Dyadic Regression

Bryan S. Graham, UC - Berkeley & NBER

October 1, 2018

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; Tabord-Meehan, 2018).

This note, drawing from material in Graham (TBD), presents some basic results on composite (pseudo-) maximum likelihood estimation of a broad class of (parametric) dyadic regression models. The approach to analyzing these models developed here is, to my knowledge, a new one (albeit building on ideas in Graham (2017)). However, the analysis does provide alternative perspectives on several known methods (e.g., the dyadic cluster robust standard error estimator introduced by Fafchamps & Gubert (2007)).

The material in this note is currently being expanded into a short paper with the same title.

Data generating process

Let $\{X_i, U_i\}_{i=1}^N$ be an independent and identically distributed (iid) random sequence. For each of the $N(N-1)$ ordered dyads, (i, j), the outcome of interest – for example total exports from country i to country j – is generated according to

$$Y_{ij} = \tilde{h}\left(X_i, X_j, U_i, U_j, V_{ij}\right), \tag{1}$$

with V_{ij} independent of $\{X_i, U_i\}_{i=1}^N$ and additionally iid across (ordered) dyads.

We parameterize the conditional mean of the outcome as

$$\mathbb{E}\left[\left.Y_{ij}\right|X_i,X_j\right]=\pi\left(R_{ij}'\theta_0\right)$$

with $\pi(\cdot)$ a known function, R_{ij} a $K\times 1$ vector of known functions of X_i and X_j, and θ_0 an unknown parameter. In many cases $\pi^{-1}(\cdot)$ will coincide with the canonical link function associated with the appropriate generalized linear model (GLM). For example $\pi(\cdot)=\exp(\cdot)/[1+\exp(\cdot)]$ and $\pi(\cdot)=\exp(\cdot)$ would be natural choices when the $\{Y_{ij}\}$ are, respectively, binary and counts.

The gravity model of trade, introduced over fifty years ago by Tinbergen (1962), belongs to the above family of models. In (a common variant of) the gravity model exports from country i to country j are given by

$$Y_{ij}=\exp\left(R_{ij}'\theta_0\right)A_iB_jV_{ij}, \tag{2}$$

with A_i, B_i and V_{ij} mean one random variables and V_{ij} independent of $\{X_i,U_i\}_{i=1}^N$ for $U_i=(A_i,B_i)'$ and iid across (ordered) dyads. Here the $\{A_i\}_{i=1}^N$ and $\{B_i\}_{i=1}^N$ sequences correspond, respectively, to exporter and importer heterogeneity terms. In this set-up θ_0 is a structural economic parameter (cf., Head & Mayer, 2014). Whether the dyadic regression function $\mathbb{E}\left[\left.Y_{ij}\right|X_i,X_j\right]$ actually equals $\exp\left(R_{ij}'\theta_0\right)$ depends on whether $\mathbb{E}\left[\left.U_i\right|X_i\right]=(1,1)'$. However θ_0 might also just index a parametric reduced form regression function. For now we will defer consideration of these identification issues, focusing instead on inference issues raised by dyadic data.

Composite maximum likelihood estimation

For concreteness we will consider the case where $\pi\left(R_{ij}'\theta_0\right)=\exp\left(R_{ij}'\theta_0\right)$. Santos Silva & Tenreyro (2006) use a conditional mean function of this form to model directed trade flows between $N=136$ countries. In their application R_{ij} includes variables such as exporter and importer GDP, distance between the two trading countries, whether they have a preferential trading agreement etc. Let $l_{ij}(\theta)=Y_{ij}R_{ij}'\theta-\exp\left(R_{ij}'\theta\right)$; this term equals (up to a term not varying with θ) the log likelihood of a Poisson random variable with mean $\exp\left(R_{ij}'\theta\right)$.

