# Policy Analysis for Dyadic Outcomes

## Econometric Methods for Social Spillovers and Networks

University of St. Gallen, September 28th to October 6th, 2020

*(see also my summer 2020 Chamberlain seminar)*

*Bryan S. Graham*

University of California - Berkeley

## Introduction

One motivation for Tinbergen's (1967) dyadic regression analysis was to evaluate the effect of preferential trade agreements on export flows (see also Rose (2004, *AER*).

Baldwin and Taglioni (2007) use gravity models to assess whether common currency zones, such as the Eurozone, promote trade.

Goodman (2017) analyzes whether Medcaid expansion states experienced in-migration.

Oretega and Peri's (2013) study the relationship between immigration entry tightness and cross-country migration.

Substantial public interest in the effects of these policies.

## Setup

Let $W_i \in \mathbb{W} = \{w_1, \ldots, w_K\}$ and $X_i \in \mathbb{X} = \{x_1, \ldots, x_L\}$ be a finite set of *ego* and *alter* policies.

$\mathbb{W}$ might enumerate different export promotion policies (e.g., tax subsidies or preferential credit schemes for exporting firms).

$\mathbb{X}$ might enumerate different combinations of protectionist policies (e.g., tariff levels).

How do different counterfactual combinations of ego and alter policy pairs map into (distributions of) outcomes?

## Potential Outcome

**Assumption 1** (Dyadic Potential Response Function) For any ego-alter pair $i, j \in \mathbb{N}$ with $i \neq j$, the potential (directed) outcome associated with adopting the pair of policies $W_i = w$ and $X_j = x$ is given by

$$Y_{ij}(w, x) = h\left(w, x, A_i, B_j, V_{ij}\right), \; x \in \mathbb{X}, \; w \in \mathbb{W} \qquad (1)$$

with $\{(A_i, B_i)\}_{i\in\mathbb{N}}$ and $\left\{\left(V_{ij}, V_{ji}\right)\right\}_{i,j\in\mathbb{N},i<j}$ both i.i.d. sequences additionally independent of each other and $h : \mathbb{W}\times\mathbb{X}\times\mathbb{A}\times\mathbb{B}\times\mathbb{V} \to \mathbb{Y}$ a measurable function. The ego and alter effects, respectively $A_i$ and $B_i$, induce dependence across any pair of potential outcomes sharing an agent in common.

Structured “interference” between units; a violation of SUTVA.

## Causal Effects

The effect on $ij$'s outcome of adopting policy pair $(w', x')$ vs. $(w, x)$ is

$$Y_{ij}(w', x') - Y_{ij}(w, x).$$

Identification of such effects at the dyad-level is infeasible.

The econometrician only observes the outcome associated with the policy pair actually adopted.

$$Y_{ij} \overset{def}{\equiv} Y_{ij}\left(W_i, X_j\right). \tag{2}$$

## Average Structural Function

(i) draw an ego unit at random from the target population and exogenously assign it policy $W_i = w$;

(ii) independently draw an alter unit at random and assign it policy $X_j = x$.

The expected outcome is

$$m^{\mathsf{ASF}}(w,x) \overset{def}{\equiv} \mathbb{E}\left[Y_{12}(w,x)\right] \overset{def}{\equiv} \int\int\int \bar{h}(w,x,a,b)\, f_{A_1}(a)\, f_{B_2}(b)\, \mathrm{d}a\mathrm{d}b, \tag{3}$$

where the second '$\overset{def}{\equiv}$' follows from $\bar{h}(w,x,a,b) \overset{def}{\equiv} \mathbb{E}\left[h(w,x,a,b,V_{12})\right] \overset{def}{\equiv} \bar{Y}_{ij}(w,x)$.

## Average Structural Function (continued)

Differences of the form

$$m^{\mathsf{ASF}}\left(w', x'\right) - m^{\mathsf{ASF}}(w, x)$$

measure the expected effects of different combinations of policies on the directed dyadic outcome.

The double difference

$$m^{\mathsf{ASF}}(1,1) - m^{\mathsf{ASF}}(0,1) - \left[m^{\mathsf{ASF}}(1,0) - m^{\mathsf{ASF}}(0,0)\right] \tag{4}$$

measures complementarity in a binary policy/treatment.

