Dyadic Regression

Econometric Methods for Networks,

University of Warwick, May 30th to June 1st, 2018

Bryan S. Graham

University of California - Berkeley

Dyadic regression analyses are abundant in social science research (see below).

In economics they date (at least) to Tinbergen's (1962) pioneering analysis of trade flows.

While frequently used by empirical researchers, dyadic regression analysis lacks inferential foundations.

Widely varying approaches to hypothesis testing used in practice.

Tinbergen (1962, SWE, Table VI-1)

FACTORS DETERMINING THE SIZE OF INTERNATIONAL TRADE FLOWS

Results of Calculations A (18 countries)

$$\log E_{ij} = \alpha_1 \log Y_i + \alpha_2 \log Y_j + \alpha_3 \log D_{ij} + \alpha_4 \log N + \alpha_5 \log P_C + \alpha_6 \log P_B + \alpha'_0$$

Calculation No.	ESTIMATED VALUE OF THE COEFFICIENTS							Correlation Coefficient
	α_1	α_2	α_3	α_4	α_5	α_6	α'_0	
A-1	0.7338 (0.0438)	0.6238 (0.0438)	–0.5981 (0.0405)	—	—	—	–0.3783	0.8248
A-2	0.7907 (0.0497)	0.6766 (0.0496)	–0.6252 (0.0460)	—	—	—	–0.4013	0.8084
A-3	0.7357 (0.0421)	0.6183 (0.0422)	–0.5570 (0.0473)	0.0191 (0.0082)	0.0496 (0.0111)	0.0406 (0.0272)	–0.4451	0.8437

E_{ij} Exports from country i to country j
Y_i GNP of exporting country
Y_j GNP of importing country
D_{ij} Distance between countries i and j
N Dummy variable for neighbor countries
P_C Dummy variable for Commonwealth preference
P_B Dummy variable for Benelux preference

In A-2 the trade amount is measured in the importing country.
Figures in brackets are standard deviations.

Year: 1958, N = 18, N(N-1) = 306 (estimation by OLS)

Tinbergen (1962, SWE, Table VI-4)

Results of Calculations B (42 countries)

$$\log E_{ij} = \alpha_1 \log Y_i + \alpha_2 \log Y_j + \alpha_3 \log D_{ij} + \alpha_4 \log N + \alpha_7 \log P + \alpha'_0$$

Calculation No.	ESTIMATED VALUE OF THE COEFFICIENTS α_1	α_2	α_3	α_4	α_7	α'_0	Correlation Coefficient
B-1	1.0240 (0.0270)	0.9395 (0.0269)	–0.8919 (0.0455)	—	—	–0.6627 (0.6802)	0.8094
B-2	1.0250 (0.0269)	0.9403 (0.0269)	–0.8225 (0.0517)	0.2581 (0.0920)	—	–0.7188 (0.6789)	0.8104
B-3	1.1832 (0.0323)	1.0752 (0.0323)	–0.9325 (0.0584)	0.2217 (0.1037)	—	–1.0296 (0.7645)	0.7987
B-4	0.9965 (0.0267)	0.9116 (0.0267)	–0.7803 (0.0511)	0.2434 (0.0903)	0.4703 (0.0588)	–0.7798 (0.6668)	0.8180
B-5	1.1567 (0.0319)	1.0486 (0.0319)	–0.9165 (0.0574)	0.2367 (0.1018)	0.8926 (0.1100)	–1.0641 (0.7505)	0.8070

E_{ij} Exports from country i to country j
Y_i GNP of exporting country
Y_j GNP of importing country
(Nominal in B-1, B-2 and B-4; real in B-3 and B-5.)
D_{ij} Distance between countries i and j
N Dummy variable for neighboring countries
P Dummy variable for preference

Because of difference in treatment of preferential relations, the coefficients are not comparable between B-4 and B-5.
Figures in brackets are standard deviations.

Year: 1959, N = 42, N(N-1) = 1,722 (estimation by OLS)

Rose (2004, AER)