We will study the estimator which chooses $\hat{\theta}_{\mathrm{DR}}$ to maximize

$$L_N(\theta)=\frac{1}{N}\frac{1}{N-1}\sum_i\sum_{j\neq i}l_{ij}(\theta). \tag{3}$$

Observe that $\hat{\theta}_{\mathrm{DR}}$ coincides with a (pseudo-) maximum likelihood estimator based on the assumption that $\{Y_{ij}\}_{i<j}$ are independent Poisson random variables conditional on $\mathbf{X} = (X_1, \ldots, X_N)'$. Typically, however, there will be dependence across any two summands in (3) which share at least one index in common. For example, trade from the United States to China likely covaries with that from the United States to Japan, even after conditioning on $\mathbf{X}$.

Dependence across dyads sharing one or more agents in common means that, even if $Y_{ij}|\,\mathbf{X}$ is well described by a Poisson distribution, equation (3) is not the correct log-likelihood function for $\{Y_{ij}\}_{i<j}\Big|\,\mathbf{X}$. Accounting for the implications of dependence across the summands in (3) is the main topic of this note. Estimation is straightforward since a conventional Poisson regression program may be used. The calculation of standard errors is, as I outline next, more complicated.

Although the form of the kernel in (3) was motivated by a Poisson assumption on the distribution of $\{Y_{ij}\}_{i<j}\Big|\,\mathbf{X}$, this is not required for consistency of $\hat{\theta}_{\mathrm{DR}}$ for θ_0. Consistency requires only that the conditional mean of Y_{ij} given $\mathbf{X}$ takes the assumed exponential form. On might call $\hat{\theta}_{\mathrm{DR}}$ a **pseudo-composite maximum likelihood estimator** (PCMLE). See Lindsey (1988) and Gourieroux & Monfort (1995) for discussions of, respectively, composite and pseudo maximum likelihood estimation.

A mean value expansion of the first order condition associated with the maximizer of (3) yields, after some re-arrangement,

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

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

$$S_N\left(\theta\right) = \frac{1}{N}\frac{1}{N-1}\sum_{i}\sum_{j\neq i} s_{ij}\left(Z_{ij},\theta\right) \tag{4}$$

with $s\left(Z_{ij},\theta\right) = \partial l_{ij}\left(\theta\right)/\partial\theta$ for $Z_{ij} = \left(Y_{ij}, X_i', X_j'\right)$ 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 we will just assume that $H_N\left(\bar{\theta}\right) \xrightarrow{p} \Gamma_0$, with Γ_0 invertible. Newey & McFadden (1994) discuss Hessian convergence; Jochmans (2017) and Graham (2017) provide specific examples similar to the present context.

In our Poisson example the "score" contribution of directed dyad (i,j) equals

$$s\left(Z_{ij},\theta\right) = \left(Y_{ij} - \exp\left(R_{ij}'\theta\right)\right)R_{ij}.$$

More generally we will work with "scores" of the form $s(Z_{ij},\theta)=\left(Y_{ij}-\pi\left(R_{ij}'\theta\right)\right)k\left(X_i,X_j;\theta\right)$. This family contains all GLM models with canonical link functions as a special case (as well as others). Most of what follows could be extended to more general score functions in any case.

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

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

so that the asymptotic sampling properties of $\sqrt{N}\left(\hat{\theta}_{\mathrm{DR}}-\theta_0\right)$ will be driven by the behavior of $\sqrt{N}S_N\left(\theta_0\right)$. As pointed out by Fafchamps & Gubert (2007) and others, (4) is not a sum of independent random variables, hence a basic central limit theorem (CLT) cannot be (directly) applied.

Variance of $\sqrt{N}S_N\left(\theta_0\right)$

My analysis of $\sqrt{N}S_N\left(\theta_0\right)$ borrows from the theory of U-Statistics (e.g., Ferguson, 2005). To make these connections clear it is convenient to re-write $S_N\left(\theta_0\right)$ as