## Parametric Example

Assume that $\mathbb{W} = \mathbb{X} = \{0, 1\}$.

A parametric form for $Y_{ij}(w, x)$ is:

$$Y_{ij}(w, x) = \alpha + w\beta + x\gamma + wx\delta + A_i + B_j + V_{ij}. \tag{5}$$

Response (5) implies that treatment effects are constant across units, for example,

$$Y_{ij}(1, 1) - Y_{ij}(0, 0) = \beta + \gamma + \delta,$$

which is constant in $i \in \mathbb{N}$.

## Identification (Proxy Variables)

**Assumption 2** (Redundancy) For $R_i \in \mathcal{R} \subseteq \mathbb{R}^{\mathsf{dim}(R)}$ a proxy variable for $A_i$, and $S_i \in \mathcal{S} \subseteq \mathbb{R}^{\mathsf{dim}(S)}$ a proxy variable for $B_i$, we have that

$$\mathbb{E}\left[Y_{ij}\left(w,x\right)\middle| W_i, X_j, A_i, B_j, R_i, S_j\right] = \mathbb{E}\left[Y_{ij}\left(w,x\right)\middle| W_i, X_j, A_i, B_j\right],$$

for any $w \in \mathbb{W}$ and $x \in \mathbb{X}$.

This is a redundancy assumption.

It asserts that the proxies $R_i$ and $S_j$ have no predictive power (in the conditional mean sense) for $Y_{ij}(w,x)$ conditional on the latent ego and alter attributes $A_i$ and $B_j$.

# <u>Identification (Strict Exogeneity)</u>

**Assumption 3** (Strict Exogeneity) The $ij$ ego-alter treatment assignment $\left(W_i, X_j\right)$ is independent of $V_{ij}$ conditional on the latent ego $A_i$ and alter $B_j$ effects:

$$V_{ij} \perp \left(W_i, X_j\right) \Big| \, A_i = a, B_i = b, \; a \in \mathbb{A}, \; b \in \mathbb{B}. \tag{6}$$

Assumption 3, which involves conditioning on *unobservables*, has no clear analog in the standard program evaluation model.

Closely related to notion of Strict Exogeneity in Chamberlain (1984).

## Relationship to Panel Data

In our parametric model (6) implies that

$$\mathbb{E}\left[Y_{ij}\middle| W_i, X_j, A_i, B_j\right] = \alpha + W_i\beta + X_j\gamma + W_iX_j\delta + A_i + B_j \quad (7)$$

since

(i) Assumption 3 gives $\mathbb{E}\left[V_{ij}\middle| W_i, X_j, A_i, B_j\right] = \mathbb{E}\left[V_{ij}\middle| A_i, B_j\right]$ and

(ii) $\mathbb{E}\left[V_{ij}\middle| A_i, B_j\right] = \mathbb{E}\left[V_{ij}\right]$ by independence of $\{(A_i, B_i)\}_{i=1}^{N}$ and $\left\{\left(V_{ij}, V_{ji}\right)\right\}_{i<j}$

(setting $\mathbb{E}\left[V_{ij}\right] = 0$ is a normalization).

## Relationship to Panel Data (continued)

Equation (7) implies, for example, that

$$\mathbb{E}\left[Y_{ij} - Y_{il} - \left(Y_{kj} - Y_{kl}\right) \middle| W_i, X_j, A_i, B_j\right] = (W_i - W_k)\left(X_j - X_l\right)\delta.$$

"Within-tetrad" variation identifies $\delta$.

This is similar to how within-group variation in a strictly exogenous regressor identifies its corresponding coefficient in the panel context.

