Further Supplemental Note to "Estimating Local Interactions Among Many Agents Who Observe Their Neighbors"

Nathan Canen, Jacob Schwartz, and Kyungchul Song
University of Houston, University of Haifa, and University of British Columbia

This note consists of five sections. In Section 1, we formally present a model of information sharing over time among many agents where the econometrician is interested in the estimation of local interactions in a particular decision problem. We explain how this extended model maps to the static model of the main paper. Section 2 explains inference based on the model with first order sophisticated agents. The secion also provides the proof of the asymptotic validity of the proposed inference, and results from Monte Carlo simulation studies. Section 3 provides details on the model selection procedure between different games Γ_0 and Γ_1. Section 4 presents a proposal on testing for information sharing on unobservables. Section 5 gives the results from the empirical application using the game with first-order sophisticated agents.

1. Information Sharing Among Many Agents Over Time

Information sharing among people takes place over time, and the econometrician usually observes part of these people as a snapshot in the process. The process involves information sharing, network formation and decision making. Agents can form a network and share information for various purposes. There is no reason to believe that each agent's particular decision problem which is of interest to the econometrician is the single ultimate concern of the agents when they form a network at an earlier time.[1] In this section, we provide an extended model of information sharing which fits the main set-up of the paper. The main idea is that people receive signals, form networks, and share information with their neighbors repeatedly. Then there is a decision making stage. The

Date: February 10, 2020.

[1]For example, it is highly implausible to assume that friendship formation among the students is done for the sole purpose of achieving maximal performance in a math exam observed by the econometrician.

network formation (of either a payoff graph or an information graph) can be made, if not exclusively, in anticipation of the decision making later. However, as we will explain later in detail, our model assumes that when the agents form an information and a payoff graph, they do not observe other agents' payoff relevant signals that are not observed by the econometrician. This is the precise sense in which the information and the payoff graph are exogenously formed.

Let us present a formal model of information sharing over time. Let N be the set of a finite yet large number of players who share their type information over time recursively, where at each stage, players go through three steps sequentially: information graph formation, type realization and information sharing. Then at the final stage, players make a decision, maximizing their expected utilities.

At Stage 0, each agent $i \in N$ is endowed with signal $\mathcal{C}_{i,0}$. The signals can be correlated across agents in an arbitrary way. Then information sharing among agents happens recursively over time as follows starting from Stage $s = 1$.

Stage s-1: (INFORMATION GRAPH FORMATION) Each player $i \in N$ receives signal $\mathcal{C}_{i,s-1}$. Using these signals, the players in N form an information sharing network $G_{I,s-1} = (N, E_{I,s-1})$ among themselves, where $G_{I,s-1} = (N, E_{I,s-1})$ is a directed graph on N.[2]

(TYPE REALIZATION) Each player $i \in N$ is given his type vector $(\tau_{i,s-1}', \eta_{i,s-1})'$, where $\eta_{i,s-1}$ is a private type which player i keeps to himself and $\tau_{i,s-1}$ a sharable type which is potentially observed by other agents.

(INFORMATION SHARING) Each player i observes the sharable type $\tau_{j,s-1}$ of each player j in his neighborhood in the information sharing network.

Stage s: (INFORMATION GRAPH FORMATION) Each player $i \in N$ receives signal $\mathcal{C}_{i,s}$ which contains part of $\mathcal{C}_{i,s-1}$ and part of the information about $G_{I,s-1}$ and $\tau_{j,s-1}$ with $j \in N_{I,s-1}(i)$. Using these signals, the players in N form an information sharing network $G_{I,s} = (N, E_{I,s})$ among themselves, where $G_{I,s} = (N, E_{I,s})$ is a directed graph on N, and receives signal $\mathcal{C}_{i,s}$.

(TYPE REALIZATION) Each player $i \in N$ is given his type vector $(\tau_{i,s}', \eta_{i,s})'$, where $\eta_{i,s}$ is a private type which player i keeps to himself and $\tau_{i,s}$ a sharable type which is potentially observed by other agents.

(INFORMATION SHARING) Each player i observes the sharable type $\tau_{j,s}$ of each player j in his neighborhood in the information sharing network $G_{I,s}$.