TABLE 1—BENCHMARK RESULTS

	Default	No industrial countries	Post 1970	With country effects
Both in GATT/WTO	−0.04	−0.21	−0.08	0.15
	(0.05)	(0.07)	(0.07)	(0.05)
One in GATT/WTO	−0.06	−0.20	−0.09	0.05
	(0.05)	(0.06)	(0.07)	(0.04)
GSP	0.86	0.04	0.84	0.70
	(0.03)	(0.10)	(0.03)	(0.03)
Log distance	−1.12	−1.23	−1.22	−1.31
	(0.02)	(0.03)	(0.02)	(0.02)
Log product real GDP	0.92	0.96	0.95	0.16
	(0.01)	(0.02)	(0.01)	(0.05)
Log product real GDP p/c	0.32	0.20	0.32	0.54
	(0.01)	(0.02)	(0.02)	(0.05)
Regional FTA	1.20	1.50	1.10	0.94
	(0.11)	(0.15)	(0.12)	(0.13)
Currency union	1.12	1.00	1.23	1.19
	(0.12)	(0.15)	(0.15)	(0.12)
Common language	0.31	0.10	0.35	0.27
	(0.04)	(0.06)	(0.04)	(0.04)
Land border	0.53	0.72	0.69	0.28
	(0.11)	(0.12)	(0.12)	(0.11)
Number landlocked	−0.27	−0.28	−0.31	−1.54
	(0.03)	(0.05)	(0.03)	(0.32)
Number islands	0.04	−0.14	0.03	−0.87
	(0.04)	(0.06)	(0.04)	(0.19)
Log product land area	−0.10	−0.17	−0.10	0.38
	(0.01)	(0.01)	(0.01)	(0.03)
Common colonizer	0.58	0.73	0.52	0.60
	(0.07)	(0.07)	(0.07)	(0.06)
Currently colonized	1.08	—	1.12	0.72
	(0.23)		(0.41)	(0.26)
Ever colony	1.16	−0.42	1.28	1.27
	(0.12)	(0.57)	(0.12)	(0.11)
Common country	−0.02	—	−0.32	0.31
	(1.08)		(1.04)	(0.58)
Observations	234,597	114,615	183,328	234,597
R^2	0.65	0.47	0.65	0.70
RMSE	1.98	2.36	2.10	1.82

Notes: Regressand: log real trade. OLS with year effects (intercepts not reported). Robust standard errors (clustering by country-pairs) are in parentheses.

Apicella, Marlowe, Fowler & Christakis (2011, Nature)

RESEARCH SUPPLEMENTARY INFORMATION

Supplementary Table S16: GEE Regression of Social Ties on Public Good Donations

	Dependent Variable: Ego Wants to Camp with Alter			***Dependent Variable: Ego Gives Gift to Alter***		
	Coef.	*S.E.*	*p*	*Coef.*	*S.E.*	*p*
Ego Public Good Donation	0.003	0.031	0.930	-0.022	0.044	0.627
Alter Public Good Donation	-0.026	0.044	0.550	-0.100	0.047	0.035
Ego-Alter Similarity in Public Good Donation	0.250	0.051	0.000	0.174	0.044	0.000
Residual		5879			2096	
Null Residual		5923			2113	
N		18054			2310	

GEE logit regression of presence of social tie from ego to alter on ego and alter attributes, clustering standard errors on each ego.

Fafchamps and Gubert (2007, AERPP)

TABLE 1—LINKS AND INCOME CORRELATION

	Coefficient estimate	Dyadic t-value
Income correlation		
Correlation of i and j's incomes[a]	1.083	1.44
Geographic proximity		
Same sitio = 1[b]	2.647	8.84
Difference in distance to road if same sitio	−0.121	−3.90
Difference in:		
Dummy = 1 if primary occupation of head is farming	0.028	0.23
Number of working members × number of activities	0.003	0.06
Age of household head	−0.010	−2.52
Health index 1–4 (1 = good health, 4 = disabled)	0.027	0.46
Years of education of household head	−0.010	−0.59
Total wealth[a]	−0.113	−2.37
Village dummies	Included but not shown	
Intercept	−5.995	−15.41
Number of observations	10,264	

Notes: The dependent variable = 1 if i cites j as the source of mutual insurance, 0 otherwise. Estimator is logit. All t-values based on standard errors corrected for dyadic correlation of errors.

[a] Instrumented variables—see text for details.

[b] Small cluster of 15–20 households.

How to Conduct Inference?

Dyads present an ironic situation in that dyadic data sets, with 100,000 cases (or often considerably more), may seem ideal for hypothesis testing. Yet, the structure of dyadic data complicates the assessment to statistical significance. Because dyadic observations are not independent events, the usual tests of significance result in overconfidence, even when the model itself appears to be correctly specified (Erikson, Pinto & Rader, 2014, p. 457).

How to Conduct Inference? (continued)

Dyadic observations are not independent. This is due to the presence of individual-specific factors common to all observations involving that individual. It is thus reasonable to assume that $\mathbb{E}\left[u_{ij}u_{ik}\right] \neq 0$ for all k and $\mathbb{E}\left[u_{ij}u_{kj}\right] \neq 0$ for all k. By the same reasoning, we also have $\mathbb{E}\left[u_{ij}u_{jk}\right] \neq 0$ and $\mathbb{E}\left[u_{ij}u_{ki}\right] \neq 0$. Provided that regressors are exogenous,...OLS...yields consistent coefficient estimates but standard errors are inconsistent, leading to incorrect inference (Fafchamps and Gubert, 2007, p. 330).