## Identification (Strict Exogeneity)

Under Assumption 3:

$$
\begin{aligned}
f_{V_{12},A_1,W_1,B_2,X_2}(v_{12},a_1,w_1,b_2,x_2) =& f_{V_{12}|A_1,W_1,B_2,X_2}(v_{12}|\, a_1,w_1,b_2,x_2) \\
& \times f_{A_1,W_1}(a_1,w_1)\, f_{B_2,X_2}(b_2,x_2) \\
=& f_{V_{12}|A_1,B_2}(v_{12}|\, a_1,b_2)\, f_{A_1,W_1}(a_1,w_1) \\
& \times f_{B_2,X_2}(b_2,x_2) \\
=& f_{V_{12}}(v_{12})\, f_{A_1,W_1}(a_1,w_1) \\
& \times f_{B_2,X_2}(b_2,x_2)
\end{aligned}
$$

units 1 and 2 independent random draws $\Rightarrow$ first $=$;

strict exogeneity $\Rightarrow$ second $=$;

independence of $\{(A_i, B_i)\}_{i=1}^{N}$ and $\left\{\left(V_{ij}, V_{ji}\right)\right\}_{i<j}$ $\Rightarrow$ third $=$.

## Identification (Strict Exogeneity)

This factorization clarifies that the effect of Assumption 3 is to ensure that all "endogeneity" in treatment choice is reflected in dependence between $W_i$ and $A_i$ and/or $B_j$ and $X_j$.

Conditional on these two latent variables, variation in treatment is "idiosyncratic" or exogenous.

## Identification (Conditional Independence)

**Assumption 4** (Conditional Independence) An ego's (alter's) treatment choice varies independently of their latent effect $A_i$ ($B_j$) given the observed proxy $R_i$ ($S_j$):

$$A_i \perp W_i |\, R_i = r,\ r \in \mathcal{R} \subseteq \mathbb{R}^{\mathrm{dim}(R)} \tag{8}$$

$$B_i \perp X_i |\, S_i = s,\ s \in \mathcal{S} \subseteq \mathbb{R}^{\mathrm{dim}(S)}. \tag{9}$$

Assumption  is a standard one in the context of single agent program evaluation problems.

It asserts – for example – that $A_i$ and $W_i$ vary independently within subpopulations homogenous in the proxy variable $R_i$.

## Proxy Variable Regression Function

Assumptions 1 to 4, plus an additional support condition described below, are sufficient to show identification of the ASF.

Let

$$q(w,x,r,s) = \mathbb{E}\left[Y_{ij}\middle| W_i = w, X_j = x, R_i = s, S_j = s\right] \qquad (10)$$

be the dyadic proxy variable regression (PVR).

## Proxy Variable Regression Function (continued)

The PVR relates to $\bar{Y}_{12}(w,x) = \bar{h}(w,x,A_1,B_2)$ as follows:

$$
\begin{aligned}
q(w,x,r,s) =& \mathbb{E}\left[ h\left(W_i, X_j, A_i, B_j, V_{ij}\right) \middle| W_i = w, X_j = x, R_i = r, S_j = s \right] \\
=& \mathbb{E}\left[ \mathbb{E}\left[ h\left(W_i, X_j, A_i, B_j, V_{ij}\right) \middle| W_i = w, X_j = x, A_i, B_j, R_i = r, S_j \right. \right. \\
& \left. \middle| W_i = w, X_j = x, R_i = r, S_j = s \right] \\
=& \mathbb{E}\left[ \mathbb{E}\left[ h\left(W_i, X_j, A_i, B_j, V_{ij}\right) \middle| W_i = w, X_j = x, A_i, B_j \right] \right. \\
& \left. \middle| W_i = w, X_j = x, R_i = r, S_j = s \right] \\
=& \mathbb{E}\left[ \bar{h}\left(w, x, A_i, B_j\right) \middle| W_i = w, X_j = x, R_i = r, S_j = s \right] \\
=& \int_a \int_b \bar{h}(w,x,a,b)\, f_{A|R}(a|\, r)\, f_{B|S}(b|\, s)\, \mathrm{d}a \mathrm{d}b \\
=& \mathbb{E}\left[ \bar{Y}_{12}(w,x) \middle| R_1 = r, S_2 = s \right]. \qquad (11)
\end{aligned}
$$