[2] The formation of an information sharing network is tantamount to each agent making a (unilateral) binary decision to share his type information with others. Details of this decision making process are not of focus in the empirical model and hence are not elaborated further. What suffices for us is that the strategy of each agent is measurable with respect to the information the agent has. This latter condition is satisfied typically when the agent chooses a pure strategy given his information.

The information sharing activities proceed up to Stage $S-1$. Then each agent faces a decision making problem.

Decision Stage: (PAYOFF GRAPH FORMATION) Each player $i \in N$ receives signal $\mathcal{C}_{i,S}$ which contains part of $\mathcal{C}_{i,S-1}$ and part of information about $G_{I,S-1}$ and $\tau_{j,S-1}$ with $j \in N_{I,S-1}(i)$. Using these signals, the players in N form a payoff graph $G_P = (N, E_P)$ among themselves, where $G_{I,S} = (N, E_{I,S})$ is a directed graph on N, and receives signal $\mathcal{C}_{i,S}$.

(TYPE REALIZATION) Each player $i \in N$ is given his type vector $(\tau'_{i,S}, \eta_{i,S})'$, where $\eta_{i,S}$ is a private type which player i keeps to himself and $\tau_{i,S}$ a sharable type which is potentially observed by other agents.

(INFORMATION SHARING) Each player i observes the sharable type $\tau_{j,S}$ of each player j in his neighborhood in the information sharing network $G_{I,S}$.

(DECISION) Each player i makes a decision which maximizes his expected utility given his beliefs about other agents' strategies using the information accumulated so far.

Now let us consider how this model of information sharing over time maps to our static local interactions model in our main paper. Our local interactions model captures the state where each player in N faces the Decision Stage as follows. The information graph G_I corresponds to $G_{I,S}$ and the payoff graph G_P as described in Decision Stage above. The type vector τ_i and η_i for each player $i \in N$ correspond to $\tau_{i,S}$ and $\eta_{i,S}$ as above. The signal $\mathcal{C}_i$ for each agent i is defined to be

$$\mathcal{C}_i = \bigvee_{s=1}^{S} \mathcal{C}_{i,s},$$

where $\mathcal{A}_1 \vee \mathcal{A}_2$ denotes the smallest σ-field that contains both $\mathcal{A}_1$ and $\mathcal{A}_2$. Therefore, each agent i's signal $\mathcal{C}_i$ contains the information that has been accumulated so far. This information contains past information sharing experiences that have happened over time.

Observe that here we deliberately separate the information sharing stage and the decision stage. This is because people typically share information without necessarily anticipating the particular decision problem that the econometrician happens to later investigate.

2. Inference for the Model with First Order Sophisticated Agents

2.1. Overview

Let us consider inference on payoff parameters using a model that assumes all the agents to be of first-order sophisticated type. The network externality is more extensive than when the agents are of simple type, and best responses involve more extensive network externality, and we require more data accordingly. In particular, we strengthen Conditions B and C as follows:

Condition B1: For each $i \in N^*$, the econometrician observes $N_{P,2}(i)$ and (Y_i, X_i) and for any $j \in N_P(i)$ and any $k \in N_{P,2}(i) \backslash N_P(i)$, the econometrician observes $n_P(j)$, $n_P(k)$, $|N_P(i) \cap N_P(j)|$ and $|N_P(j) \cap N_P(k)|$, X_j and X_k.

Condition C1: Either of the following two conditions is satisfied.

(a) For any $i, j \in N^*$ such that $i \neq j$, $N_{P,2}(i) \cap N_{P,2}(j) = \varnothing$.

(b) For each agent $i \in N^*$, and for any agent $j \in N^*$ such that $N_{P,2}(i) \cap N_{P,2}(j) \neq \varnothing$, the econometrician observes Y_j, $|N_{P,2}(j) \cap N_{P,2}(k)|$, $n_P(k)$ and X_k for all $k \in N_P(j)$.

Condition B1 requires that the data contain many agents such that $N_{P,2}(i)$ for each i of such agents is available together with the number of common G_P-neighbors between each agent $k \in N_{P,2}(i)$ and agent i and between each agent $j \in N_P(i)$ and agent i. Condition C1 is again trivially satisfied if data contain many agents such that G_P-neighbors of G_P-neighbors do not overlap. In this case, we can select N^* to include only those agents.