Existing suggestions

1. Permutation approaches: quadratic assignment procedure (QAP) of Hubert (1985, PM), Krackhardt (1988, SN)

2. Integrated likelihood/MCMC: p_2 model of van Duijn, Snijders and Zijlstra (2004, SN), Zijlstra, van Duijn and Snijders (2009, BJMSP), Krivitsky, Handcock, Raftery and Hoff (2009, SN)

Existing suggestions (continued)

3. Pairwise/composite likelihood: Bellio and Varin (2005, SM)

4. Dyadic cluster-robust s.e.: Fafchamps and Gubert (2007, JDE), Cameron and Miller (2014, WP), Aronow, Samii and Assenova (2015, PA), Tabord-Meehan (2017, WP)

Dyadic Regression: Notation & Setup

Let $Y_{ij} = Y_{ji}$ be an *undirected* outcome of interest associated with dyad $\{i, j\}$ (directed case poses few additional challenges).

- will focus on binary case with $Y_{ij} = D_{ij} \in \{0, 1\}$

Let X_i be a vector of agent-level covariates.

Let U_i be unobserved agent-level heterogeneity.

Dyadic Regression: Notation & Setup (continued)

The dyadic regression function (symmetric in its two arguments) is

$$g\left(x, x'\right) = \mathbb{E}\left[Y_{ij} \middle| X_i = x, X_j = x'\right]$$

Here i and j denote two independent random draws from the population of interest.

Dyadic Regression: Nonparametric DGP

We will assume that

$$D_{ij} \Big| X_i, X_j, U_i, U_j \sim \text{Bernoulli}\left(h\left(X_i, X_j, U_i, U_j\right)\right)$$

for some function $h(\cdot)$, symmetric in its first and second, as well as its third and fourth, arguments.

May be possible to motivate this DGP via exchangeability arguments (e.g., Aldous-Hoover Theorem); cf., Menzel (2018).

Iterated expectations gives

$$g\left(x, x'\right) = \int \int h\left(x, x', u, v\right) f_{U|X}\left(u \middle| x\right) f_{U|X}\left(v \middle| x'\right) \mathrm{d}u \mathrm{d}v.$$

Dyadic Regression: Nonparametric DGP (continued)

Elements of $\mathbf{D} = \left[D_{ij}\right]$ are conditionally independent given $\mathbf{X}$ *and* the latent $\mathbf{U}$, but may be dependent conditional on $\mathbf{X}$ alone.

Captures types of dependence structures typically assumed in empirical work (e.g., Frank and Strauss, 1986, JASA; Fafchamps and Gubert, 2007, JDE).

Will defer question of whether $g\left(x, x'\right)$ has a structural interpretation until later.

Dyadic Regression: Parametric estimation

A prototypical specification for a binary outcome is $\pi\left(R_{ij}'\theta_0\right) = g\left(X_i, X_j\right)$ where, for $\theta = (\alpha, \beta', \gamma')'$,

$$\text{logit}\left[\pi\left(R_{ij}'\theta\right)\right] = \alpha + \left[t\left(X_i\right) + t\left(X_j\right)\right]'\beta + \omega\left(X_i, X_j\right)'\gamma$$

1. $t(X)$ a vector of linear independent and known functions of X;

2. $\omega\left(X_i, X_j\right) = \omega\left(X_j, X_i\right)$ dyadic-specific regressors;

3. $R_{ij} = r\left(X_i, X_j\right) = \left(1, \left(t\left(X_i\right) + t\left(X_j\right)\right)', \omega\left(X_i, X_j\right)'\right)'$.

Dyadic Regression: Parametric estimation (continued)

Estimate θ_0 by maximizing the Bernoulli pseudo-likelihood function

$$L_N(\theta) = \binom{N}{2}^{-1} \sum_{i<j} l\left(Z_{ij};\theta\right)$$

with $Z_{ij} = \left(X_i', X_j', D_{ij}\right)'$ and $l\left(Z_{ij};\theta\right)$ equal to

$$l\left(Z_{ij};\theta\right) = D_{ij} \ln\left[\pi\left(R_{ij}'\theta\right)\right] + \left(1 - D_{ij}\right) \ln\left[1 - \pi\left(R_{ij}'\theta\right)\right].$$

This can be done using standard software (see examples above).

Dyadic Regression: Parametric estimation (continued)

Under some basic conditions

$$\sqrt{N}\left(\hat{\theta}_{\mathsf{DR}} - \theta_0\right) = \underbrace{\left[-H_N\left(\bar{\theta}\right)\right]^{+}}_{\text{Inverse Hessian}} \times \sqrt{N} S_N\left(\theta_0\right)$$

where

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

for $s\left(Z_{ij}; \theta\right) = \frac{\partial l(Z_{ij};\theta)}{\partial \theta}$ and $H_N\left(\theta\right) = \binom{N}{2}^{-1} \sum_{i<j} \frac{\partial^2 l(Z_{ij};\theta)}{\partial \theta \partial \theta'}$.

Dyadic Regression: Parametric estimation (continued)

$S_N(\theta)$ is not the sum of independent components.

...also not a U-Statistic (D_{ij} is a dyad-level random variable), but it is "U-Statistic like".