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

where $\bar{s}\left(Z_{ij},\theta\right)$ is invariant to permutations of its indices. Note that $S_N\left(\theta\right)$ is not a U-Statistics since Z_{ij} includes both dyadic-specific (Y_{ij}) and agent-specific (X_i) random variables. Nevertheless tools from the literature on U-Statistics can be applied. In this spirit, a Hoeffding (1948) variance decomposition gives

$$\mathbb{V}\left(\sqrt{N}S_N\left(\theta_0\right)\right)=4\Sigma_1+\frac{2}{N-1}\left(\Sigma_2-2\Sigma_1\right) \tag{5}$$

where

$$\Sigma_q=\mathbb{E}\left[\bar{s}\left(Z_{i_1i_2};\theta_0\right)\bar{s}\left(Z_{j_1j_2};\theta_0\right)'\right] \tag{6}$$

when the dyads $\{i_1,i_2\}$ and $\{j_1,j_2\}$ share $q=0,1,2$ agents in common and the sampling assumption and (4) ensure independence across dyads with *no* agents in common (i.e., $\Sigma_0=0$). Calculation (5) suggests the finite sample variance approximation

$$\mathbb{V}\left(\sqrt{N}\left(\hat{\theta}_{\mathrm{DR}}-\theta_0\right)\right)\approx\Gamma_0^{-1}\left[4\Sigma_1+\frac{2}{N-1}\left(\Sigma_2-2\Sigma_1\right)\right]\Gamma_0^{-1}. \tag{7}$$

Limit distribution of $\sqrt{N}S_N(\theta_0)$

To characterize the sampling distribution of $\sqrt{N}S_N(\theta_0)$ when N grows large, I use a double projection argument. First I project $s(Z_{ij};\theta_0)$ onto an arbitrary function of X_i, X_j, U_i and U_j. This projection allows me to represent $S_N(\theta_0)$ as the sum of a U-statistic and an (asymptotically normal, but negligible) remainder term. Normal Hajek Projection type arguments can then be applied to the U-Statistics component of the full statistic. A similar method features in Graham (2017).

Write $S_N(\theta) = V_N(\theta) + T_N(\theta)$ where $V_N(\theta)$ is the 2nd order U-Statistic

$$V_N(\theta) = \binom{N}{2}^{-1} \sum_{i<j} v(X_i, U_i, X_j, U_j; \theta) \tag{8}$$

with kernel, letting $h(x, x', u, u') = \mathbb{E}\left[\tilde{h}(x, x', u, u', V_{ij})\right]$,

$$\begin{aligned} v(X_i, U_i, X_j, U_j; \theta) &\stackrel{def}{\equiv} \mathbb{E}\left[\left.\bar{s}(Z_{ij}, \theta)\right| X_i, U_i, X_j, U_j\right] \\ &= \frac{1}{2}\left\{\left(h(X_i, X_j, U_i, U_j) - \pi\left(R_{ij}'\theta\right)\right) k(X_i, X_j; \theta)\right. \\ &\quad \left. + \left(h(X_j, X_i, U_j, U_i) - \pi\left(R_{ji}'\theta\right)\right) k(X_j, X_i; \theta)\right\}. \end{aligned}$$

and

$$T_N(\theta) = \binom{N}{2}^{-1} \sum_{i<j} t(X_i, U_i, X_j, U_j, Y_{ij}; \theta). \tag{9}$$

with

$$\begin{aligned} t(X_i, U_i, X_j, U_j, Y_{ij}; \theta) &\stackrel{def}{\equiv} \bar{s}(Z_{ij}, \theta) - \mathbb{E}\left[\left.\bar{s}(Z_{ij}, \theta)\right| X_i, U_i, X_j, U_j\right] \\ &= (Y_{ij} - h(X_i, X_j, U_i, U_j)) k(X_i, X_j; \theta) \\ &\quad + (Y_{ji} - h(X_j, X_i, U_j, U_i)) k(X_j, X_i; \theta). \end{aligned}$$