The inference is similar as in the case with agents of simple type, except that we redefine Z_i and v_i into Z_i^{FS}, and v_i^{FS} as we explain below. Define

$$
\begin{aligned}
(2.1) \qquad Z_i^{\mathsf{FS}} &= \left(1 + \frac{\beta_0}{n_P(i)} \sum_{j \in N_P(i)} w_{ji}^{[0]}\right) X_i \\
&\quad + \sum_{j \in N_{P,2}(i)} \left(\frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} w_{kj}^{[0]} 1\{j \in \overline{N}_P(k)\}\right) X_j.
\end{aligned}
$$

Then, by the previous results (see (2.10) of Canen, Schwartz, and Song (2019)), we can write

$$(2.2) \qquad Y_i = Z_i^{\mathsf{FS}\prime} \rho_0 + v_i^{\mathsf{FS}},$$

where

$$v_i^{\mathsf{FS}} = R_i^{\mathsf{FS}}(\varepsilon) + \eta_i,$$

with

$$
\begin{aligned}
R_i^{\mathsf{FS}}(\varepsilon) &= \left(1 + \frac{\beta_0}{n_P(i)} \sum_{j \in N_P(i)} w_{ji}^{[0]}\right) \varepsilon_i \\
&\quad + \sum_{j \in N_{P,2}(i)} \left(\frac{\beta_0}{n_P(i)} \sum_{k \in N_P(i)} w_{kj}^{[0]} 1\{j \in \overline{N}_P(k)\}\right) \varepsilon_j.
\end{aligned}
$$

Using this reformulation, we can develop inference similarly as before. More specifically, let us define

$$
\Lambda^{\mathsf{FS}} = \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N^*} \mathbf{E}[v_i^{\mathsf{FS}} v_j^{\mathsf{FS}} | \mathcal{F}] \tilde{\varphi}_i \tilde{\varphi}_j', \tag{2.3}
$$

and let $\hat{\Lambda}^{\mathsf{FS}}$ be a consistent estimator of Λ^{FS}. (See the next subsection of the construction of the estimator.) Define

$$
\hat{\rho}^{\mathsf{FS}} = \left[S_{Z\tilde{\varphi}}^{\mathsf{FS}} (\hat{\Lambda}^{\mathsf{FS}})^{-1} (S_{Z\tilde{\varphi}}^{\mathsf{FS}})'\right]^{-1} S_{Z\tilde{\varphi}}^{\mathsf{FS}} (\hat{\Lambda}^{\mathsf{FS}})^{-1} S_{\tilde{\varphi} y}^{\mathsf{FS}}, \tag{2.4}
$$

where $S_{Z\tilde{\varphi}}^{\mathsf{FS}}$ and $S_{\tilde{\varphi} y}^{\mathsf{FS}}$ are the same as $S_{Z\tilde{\varphi}}^{\mathsf{FS}}$ and $S_{\tilde{\varphi} y}^{\mathsf{FS}}$ except that we use Z^{FS} in place of Z. Using this, we construct the estimator

$$
\hat{V}^{\mathsf{FS}} = \left[S_{Z\tilde{\varphi}}^{\mathsf{FS}} (\hat{\Lambda}^{\mathsf{FS}})^{-1} S_{\tilde{\varphi} Z}^{\mathsf{FS}}\right]^{-1}. \tag{2.5}
$$

We construct a vector of residuals $\hat{v}^{\mathsf{FS}} = [\hat{v}_i^{\mathsf{FS}}]_{i \in N^*}$, where

$$
\hat{v}_i^{\mathsf{FS}} = Y_i - Z_i^{\mathsf{FS}\prime} \hat{\rho}^{\mathsf{FS}}. \tag{2.6}
$$

Finally, we form a profiled test statistic as follows:

$$
T^{\mathsf{FS}}(\beta_0) = \frac{(\hat{v}^{\mathsf{FS}})' \tilde{\varphi} (\hat{\Lambda}^{\mathsf{FS}})^{-1} \tilde{\varphi}' \hat{v}^{\mathsf{FS}}}{n^*}. \tag{2.7}
$$

Then, we construct confidence intervals

$$
C_{1-\alpha}^{\beta,\mathsf{FS}} \equiv \left\{\beta \in (-1,1) : T^{\mathsf{FS}}(\beta) \leq c_{1-\alpha}\right\},
$$

where $c_{1-\alpha}$ is the $(1-\alpha)$-quantile of χ_{M-d}^2.