A Hoeffding (1948) variance decomposition gives

$$\mathbb{V}\left(\sqrt{N}S_N(\theta_0)\right) = 4\Sigma_1 + \frac{2}{N-1}(\Sigma_2 - 2\Sigma_1)$$

where $\Sigma_p = \mathbb{E}\left[s\left(Z_{i_1 i_2};\theta_0\right)s\left(Z_{j_1 j_2};\theta_0\right)'\right]$ when the dyads $\{i_1, i_2\}$ and $\{j_1, j_2\}$ share $p = 0, 1, 2$ agents in common.

Dyadic Regression: Variance estimation

Fafchamps and Gubert (2007, JDE) propose a now widely-used dyadic-clustered covariance estimator (cf., Cameron and Miller, 2014, WP; Aronow et al., 2017, PA).

It turns out their estimator is equivalent to a natural analog estimate of $4\Sigma_1 + \frac{2}{N-1}(\Sigma_2 - 2\Sigma_1)$.

Showing this involves tedious counting arguments.

Dyadic Regression: Variance estimation (continued)

The standard "econometrician's estimate" focuses on the leading term only:

$$\tilde{\Sigma}_1 = \frac{1}{N} \sum_{i=1}^{N} \hat{\bar{s}}_i(\theta) \, \hat{\bar{s}}_i(\theta)'$$

with $\hat{\bar{s}}_i(\theta) = \frac{1}{N-1} \sum_{j \neq i} s\left(Z_{ij}; \theta\right)$.

This "Jackknife" estimate is biased (e.g., Efron and Stein, 1979, AS).

It turns out that the Fafchamps and Gubert (2007, JDE) estimate is "bias-corrected" (albeit computationally inefficient).

When network is sparse these differences appear to be important.

Dyadic Regression: Asymptotic Normality

Let $k\left(X_i, X_j; \theta_0\right) = \frac{\pi_1\left(R_{ij}'\theta_0\right)R_{ij}}{\pi\left(R_{ij}'\theta_0\right)\left[1-\pi\left(R_{ij}'\theta_0\right)\right]}$ for $\pi_1(v) = \partial\pi(v)/\partial v$ and consider the following decomposition of S_N:

$$\begin{aligned}
\sqrt{N}S_N &= \sqrt{N}\binom{N}{2}^{-1}\sum_{i<j}\left\{D_{ij} - \pi\left(R_{ij}'\theta_0\right)\right\}k\left(X_i, X_j; \theta_0\right) \\
&= \underbrace{\sqrt{N}\binom{N}{2}^{-1}\sum_{i<j}\left\{h\left(X_i, X_j, U_i, U_j\right) - \pi\left(R_{ij}'\theta_0\right)\right\}k\left(X_i, X_j; \theta_0\right)}_{V_N} \\
&+ \underbrace{\sqrt{N}\binom{N}{2}^{-1}\sum_{i<j}\left\{D_{ij} - h\left(X_i, X_j, U_i, U_j\right)\right\}k\left(X_i, X_j; \theta_0\right)}_{T_N} \\
&= \sqrt{N}V_N + \sqrt{N}T_N. \qquad (1)
\end{aligned}$$

Dyadic Regression: Asymptotic Normality

Begin with the second term, T_N, which equals the sum of $\binom{N}{2}$ conditionally independent random variables, each with variance

$$\begin{aligned}\Omega_3 = & \mathbb{E}\left[h\left(X_1, X_2, U_1, U_2\right)\left[1 - h\left(X_1, X_2, U_1, U_2\right)\right]\right. \\ & \left.\times k\left(X_1, X_2; \theta_0\right) k\left(X_1, X_2; \theta_0\right)'\right].\end{aligned}$$

Theorem 8 of Rao (2009) then gives $\binom{N}{2}^{1/2} T_N \xrightarrow{D} N\left(0, \Omega_3\right)$.

Dyadic Regression: Asymptotic Normality (continued)

V_N is a 2^{nd} order U-Statistic with kernel

$$v(x, y, u, v) = \left\{h(x, y, u, v) - \pi\left(r(x, y)'\theta_0\right)\right\} k(x, y; \theta_0)$$

Decompose V_N as $V_N = V_{1N} + V_{2N}$ with

$$V_{1N} = \frac{2}{N}\sum_{i=1}^{N} v_1(X_i, U_i)$$

$$V_{2N} = \binom{N}{2}^{-1}\sum_{i<j}\left\{v\left(X_i, X_j, U_i, U_j\right) - v_1(X_i, U_i) - v_1\left(X_j, U_j\right)\right\}$$

where $v_1(x, u) = \mathbb{E}\left[\left\{h(x, X_1, u, U_1) - \pi\left(r(x, X_1)'\theta_0\right)\right\} k(x, X_1; \theta_0)\right]$.

Note: $\mathbb{C}(T_N, V_{1N}) = \mathbb{C}(T_N, V_{2N}) = \mathbb{C}(V_{1N}, V_{2N}) = 0$.