Let $V_N = V_N(\theta_0)$ and $T_N = T_N(\theta_0)$. The L^2 projection of $V_N(\theta)$ onto (X_1, U_1) is

$$\begin{aligned} \mathbb{E}\left[\left.V_N(\theta)\right| X_1, U_1\right] &= \frac{2}{N(N-1)} \sum_{i<j} \mathbb{E}\left[\left.v(X_i, U_i, X_j, U_j; \theta)\right| X_1, U_1\right] \\ &= \frac{2}{N(N-1)} \left\{(N-1)\bar{v}_1(X_1, U_1; \theta)\left[\frac{N(N-1)}{2} - (N-1)\right] \cdot \bar{v}_0(\theta)\right\} \\ &= \frac{2}{N}\left\{\bar{v}_1(X_1, U_1; \theta) - \bar{v}_0(\theta)\right\} + \bar{v}_0(\theta) \end{aligned}$$

with $\bar{v}_0(\theta) \stackrel{def}{\equiv} \mathbb{E}[v(X_1,U_1,X_2,U_2;\theta)]$, $\bar{v}_1(x,u;\theta) \stackrel{def}{\equiv} \mathbb{E}[v(x,u,X,U;\theta)]$. Note that $\bar{v}_0(\theta_0)=0$. From the above calculations we get a Hájek projection of

$$\Pi\left(V_N(\theta)-\bar{v}_0(\theta)\middle|\mathcal{L}\right)=\frac{2}{N}\sum_{i=1}^{N}\left\{\bar{v}_1(X_i,U_i;\theta)-\bar{v}_0(\theta)\right\}. \tag{10}$$

Next observe that (10), at $\theta=\theta_0$, is a sum of iid random variables. A CLT (and Slutsky's Theorem) give

$$\sqrt{N}\left(\hat{\theta}_{\mathrm{DR}}-\theta_0\right)\xrightarrow{D}\mathcal{N}\left(0,4\left(\Gamma_0'\Omega_1^{-1}\Gamma_0\right)^{-1}\right), \tag{11}$$

where $\Omega_1=\mathbb{E}\left[\bar{v}_1(X_i,U_i;\theta_0)\,\bar{v}_1(X_i,U_i;\theta_0)'\right]$. By direct calculation it is possible to show that $\Omega_1=\Sigma_1$ for Σ_1 as defined in (6) above. This suggest the variance estimator

$$\hat{\Omega}_1=\frac{1}{N}\sum_{i=1}^{N}\hat{\bar{s}}_{1,i}\hat{\bar{s}}_{1,i}'$$

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

The variance expression, equation (5), indicates that inference based upon the limit distribution (11) ignores higher order variance terms. 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 and power properties.

Bibliographic notes

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 are rare. Within the empirical trade literature, when working with directed outcome data, some authors do cluster on pairs of countries, allowing for dependence across observations (i,j) and (j,i), but ruling out dependence across, for example, observations (i,j) and (i,k). More commonly simple heteroscedastic robust standard errors are reported (see for example the *Handbook of International Economics* chapter by Head & Mayer (2014)). This approach to inference implicitly assumes that all $N(N-1)$ "score" terms are uncorrelated with one another (an assumption violated by the data generating process used in the simulations reported in Head & Mayer (2014); see, for example, Table 3.3).

Fafchamps & Gubert (2007) proposed a variance-covariance estimator which allows for dyadic-dependence. Their estimator coincides with the bias-corrected discussed in Graham (TBD). Additional versions (and analyses) of this estimator are provided by Cameron & Miller (2014) and Aronow et al. (2017). None of these authors demonstrate asymptotic normality of $\sqrt{N}\left(\hat{\theta}_{\mathrm{DR}}-\theta_0\right)$ as is done here. Tabord-Meehan (2018) does show asymptotic normality using a rather different approach (and also covering more cases than the results sketched).

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 $\hat{\Omega}_1$ defined above.

References

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.

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.

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. (1995). *Statistics and Econometric Models*, volume 1. Cambridge: Cambridge University Press.

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.

Head, K. & Mayer, T. (2014). *Handbook of International Economics*, volume 4, chapter Gravity equations: workhorse, toolkit, and cookbook, (pp. 131 – 195). North-Holland: Amsterdam.

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

Jochmans, K. (2017). Two-way models for gravity. *Review of Economics and Statistics*, 99(3), 478 – 485.

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

Newey, W. K. & McFadden, D. (1994). *Handbook of Econometrics*, volume 4, chapter Large sample estimation and hypothesis testing, (pp. 2111 – 2245). North-Holland: Amsterdam.

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

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

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

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.