The confidence intervals for $a'\rho$ can be similarly constructed as in Section 3.1.4 of Canen, Schwartz, and Song (2019). More specifically, let

$$
\hat{V}^{\mathsf{FS}} = \left[S_{Z\tilde{\varphi}}^{\mathsf{FS}} (\hat{\Lambda}^{\mathsf{FS}})^{-1} S_{Z\tilde{\varphi}}^{\mathsf{FS}\prime}\right]^{-1},
$$

and define

$$
\hat{\sigma}^{\mathsf{FS}}(a) = \sqrt{a' \hat{V}^{\mathsf{FS}} a}.
$$

Then, the confidence interval for $a'\rho$ is found as

$$C_{1-\alpha}^{\rho,\mathsf{FS}}(a) = \bigcup_{\beta \in C_{1-(\alpha/2)}^{\rho,\mathsf{FS}}} C_{1-(\alpha/2)}^{\beta,\mathsf{FS}}(\beta, a),$$

where

$$C_{1-(\alpha/2)}^{\rho,\mathsf{FS}}(\beta_0, a) = \left[a'\hat{\rho}^{\mathsf{FS}} - \frac{z_{1-(\alpha/4)}\hat{\sigma}^{\mathsf{FS}}(a)}{\sqrt{n}}, a'\hat{\rho}^{\mathsf{FS}} + \frac{z_{1-(\alpha/4)}\hat{\sigma}^{\mathsf{FS}}(a)}{\sqrt{n}} \right].$$

and $z_{1-(\alpha/4)}$ is the $(1-(\alpha/4))$-percentile of $N(0,1)$.

2.2. Estimation of the Asymptotic Covariance Matrix

We first construct a consistent estimator $\hat{\Lambda}^{\mathsf{FS}}$ of $\hat{\Lambda}$. Define for $i, j \in N$,

$$e_{ij}^{\mathsf{FS}} = \mathbf{E}[R_i^{\mathsf{FS}}(\varepsilon) R_j^{\mathsf{FS}}(\varepsilon) | \mathcal{F}] / \sigma_\varepsilon^2.$$

If we let

$$\overline{w}_{ij}^{[0]} = \frac{1}{n_P(i)} \sum_{k \in N_P(i)} w_{kj}^{[0]} 1\{j \in \overline{N}_P(k)\}, \tag{2.8}$$

we can rewrite

$$e_{ii}^{\mathsf{FS}} = \left(1 + \beta_0 \overline{w}_{ii}^{[0]}\right)^2 + \beta_0^2 \sum_{j \in N_{P,2}(i)} \left(\overline{w}_{ij}^{[0]}\right)^2, \tag{2.9}$$

and for $i \neq j$, $e_{ij}^{\mathsf{FS}} = \beta_0 q_{\varepsilon,ij}^{\mathsf{FS}}$, where

$$\begin{aligned} q_{\varepsilon,ij}^{\mathsf{FS}} &= \overline{w}_{ji}^{[0]} \left(1 + \beta_0 \overline{w}_{ii}^{[0]}\right) 1\{i \in N_{P,2}(j)\} + \overline{w}_{ij}^{[0]} \left(1 + \beta_0 \overline{w}_{jj}^{[0]}\right) 1\{j \in N_{P,2}(i)\} \\ &\quad + \beta_0 \sum_{s \in N_{P,2}(i) \cap N_{P,2}(j)} \overline{w}_{is}^{[0]} \overline{w}_{js}^{[0]}, \end{aligned} \tag{2.10}$$

where the last term is zero if $N_{P,2}(i) \cap N_{P,2}(j)$ is empty. Similarly, sums over empty sets in any of the terms above are zero. Let us now write

$$\Lambda^{\mathsf{FS}} = \Lambda_1^{\mathsf{FS}} + \Lambda_2^{\mathsf{FS}},$$

where

$$\begin{aligned} \Lambda_1^{\mathsf{FS}} &= \frac{1}{n^*} \sum_{i \in N^*} \mathbf{E}[(v_i^{\mathsf{FS}})^2 | \mathcal{F}] \tilde{\varphi}_i \tilde{\varphi}_i', \text{ and} \\ \Lambda_2^{\mathsf{FS}} &= \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in N_{-i}^*} \mathbf{E}[v_i^{\mathsf{FS}} v_j^{\mathsf{FS}} | \mathcal{F}] \tilde{\varphi}_i \tilde{\varphi}_j'. \end{aligned}$$