Dyadic Regression: Asymptotic Normality (continued)

A variance calculation then gives

$$\mathbb{V}\left(\sqrt{N}\left(\begin{array}{c} V_{1N} \\ \left[\frac{N-1}{2}\right]^{1/2} V_{2N} \end{array}\right)\right) = \left(\begin{array}{cc} 4\Omega_1 & 0 \\ 0 & \Omega_2 - 2\Omega_1 \end{array}\right)$$

where

$$\Omega_q = \mathbb{E}\left[v\left(X_{i_1}, X_{i_2}, U_{i_1}, U_{i_2}\right) v\left(X_{j_1}, X_{j_2}, U_{j_1}, U_{j_2}\right)'\right]$$

when the dyads $\{i_1, i_2\}$ and $\{j_1, j_2\}$ share $q = 0, 1, 2$ agents.

Direct calculation reveals that $\Omega_1 = \Sigma_1$, as defined earlier, and also that $\Sigma_2 = \Omega_2 + \Omega_3$.

Dyadic Regression: Asymptotic Normality (continued)

Putting each of the above pieces together suggests that

$$\mathbb{V}\left(\sqrt{N}\begin{pmatrix} V_{1N} \\ \left[\frac{N-1}{2}\right]^{1/2} V_{2N} \\ \left[\frac{N-1}{2}\right]^{1/2} T_N \end{pmatrix}\right) = \begin{pmatrix} 4\Omega_1 & 0 & 0 \\ 0 & \Omega_2 - 2\Omega_1 & 0 \\ 0 & 0 & \Omega_3 \end{pmatrix}. \qquad (2)$$

Assume: support of $R_{ij} = r\left(X_i, X_j\right)$ is a compact subset of $\mathbb{R}^{\dim(\theta)}$ and $\theta \in \Theta \subset \mathbb{R}^{\dim(\theta)}$ with $\theta_0 \in \mathrm{int}(\Theta)$.

Under these regularity conditions the probability that any particular dyad links will be bounded away from both zero and one.

Dyadic Regression: Asymptotic Normality (continued)

The network will be dense in the limit.

The sampling properties of $\sqrt{N}\left(\hat{\theta}_{\mathrm{DR}}-\theta_0\right)$ will be driven by those of $\sqrt{N}V_{1N}$. A projection argument, CLT and Slutsky Theorem give in this case

$$\sqrt{N}\left(\hat{\theta}_{\mathrm{DR}}-\theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, 4\Gamma_0^{-1}\Omega_1\Gamma_0^{-1}\right). \tag{3}$$

Dyadic Regression: Asymptotic Normality (continued)

Suppose we wish to construct a 95 percent confidence interval for (a component of) θ_0.

I suggest using an approximate variance of

$$\mathbb{V}_0 = \frac{1}{N}\Gamma_0^{-1}\left[4\Sigma_1 + \frac{2}{N-1}\left(\Sigma_2 - 2\Sigma_1\right)\right]\Gamma_0^{-1}$$

(i.e., keep asymptotically negligible terms).

Replacing Γ_0, Σ_1 and Σ_2 with plug-in estimates $\hat{\Gamma}$, $\hat{\Sigma}_1$, and $\hat{\Sigma}_2$ yields a feasible covariance estimate, $\hat{\mathbb{V}}$, with associated confidence interval $\hat{\theta}_{\mathrm{k,DR}} \pm 1.96\sqrt{\hat{\mathbb{V}}_{k,k}}$.

Dyadic Regression

Applying some basic ideas/tools on exchangeable random graphs, network moments etc...

...puts dyadic regression on a much sounder inferential basis.

Potential to make a large empirical literature much more coherent.

It turns out that (one) emerging practice in economics has a coherent foundation.

Average Partial Effects

Do trade agreements increase trade (e.g., Tinbergen, 1962; Rose, 2004, AER)?

1. draw agent i at random and exogenously assign her covariate value $X_i = x$

2. draw a second independent agent j at random and assign her covariate value $X_j = x'$.

The (ex ante) expected outcome associated with these assignments is

$$m^{\mathsf{ASF}}\left(x, x'\right) = \int h\left(x, x', u, v\right) f_U\left(u\right) f_U\left(v\right) \mathsf{d}u\mathsf{d}v$$

Average Partial Effects (continued)

If $X_i \in \{0,1\}$ is a binary indicator for GATT/WHO membership as in Rose (2004), then the contrast

$$m^{\mathsf{ASF}}(1,1) - m^{\mathsf{ASF}}(0,0)$$

gives differences in the probability of trade between a random pair of countries in the GATT/WHO vs. non-GATT/WHO states of the world.

The dyadic setting also raises new questions. For example the double difference

$$m^{\mathsf{ASF}}(1,1) - m^{\mathsf{ASF}}(0,1) - \left[m^{\mathsf{ASF}}(1,0) - m^{\mathsf{ASF}}(0,0)\right]$$

measures complementarity in a binary policy/treatment across the two agents in the dyad.

