

Policy Analysis for Dyadic Outcomes

Jinan University Econometrics Camp
Guangzhou, China, December 9-12, 2019

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 Medicaid 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 $W_i \in \mathbb{W} = \{w_1, \dots, w_K\}$ and $X_i \in \mathbb{X} = \{x_1, \dots, x_L\}$ 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

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

with $\{(A_i, B_i)\}_{i \in \mathbb{N}}$ and $\{(V_{ij}, V_{ji})\}_{i, j \in \mathbb{N}, i < j}$ both i.i.d. sequences additionally independent of each other and $h : \mathbb{W} \times \mathbb{X} \times \mathbb{A} \times \mathbb{B} \times \mathbb{V} \rightarrow \mathbb{Y}$ 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.

$$Y_{ij} \stackrel{def}{=} Y_{ij}(W_i, X_j). \quad (2)$$

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

$$m^{\text{ASF}}(w, x) \stackrel{\text{def}}{=} \mathbb{E}[Y_{12}(w, x)] \tag{3}$$
$$\stackrel{\text{def}}{=} \int \int \int \bar{h}(w, x, a, b) f_{A_1}(a) f_{B_2}(b) \text{d}a \text{d}b,$$

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

Average Structural Function (continued)

Differences of the form

$$m^{\text{ASF}}(w', x') - m^{\text{ASF}}(w, x)$$

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

The double difference

$$m^{\text{ASF}}(1, 1) - m^{\text{ASF}}(0, 1) - [m^{\text{ASF}}(1, 0) - m^{\text{ASF}}(0, 0)] \quad (4)$$

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:

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

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

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

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 .

Identification (Strict Exogeneity)

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

$$V_{ij} \perp (W_i, X_j) \mid A_i = a, B_j = b, a \in \mathbb{A}, b \in \mathbb{B}. \quad (6)$$

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

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

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 $\{(V_{ij}, V_{ji})\}_{i < j}$

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

Relationship to Panel Data (continued)

Equation (7) implies, for example, that

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

“Within-tetrad” variation identifies δ .

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:

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

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

strict exogeneity \Rightarrow second =;

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

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

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

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

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

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:

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

Main Identification Result

Equation (11) gives the identification result

$$\mathbb{E}_R \left[\mathbb{E}_S \left[q \left(w, x, R_i, S_j \right) \right] \right] = \int_r \int_s \left[\int_a \int_b \bar{h} (w, x, a, b) f_{A|R} (a | r) f_{B|S} (b | s) da db \right] f_R (r) f_S (s) dr ds \quad (12)$$

$$\begin{aligned} &= \int_a \int_b \bar{h} (w, x, a, b) f_A (a) f_B (b) da db \\ &= \mathbb{E} [\bar{Y}_{12} (w, x)] \\ &= m^{\text{ASF}} (w, x) . \end{aligned}$$

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:

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

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 $p_w(r) \stackrel{def}{=} \Pr(W_i = w | R_i = r)$ and $p_x(s) \stackrel{def}{=} \Pr(X_i = x | S_i = s)$.

Main Identification Result (Summary)

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

$$m^{\text{ASF}}(w, x) = \int \int q(w, x, r, s) f_R(r) f_S(s) dr ds. \quad (14)$$

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 $Q_i = (W'_i, X'_i, R'_i, S'_i)'$ 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 γ using the Poisson dyadic regression estimator described earlier.

Estimation (continued)

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

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

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 γ in hand, form the fitted values $\{q(w, x, R_i, S_j; \hat{\gamma})\}_{i < j}$ and, invoking Theorem 2, compute the analog estimate

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

Two step procedure:

1. Estimate γ_0 by dyadic regression;
2. Compute $\hat{m}^{\text{ASF}}(w, x; \hat{\gamma})$

(U-Statistic with estimated nuisance parameter).

Estimation (continued)

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

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

where

$$\begin{aligned}\psi_0(w, x, R_1, S_1; \gamma) &= \frac{q^e(w, x, R_1; \gamma) + q^a(w, x, S_1; \gamma_0)}{2} - m^{\text{ASF}}(w, x; \gamma) \\ M_0(w, x) &= \frac{1}{2} \mathbb{E} \left[\frac{\partial q(w, x, R_1, S_2; \gamma_0)}{\partial \gamma'} + \frac{\partial q(w, x, R_2, S_1; \gamma_0)}{\partial \gamma'} \right]\end{aligned}$$

Influence Function

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

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

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

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

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

$$\sqrt{N} \left(\hat{m}^{\text{ASF}}(w, x; \hat{\gamma}) - m^{\text{ASF}}(w, x; \gamma_0) \right) \xrightarrow{D} \mathcal{N} \left(0, 4\Xi_0(w, x) + 4M_0(w, x) \left(\Gamma'_0 \Sigma_1^{-1} \Gamma_0 \right)^{-1} M_0(w, x)' \right)$$

with

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

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 $(R'_i, S'_i)'$.

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.