To motivate estimation of Λ_2^{FS}, we rewrite

$$\Lambda_2^{\mathsf{FS}} = \frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N^*_{-i}} (e_{ij}^{\mathsf{FS}})\sigma_\varepsilon^2 \tilde{\varphi}_i \tilde{\varphi}_j' = \frac{\beta_0}{n^*}\sum_{i\in N^*}\sum_{j\in N^*_{-i}} q_{\varepsilon,ij}^{\mathsf{FS}} \sigma_\varepsilon^2 \tilde{\varphi}_i \tilde{\varphi}_j'. \tag{2.11}$$

Let us find an expression for σ_ε^2. Note that

$$\frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*} \mathbf{E}[v_i^{\mathsf{FS}} v_j^{\mathsf{FS}} | \mathcal{F}] = \beta_0 b_\varepsilon^{\mathsf{FS}} \sigma_\varepsilon^2,$$

where

$$b_\varepsilon^{\mathsf{FS}} = \frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*} q_{\varepsilon,ij}^{\mathsf{FS}}.$$

Hence if we let

$$s_\varepsilon^{\mathsf{FS}} = \frac{\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*} \mathbf{E}[v_i^{\mathsf{FS}} v_j^{\mathsf{FS}} | \mathcal{F}]}{\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*} q_{\varepsilon,ij}^{\mathsf{FS}}},$$

we have

$$\sigma_\varepsilon^2 \beta_0 = s_\varepsilon^{\mathsf{FS}}.$$

Plugging this in the last term in (2.11), we obtain that

$$\Lambda_2^{\mathsf{FS}} = \frac{s_\varepsilon^{\mathsf{FS}}}{n^*}\sum_{i\in N^*}\sum_{j\in N^*_{-i}} q_{\varepsilon,ij}^{\mathsf{FS}} \tilde{\varphi}_i \tilde{\varphi}_j'.$$

Our estimator then uses the empirical analogues to find $\hat{\Lambda}^{\mathsf{FS}}$.

First define

$$\tilde{\rho}^{\mathsf{FS}} = \left[(S_{Z\tilde{\varphi}}^{\mathsf{FS}})(S_{Z\tilde{\varphi}}^{\mathsf{FS}})'\right]^{-1} S_{Z\tilde{\varphi}}^{\mathsf{FS}} S_{\tilde{\varphi}y}^{\mathsf{FS}}, \tag{2.12}$$

and let

$$\tilde{v}_i^{\mathsf{FS}} = Y_i - Z_i^{\mathsf{FS}\prime} \tilde{\rho}^{\mathsf{FS}}. \tag{2.13}$$

We now present a consistent estimator $\hat{\Lambda}^{\mathsf{FS}}$:

$$\hat{\Lambda}^{\mathsf{FS}} = \hat{\Lambda}_1^{\mathsf{FS}} + \hat{\Lambda}_2^{\mathsf{FS}},$$

where

$$\hat{\Lambda}_1^{\mathsf{FS}} = \frac{1}{n^*} \sum_{i \in N^*} (\tilde{v}_i^{\mathsf{FS}})^2 \tilde{\varphi}_i \tilde{\varphi}_i', \text{ and}$$
$$\hat{\Lambda}_2^{\mathsf{FS}} = \frac{\hat{s}_\varepsilon^{\mathsf{FS}}}{n^*} \sum_{i \in N^*} \sum_{j \in N^*_{-i} : N_{P,2}(i) \cap N_{P,2}(j) \neq \varnothing} q_{\varepsilon,ij}^{\mathsf{FS}} \tilde{\varphi}_i \tilde{\varphi}_j',$$

and

$$\hat{s}_\varepsilon^{\mathsf{FS}} = \frac{\sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} \tilde{v}_i^{\mathsf{FS}} \tilde{v}_j^{\mathsf{FS}}}{\sum_{i \in N^*} \sum_{j \in N_P(i) \cap N^*} q_{\varepsilon,ij}^{\mathsf{FS}}},$$

with $q_{\varepsilon,ij}^{\mathsf{FS}}$ as defined in (2.10). For this, we construct $\tilde{v}_i^{\mathsf{FS}}$ as we constructed $\tilde{v}_i$ using Z^{FS} in place of Z.