Average Partial Effects: Identification

A simple identification result under "selection on observations" type assumptions follows if there is a proxy W_i for U_i such that:

1. [*redundancy*] $\mathbb{E}\left[D_{ij}\middle| X_i, X_j, U_i, U_j, W_i, W_j\right] = h\left(X_i, X_j, U_i, U_j\right)$;

2. [*conditional independence*] $U_i \perp X_i | \, W_i = w, \; w \in \mathbb{W}$;

3. [*support*] a support condition holds.

Dyadic proxy variable regression

Define the dyadic proxy variable regression (PVR) function as

$$q\left(x, x', w, w'\right) = \mathbb{E}\left[D_{ij} \middle| X_i = x, X_j = x', W_i = w, W_j = w'\right]$$

Under the first two conditions (*and random sampling*)

$$\begin{aligned} q\left(X_i, X_j, W_i, W_j\right) &= \mathbb{E}\left[\mathbb{E}\left[D_{ij} \middle| X_i, X_j, U_i, U_j, W_i, W_j\right] \middle| X_i, X_j, W_i, W_j\right] \\ &= \mathbb{E}\left[h\left(X_i, X_j, U_i, U_j\right) \middle| X_i, X_j, W_i, W_j\right] \\ &= \int h\left(X_i, X_j, u, v\right) f_{U|W}\left(u \middle| W_i\right) f_{U|W}\left(v \middle| W_j\right) \mathrm{d}u\mathrm{d}v \end{aligned}$$

Double marginal integration

Putting things together we have

$$
\begin{aligned}
\mathbb{E}_{W_i}\left[\mathbb{E}_{W_j}\left[q\left(x,x',W_i,W_j\right)\right]\right] =& \int \left[\int h\left(x,x',u,v\right)\right.\\
&\left.\times f_{U|W}\left(u|\, w\right) f_{U|W}\left(v|\, w'\right) \mathsf{d}u\mathsf{d}v\right]\\
&\times\, f_W\left(w\right) f_W\left(w'\right) \mathsf{d}w\mathsf{d}w'\\
=& \int h\left(x,x',u,v\right) f_U\left(u\right) f_U\left(v\right) \mathsf{d}u\mathsf{d}v\\
=& m^{\mathsf{ASF}}\left(x,x'\right).
\end{aligned}
$$

Support Condition

Since $q(x, x', w, w')$ is only identified at those points where

$$f_{W|X}(w|\, x)\, f_{W|X}\left(w'\middle|\, x'\right) > 0$$

while the integral

$$m^{\mathsf{ASF}}\left(x, x'\right) = \int\int q\left(x, x', w, w'\right) f_W(w)\, f_W\left(w'\right) \mathrm{d}w \mathrm{d}w'$$

is over $\mathbb{W} \times \mathbb{W}$ (need support condition!).

Support Condition (continued)

The needed condition is:

$$\mathbb{S}\left(x, x'\right) \overset{def}{\equiv} \left\{w, w' : f_{W|X}\left(w|\, x\right) f_{W|X}\left(w'\middle|\, x'\right) > 0\right\} = \mathbb{W} \times \mathbb{W}.$$

When X_i is discretely-valued we can express the support conditioning in a form similar to the overlap condition from program evaluation:

$$p_x\left(w\right) p_{x'}\left(w'\right) \geq \kappa > 0 \text{ for all } \left(w, w'\right) \in \mathbb{W} \times \mathbb{W}$$

where $p_x\left(w\right) \overset{def}{\equiv} \Pr\left(X_i = x|\, W_i = w\right).$

Identification Wrap-up

Estimation of, and inference on, the ASF are straightforward when the proxy variable regression function is "flexible parametric".

Provides a framework for thinking about causal effects in dyadic settings (both experimental and observational).

When $X \in \{0, 1\}$ there are interesting connections to the program evaluation literature.

Semiparametric efficiency bound...

Estimation

Identification result suggests the analog estimator

$$\widehat{m}^{\mathsf{ASF}}\left(x,x'\right)=\binom{N}{2}^{-1}\sum_{i<j}\widehat{q}\left(x,x',W_i,W_j\right), \tag{4}$$

with $\widehat{q}\left(x,x',w,w'\right)$ a preliminary estimate of the dyadic proxy variable regression function.

One approach would be to estimate $q\left(x,x',w,w'\right)$ non-parametrically.

An exploration of such an approach would be an interesting topic for future research.

A Correlated Random Effects Specification

Dyadic logit is 'reduced form' by construction.

Source of dependence across (i,j) and (i,k) is left unspecified.

Can we write down a likelihood and work backwards?

cf., p_2 model of van Duijn, Snijders and Zijlstra (2004, SN).

cf., 'fixed effects' models studied in Graham (2017, EM).

