. & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3), 729 – 737.

Crane, H. & Towsner, H. (2018). Relatively exchangeable structures. *Journal of Symbolic Logic*, 83(2), 416 – 442.

Davezies, L., d'Haultfoeuille, X., & Guyonvarch, Y. (2019). *Empirical process results for exchangeable arrayes*. Technical report, CREST-ENSAE.

de Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale*, 4, 251 – 299.

Diaconis, P. & Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica*, 28(1), 33 – 61.

Eagleson, G. K. & Weber, N. C. (1978). Limit theorems for weakly exchangeable arrays. *Mathematical Proceedings of the Cambridge Philosophical Society*, 84(1), 123 – 130.

Fafchamps, M. & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2), 326 – 350.

Ferguson, T. S. (2005). U-statistics. University of California - Los Angeles.

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: applications to poisson models. *Econometrica*, 52(3), 701 – 720.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4), 1033 – 1063.

Graham, B. S. (TBD). *Handbook of Econometrics*, volume 7, chapter The econometric analysis of networks. North-Holland: Amsterdam.

Graham, B. S., Imbens, G. W., & Ridder, G. (2014). Complementarity and aggregate implications of assortative matching: a nonparametric analysis. *Quantitative Economics*, 5(1), 29 – 66.

Graham, B. S., Niu, F., & Powell, J. L. (2019). *Kernel density estimation for undirected dyadic data*. Technical report, University of California - Berkeley.

Head, K. & Mayer, T. (2014). *Handbook of International Economics*, volume 4, chapter Gravity equations: workhorse, toolkit, and cookbook, (pp. 131 – 191). North-Holland: Amsterdam.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3), 293 – 325.

Holland, P. W. & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7, 1 – 45.

Hoover, D. N. (1979). *Relations on probability spaces and arrays of random variables*. Technical report, Institute for Advanced Study, Princeton, NJ.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221 – 233.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data*. New York: Springer.

Lindsey, B. G. (1988). Composite likelihood. *Contemporary Mathematics*, 80, 221 – 239.

Menzel, K. (2017). *Bootstrap with clustering in two or more dimensions*. Technical Report 1703.03043v2, arXiv.

Oneal, J. R. & Russett, B. (1999). The kantian peace: the pacific benefits of democracy, interdependence, and international organizations. *World Politics*, 52(1), 1 – 37.

Picard, F., Daudin, J. J., Koskas, M., Schbath, S., & Robin, S. (2008). Assessing the exceptionality of network motifs. *Journal of Computational Biology*, 15(1), 1 – 20.

Portes, R. & Rey, H. (2005). The determinants of cross-border equity flows. *Journal of International Economics*, 65(2), 269 – 296.

Rose, A. K. (2004). Do we really know that the wto increases trade? *American Economic Review*, 94(1), 98 – 114.

Santos Silva, J. & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4), 641 – 658.

Santos Silva, J. & Tenreyro, S. (2010). Currency unions in prospect and retrospect. *Annual Review of Economics*, 2, 51 – 74.

Shalizi, C. R. (2016). *Lecture 1: Conditionally-independent dyad models*. Lecture note, Carnegie Mellon University.

Snijders, T. A. B. & Borgatti, S. P. (1999). Non-parametric standard errors and tests for network statistics. *Connections*, 22(2), 61 – 70.

Tabord-Meehan, M. (2018). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business and Economic Statistics*.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge University Press.