## Main Identification Result

Equation (11) gives the identification result

$$\mathbb{E}_R\left[\mathbb{E}_S\left[q\left(w,x,R_i,S_j\right)\right]\right]=\int_r\int_s\left[\int_a\int_b\bar{h}\left(w,x,a,b\right)f_{A|R}\left(a|\,r\right)f_{B|S}\left(b|\,s\right)\mathsf{d}a\mathsf{d}b\right] \tag{12}$$

$$\begin{aligned}&f_R\left(r\right)f_S\left(s\right)\mathsf{d}r\mathsf{d}s\\ =&\int_a\int_b\bar{h}\left(w,x,a,b\right)f_A\left(a\right)f_B\left(b\right)\mathsf{d}a\mathsf{d}b\\ =&\mathbb{E}\left[\bar{Y}_{12}\left(w,x\right)\right]\\ =&m^{\mathsf{ASF}}\left(w,x\right).\end{aligned}$$

## Overlap

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

$$f_{R|W}(r|\,x)\,f_{S|X}(s|\,x) > 0,$$

while the integration in (12) is over $\mathcal{R} \times \mathcal{S}$, we require a formal support condition:

$$\mathbb{S}(w,x) \overset{def}{\equiv} \Big\{r, s \,:\, f_{R|W}(r|\,w)\,f_{S|X}(s|\,x) > 0\Big\} = \mathcal{R} \times \mathcal{S}. \qquad (13)$$

## Overlap (Discrete Policies)

When $W_i$ and $X_j$ are discretely-valued, with a finite number of support points, as assumed here, (13) can be expressed in a form similar to the overlap condition familiar from the program evaluation literature.

**Assumption 5** (Overlap) For $(w, x)$ the ego-alter treatment combination of interest

$$p_w(r)\, p_x(s) \geq \kappa > 0 \text{ for all } (r, s) \in \mathcal{R} \times \mathcal{S}$$

where $p_w(r) \overset{def}{\equiv} \Pr\left(W_i = w \mid R_i = r\right)$ and $p_x(s) \overset{def}{\equiv} \Pr\left(X_i = x \mid S_i = s\right)$.

## Main Identification Result (Summary)

**Theorem 2** Under Assumptions 1 through 5 the ASF is identified by

$$m^{\mathsf{ASF}}(w,x) = \int\int q(w,x,r,s)\, f_R(r)\, f_S(s)\, \mathrm{d}r\mathrm{d}s. \tag{14}$$

Theorem 2 shows that the ASF is identified by double marginal integration over the dyadic proxy variable regression function.

Double marginal integration also features in Graham, Imbens and Ridder (2018, *JBES*), in the context of identifying an average match function (AMF).

# Estimation

Let $q(w, x, r, s; \gamma)$ be a parametric model for the dyadic proxy variable regression function.

If the outcome of interest is export flows, we might specify that

$$q(w, x, r, s; \gamma) = \exp\left(t(Q_i)' \gamma\right),$$

with $Q_i = \left(W_i', X_i', R_i', S_i'\right)'$ and $t(Q_i)$ a finite (and pre-specified) set of basis functions (preferably including interactions of terms in the treatment variables – $W, X$ – and proxy variables – $R, S$).

We can estimate $\gamma$ using the Poisson dyadic regression estimator described earlier.

## Estimation (continued)

From our general results on dyadic regression we get the asymptotically linear representation

$$\sqrt{N}\left(\widehat{\gamma}-\gamma_0\right)=-\Gamma_0^{-1}\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\overline{s}_1^e\left(Q_i,U_i;\gamma_0\right)+\overline{s}_1^a\left(Q_i,U_i;\gamma_0\right)}{2}\right\}+o_p\left(1\right) \tag{15}$$

with $U_i=(A_i,B_i)'$ and $Q_i$ playing the role of "$X_i$" in our earlier results.

## Estimation (continued)

With an estimate of $\gamma$ in hand, form the fitted values $\left\{q\left(w, x, R_i, S_j; \widehat{\gamma}\right)\right\}_{i<}$ and, invoking Theorem 2, compute the analog estimate

$$\widehat{m}^{\text{ASF}}(w, x; \widehat{\gamma}) = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{q\left(w, x, R_i, S_j; \widehat{\gamma}\right) + q\left(w, x, R_j, S_i; \widehat{\gamma}\right)}{2}. \tag{16}$$

Two step procedure:

1. Estimate $\gamma_0$ by dyadic regression;

2. Compute $\widehat{m}^{\text{ASF}}(w, x; \widehat{\gamma})$

(U-Statistic with estimated nuisance parameter).