A Correlated Random Effects Specification (continued)

Links form according to

$$D_{ij} = \mathbf{1}\left(\left[t\left(X_i\right) + t\left(X_j\right)\right]' \beta_0 + \omega\left(X_i, X_j\right)' \gamma_0 + A_i + A_j - U_{ij} \leq 0\right)$$

with

$$U_{ij} \Big| X_i, X_j, W_i, W_j, A_i, A_j \sim \mathcal{N}\left(0, 1\right)$$

and independently distributed across dyads.

A Correlated Random Effects Specification (continued)

Posit the correlated random effects specification

$$A_i | X_i, W_i \sim N\left(\frac{\alpha_0}{2} + k(W_i)' \delta_0, \sigma_A^2\right)$$

with $k(W_i)$ a vector of known functions of the proxy variables.

A Correlated Random Effects Specification (continued)

Averaging over A_i and A_j gives a dyadic proxy variable regression function of

$$q\left(X_i, X_j, W_i, W_j; \pi_0\right) = \Phi\left(R_{ij}'\pi_0\right) \tag{5}$$

for

$$\pi_0 = \left(1 + 2\sigma_A^2\right)^{-1/2}\left(\alpha_0, \beta_0', \gamma_0', \delta_0'\right)'$$

and (redefining)

$$R_{ij} = \left(1, \left[t\left(X_i\right) + t\left(X_j\right)\right]', \omega\left(X_i, X_j\right)', \left[k\left(W_i\right) + k\left(W_j\right)\right]'\right)'$$

A Correlated Random Effects Estimation

1. Use $q\left(X_i, X_j, W_i, W_j; \pi_0\right) = \Phi\left(R'_{ij}\pi_0\right)$ and proceed as in logit case above

   (a) computationally straightforward

   (b) does not recover estimate of $\rho_0 = \sigma_A^2\left(1 + 2\sigma_A^2\right)^{-1}$

2. Maximize integrated likelihood (high dimensional integral, MCMC, efficient?)

3. Use composite likelihood ideas (“Triad Probit”, how inefficient?)

Triad Probit

Let $\eta_0 = \left(\alpha_0, \beta_0', \gamma_0', \delta_0'\right)'$ and $S_{ij} = 2D_{ij} - 1$. Consider the log-likelihood associated with the *pair* $\left(D_{ij}, D_{ik}\right)$:

$$\begin{aligned}\ln \Pr\left(D_{ij}, D_{ik} \middle| \mathbf{X}, \mathbf{W}; \theta_0\right) &= \ln \Phi\left(S_{ij}\frac{R_{ij}'\eta_0}{\sqrt{1+2\sigma_A^2}}, S_{ik}\frac{R_{ik}'\eta_0}{\sqrt{1+2\sigma_A^2}}; S_{ij}S_{ik}\rho_0\right) \\ &= l_{ijk}^*\end{aligned}$$

for $\theta_0 = \left(\eta_0', \rho_0\right)'$ and $Z_{ij} = \left(D_{ij}, R_{ij}'\right)'$.

Note $\left(1 + 2\sigma_A^2\right)^{-1} = 1 - 2\rho_0$.

Pairwise likelihood depends non-trivially on the distribution of the random effects $\{A_i\}_{i=1}^{\infty}$.

Triad Probit (continued)

Pairwise likelihood is not invariant to permutations of i, j and k.

Define the permutation invariant kernel

$$l_{ijk}(\theta) = \frac{1}{3}\left[l^*_{ijk} + l^*_{jik} + l^*_{kij}\right]$$

and associated criterion function

$$L_N(\theta) = \binom{N}{3}^{-1} \sum_{i<j<k} l_{ijk}(\theta).$$

Similar to a third-order U-process maximizer (e.g., Honore and Powell, 1994, JE).

Also like a composite likelihood (cf., Bellio and Varin, 2005, SM).

Triad Probit: Asymptotic Distribution

Quick outline

Let:

1. $S_N(\theta) = \binom{N}{3}^{-1} \sum_{i<j<k} s_{ijk}(\theta)$ with $s_{ijk}(\theta) = \frac{\partial l_{ijk}(\theta)}{\partial \theta}$.

2. Define $\Gamma_0 = \mathbb{E}\left[\frac{\partial^2 l_{ijk}(\theta)}{\partial \theta \partial \theta'}\right]$.

3. As earlier, $\Sigma_q = \mathbb{E}\left[s_{i_1 i_2 i_3} s'_{j_1 j_2 j_3}\right]$ equals the covariance of $s_{i_1 i_2 i_3}$ and $s_{j_1 j_2 j_3}$ when they share $q = 0, 1, 2, 3$ indices in common.