2.3. Asymptotic Theory for Inference from the Model with First-Order Sophisticated Agents

In this section, we develop asymptotic theory for the game with first-order sophisticated agents. Recall that each player i's best response $s_i^{[1]}$ takes the following form: (recall the definition of $\overline{w}_{ij}^{[0]}$ in (2.8))

$$s_i^{[1]}(\mathcal{I}_{i,1}) = \left(1 + \beta_0 \overline{w}_{ii}^{[0]}\right) X_i' \rho_0 + \beta_0 \sum_{j \in N_{P,2}(i)} \overline{w}_{ij}^{[0]} X_j' \rho_0 + R_i^{\mathsf{FS}}(\varepsilon) + \eta_i,$$

where,

$$R_i^{\mathsf{FS}}(\varepsilon) = \left(1 + \beta_0 \overline{w}_{ii}^{[0]}\right) \varepsilon_i + \beta_0 \sum_{j \in N_{P,2}(i)} \overline{w}_{ij}^{[0]} \varepsilon_j.$$

We make the following assumptions.

Assumption 2.1. There exists $c > 0$ such that for all $n^* \geq 1$,

$$\lambda_{\min}(S_{\varphi\varphi}) \geq c, \lambda_{\min}(\Lambda^{\mathsf{FS}}) \geq c,$$
$$\lambda_{\min}((S_{Z\tilde{\varphi}}^{\mathsf{FS}})(S_{Z\tilde{\varphi}}^{\mathsf{FS}})') \geq c, \text{ and}$$
$$\lambda_{\min}((S_{Z\tilde{\varphi}}^{\mathsf{FS}})(\Lambda^{\mathsf{FS}})^{-1}(S_{Z\tilde{\varphi}}^{\mathsf{FS}})') \geq c.$$

Assumption 2.2. There exists a constant $C > 0$ such that for all $n^* \geq 1$,

$$\max_{i \in N_2^\circ} ||X_i|| + \max_{i \in N_2^\circ} ||\tilde{\varphi}_i|| \leq C$$

and $\mathbf{E}[\varepsilon_i^4|\mathcal{F}] + \mathbf{E}[\eta_i^4|\mathcal{F}] < C$, where $n_2^\circ = |N_2^\circ|$ and

$$N_2^\circ = \bigcup_{i \in N^*} \overline{N}_{P,2}(i).$$

Then the asymptotic results are summarized in the following theorem.

Theorem 2.1. *Suppose that the conditions of Theorem 2.2 and Assumption 3.5 in Canen, Schwartz, and Song (2019), and Assumptions 2.1 - 2.2 hold. Then,*

$$T^{\mathsf{FS}}(\beta_0) \to_d \chi^2_{M-d}, \text{ and } (\hat{V}^{\mathsf{FS}})^{-1/2}\sqrt{n^*}(\hat{\rho}^{\mathsf{FS}} - \rho_0) \to_d N(0, I_d),$$

as $n^* \to \infty$.

The proofs follow similar steps as in the proof of Theorem 3.2 of Canen, Schwartz, and Song (2019). For the sake of transparency, we provide complete proofs here.