## Estimation (continued)

With the PVR satisfying some conditions (see Assumption 6 in Handbook Chapter), we have that

$$\begin{aligned}\sqrt{N}\left(\widehat{m}^{\mathsf{ASF}}(w,x;\widehat{\gamma})-m^{\mathsf{ASF}}(w,x;\gamma_0)\right)=&\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\psi_0(w,x,R_i,S_i;\gamma_0)\\&+M_0(w,x)\sqrt{N}(\widehat{\gamma}-\gamma_0)+o_p(1)\end{aligned}$$

where

$$\psi_0(w,x,R_1,S_1;\gamma)=\frac{q^e(w,x,R_1;\gamma)+q^a(w,x,S_1;\gamma_0)}{2}-m^{\mathsf{ASF}}(w,x;\gamma)$$

$$M_0(w,x)=\frac{1}{2}\mathbb{E}\left[\frac{\partial q(w,x,R_1,S_2;\gamma_0)}{\partial\gamma'}+\frac{\partial q(w,x,R_2,S_1;\gamma_0)}{\partial\gamma'}\right]$$

## Influence Function

with the $q^e(w,x,r;\gamma)$ and $q^a(w,x,s;\gamma)$ terms in the previous slide equal to

$$\begin{aligned} q^e(w,x,r;\gamma) &= \mathbb{E}_S\left[q(w,x,r,S;\gamma)\right] \\ q^a(w,x,s;\gamma) &= \mathbb{E}_R\left[q(w,x,R,s;\gamma)\right]. \end{aligned}$$

Plugging on our results for $\sqrt{N}(\widehat{\gamma}-\gamma_0)$ we get:

$$\begin{aligned} &\sqrt{N}\left(\widehat{m}^{\mathsf{ASF}}(w,x;\widehat{\gamma}) - m^{\mathsf{ASF}}(w,x;\gamma_0)\right) \\ =&\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\psi_0(w,x,R_1,S_1;\gamma_0) \\ &- M_0(w,x)\,\Gamma_0^{-1} \\ &\times\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\overline{s}_1^e(Q_i,U_i;\gamma_0)+\overline{s}_1^a(Q_i,U_i;\gamma_0)}{2}\right\}+o_p(1). \end{aligned}$$

## Limit Distribution

Under correct (enough) specification of the composite likelihood, which will typically follow if the parametric form of the the PVR function is itself correctly specified, both $\bar{s}_1^e(Q_1, U_1; \gamma_0)$ and $\bar{s}_1^a(Q_1, U_1; \gamma_0)$ will be conditional mean zero given $Q_1$.

The first and second terms in the influence function on the previous slide will be uncorrelated with each other.

# Limit Distribution (continued)

We get a limit distribution of

$$\sqrt{N}\left(\widehat{m}^{\mathsf{ASF}}(w,x;\widehat{\gamma}) - m^{\mathsf{ASF}}(w,x;\gamma_0)\right) \xrightarrow{D}$$
$$\mathcal{N}\left(0, 4\Xi_0(w,x) + 4M_0(w,x)\left(\Gamma_0'\Sigma_1^{-1}\Gamma_0\right)^{-1} M_0(w,x)'\right)$$

with

$$\Xi_0(w,x) = \mathbb{V}\left(\psi_0(w,x,R_1,S_1;\gamma_0)\right)$$

and

$$\Sigma_1 = \mathbb{V}\left(\frac{\bar{s}_1^e(Q_i,U_i;\gamma_0) + \bar{s}_1^a(Q_i,U_i;\gamma_0)}{2}\right).$$

## Limit Distribution (continued)

The first term in the asymptotic variance reflects the econometrician's imperfect knowledge of the distribution of the proxy variables $\left(R_i', S_i'\right)'$.

The second term reflects the asymptotic penalty associated with not knowing the conditional distribution of $Y_{12}$ given $W_1, X_2, R_1, S_2$.

See Graham (2011, *Econometrica*) and Graham, Imbens and Ridder (2018, *JBES*) for more expansive discussions in related contexts.

## Practical Implications

In practice may want to use a variance estimate that includes estimates of asymptotically negligible terms.

Can also use bootstrap discussed earlier.

Many potential applications in international trade and other fields.