Triad Probit: Asymptotic Distribution (continued)

Calculation then gives

$$\mathbb{V}\left(\sqrt{N}S_N\left(\theta\right)\right) = 9\Sigma_1 + \frac{18}{N-1}\left(\Sigma_2 - 2\Sigma_1\right) + \frac{6}{\left(N-1\right)\left(N-2\right)}\left(\Sigma_3 + 3\Sigma_1\right)$$

which suggests, under regularity conditions, the limiting distribution

$$\sqrt{N}\left(\hat{\theta}_{\mathsf{TP}} - \theta_0\right) \xrightarrow{D} N\left(0, 9\Gamma_0^{-1}\Sigma_1\Gamma_0^{-1}\right). \tag{6}$$

Nyakatoke Example

	Dyadic Logit	Triad Probit
Lutheran	0.0674 (0.1042)	0.0404 (0.0445)
Muslim	0.0647 (0.1759)	0.0271 (0.0656)
Same religion	0.3836 (0.1274)	0.1940 (0.0461)
Other blood	1.5701 (0.2321)	0.8785 (0.1027)
Cousin, etc.	2.1031 (0.3090)	1.2227 (0.1889)
Child, etc.	3.4068 (0.2145)	2.0966 (0.1214)
ρ_0	-	0.0651 (0.0207)

Nyakatoke Example (continued)

cf., de Weerdt (2004, IAP)

Standard errors include higher-order variance terms.

$(1 - 2\widehat{\rho})^{1/2} = 0.9327$ and $\pi/\sqrt{3} = 1.8138$

Triad probit coefficients $\times 1.8138 \times 0.9327 \approx$ dyadic logit coefficients.

ASF Estimation

A flexible parametric ASF estimate is

$$\widehat{m}^{\mathsf{ASF}}\left(x,x';\widehat{\theta}_{\mathsf{TP}}\right)=\binom{N}{2}^{-1}\sum_{i<j}q\left(x,x',W_i,W_j;\widehat{\theta}_{\mathsf{TP}}\right). \qquad (7)$$

Sampling variability in (7) stems from two sources:

1. uncertainty about the marginal distribution function of the proxy variable, F_W,

2. uncertainty about the form of the proxy variable regression function (in this case its parameter θ_0).

ASF Estimation (continued)

A mean value expansion yields

$$\begin{aligned}\sqrt{N}\left(\hat{m}^{\mathsf{ASF}}\left(x,x';\hat{\theta}_{\mathsf{TP}}\right)-m^{\mathsf{ASF}}\left(x,x';\theta_0\right)\right)=&\sqrt{N}V_N\left(x,x'\right)\\&+\hat{M}_\theta\left(x,x'\right)\sqrt{N}\left(\hat{\theta}_{\mathsf{TP}}-\theta_0\right)\end{aligned}$$

with

$$V_N\left(x,x'\right)=\binom{N}{2}^{-1}\sum_{i<j}\left[q\left(x,x',W_i,W_j;\theta_0\right)-m^{\mathsf{ASF}}\left(x,x';\theta_0\right)\right]$$

$$\hat{M}_\theta\left(x,x'\right)=\binom{N}{2}^{-1}\sum_{i<j}\phi\left(L_{ij}\left(x,x'\right)'\bar{\theta}\right)L_{ij}\left(x,x'\right)$$

where $L_{ij}(x,x')=\left(1,\left[t(x)+t(x')\right]',\omega(x,x')',\left[k(W_i)+k\left(W_j\right)\right]'\right)'$, $\phi(\cdot)$ is the standard normal pdf, and $\bar{\theta}$ is a mean value between $\hat{\theta}_{\mathsf{TP}}$ and θ_0 which may vary from row to row.

ASF Estimation (continued)

Next observe that $V_N(x, x')$ is a second order U-Statistic and define

$$\begin{aligned}\Delta_q\left(x, x'\right) =& \mathbb{E}\left[\left(q\left(x, x', W_{i_1}, W_{i_2}; \theta_0\right) - m^{\mathsf{ASF}}\left(x, x'; \theta_0\right)\right)\right.\\ &\left.\times\left(q\left(x, x', W_{j_1}, W_{j_2}; \theta_0\right) - m^{\mathsf{ASF}}\left(x, x'; \theta_0\right)\right)'\right]\end{aligned}$$

when $\{i_1, i_2\}$ and $\{j_1, j_2\}$ share $q = 1, 2$ indices in common.

ASF Estimation (continued)

Standard projection arguments for U-statistics, as well as the limiting distribution for $\widehat{\theta}_{\mathsf{TP}}$ outlined above, suggest an asymptotic distribution for the ASF of, letting $\widehat{M}_\theta(x, x') \xrightarrow{p} M_\theta(x, x')$,

$$\sqrt{N}\left(\widehat{m}^{\mathsf{ASF}}\left(x, x'; \widehat{\theta}_{\mathsf{TP}}\right) - m^{\mathsf{ASF}}\left(x, x'; \theta_0\right)\right)$$
$$\xrightarrow{D} \mathcal{N}\left(0, 4\Delta_1\left(x, x'\right) + 9M_\theta\left(x, x'\right)\Gamma_0^{-1}\Sigma_1\Gamma_0^{-1}M_\theta\left(x, x'\right)'\right).$$

Dyadic regression wrap-up

For "fixed effect" estimation see Graham (2017, EM), Jochmans (2017, JBES) and Dzemski (2014, WP).

Other settings with group production.

Several theoretical questions are open.