Lemma 2.1. *Suppose that the conditions of Theorem 2.1 hold. Then, as* $n^* \to \infty$,

$$(\Lambda^{\mathsf{FS}})^{-1/2} \frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \tilde{\varphi}_i v_i^{\mathsf{FS}} \to_d N(0, I_M).$$

Proof: Choose any vector $b \in \mathbf{R}^M$ such that $||b|| = 1$ and let $\tilde{\varphi}_{i,b} = b'\tilde{\varphi}_i$. Define

$$a_i^{\mathsf{FS}} = \left(1 + \beta_0 \overline{w}_{ii}^{[0]}\right) \tilde{\varphi}_{i,b} 1\{i \in N^*\} + \beta_0 \sum_{j \in N_{P,2}(i) \cap N^*} \tilde{\varphi}_{j,b} \overline{w}_{ji}^{[0]}.$$

By (B.1) of Canen, Schwartz, and Song (2019), we have

$$0 \le \overline{w}_{ii}^{[0]} \le 1 + \frac{\beta_0^2}{1 - \beta_0^2}, \text{ and } \left|\overline{w}_{ij}^{[0]}\right| \le \frac{|\beta_0|}{n_P(i)(1 - |\beta_0|)} \left(1 + \frac{\beta_0^2}{1 - \beta_0^2}\right). \tag{2.14}$$

Then we can write

$$\frac{1}{\sqrt{n^*}} \sum_{i \in N^*} \tilde{\varphi}_{i,b} v_i^{\mathsf{FS}} = \sum_{i \in N_2^\circ} \xi_i^{\mathsf{FS}}, \tag{2.15}$$

where $\xi_i^{\mathsf{FS}} = (a_i^{\mathsf{FS}} \varepsilon_i + \tilde{\varphi}_{i,b} \eta_i 1\{i \in N^*\})/\sqrt{n^*}$. By the Berry-Esseen Lemma (e.g., Shorack (2000), p.259),

$$\sup_{t \in \mathbf{R}} \left| P\left\{ \sum_{i \in N_2^\circ} \frac{\xi_i^{\mathsf{FS}}}{\sigma_{\xi,i}^{\mathsf{FS}}} \le t | \mathcal{F} \right\} - \Phi(t) \right| \le \frac{9\mathbf{E}\left[\sum_{i \in N_2^\circ} |\xi_i^{\mathsf{FS}}|^3 | \mathcal{F}\right]}{\left(\sum_{i \in N_2^\circ} (\sigma_{\xi,i}^{\mathsf{FS}})^2\right)^{3/2}}, \tag{2.16}$$

where $(\sigma_{\xi,i}^{\mathsf{FS}})^2 = \mathrm{Var}(\xi_i^{\mathsf{FS}}|\mathcal{F})$. It suffices to show that the last bound vanishes in probability as $n^* \to \infty$. Again, since ε_i's and η_i's are independent,

$$\sum_{i \in N_2^\circ} \sigma_{\xi,i}^2 \geq \sigma_\eta^2 > 0.$$

Observe that as in (D.20) of Canen, Schwartz, and Song (2019), for some constant $C_1 > 0$,

$$\mathbf{E}\left[\sum_{i \in N_2^\circ} |\xi_i^{\mathsf{FS}}|^3 |\mathcal{F}\right] \leq \frac{C_1 \max_{i \in N} \mathbf{E}[|\varepsilon_i|^3|\mathcal{F}]}{(n^*)^{3/2}} \sum_{i \in N_2^\circ} |a_i^{\mathsf{FS}}|^3 + \frac{C_1 n_2^\circ \max_{i \in N} \mathbf{E}[|\eta_i|^3|\mathcal{F}]}{(n^*)^{3/2}}.$$

Now, as for the leading term, note that by (2.14), Assumption 2.2 of Canen, Schwartz, and Song (2019) and the assumption that $|\tilde{\varphi}_{i,b}| < C$ for all $i \in N^*$ for some $C > 0$, we have for some constant $C_4 > 0$ that does not depend on n,

$$\frac{1}{n^*} \sum_{i \in N_2^\circ} |a_i^{\mathsf{FS}}|^3 \leq C.$$

Hence, we find that for some $C_1 > 0$,

$$\mathbf{E}\left[\sum_{i \in N^*} |\xi_i^{\mathsf{FS}}|^3 |\mathcal{F}\right] \leq \frac{C_1}{\sqrt{n^*}} \max_{i \in N} \mathbf{E}[|\varepsilon_i|^3|\mathcal{F}] + \frac{C_1 n_2^\circ}{(n^*)^{3/2}} \max_{i \in N} \mathbf{E}[|\eta_i|^3|\mathcal{F}].$$

Since $n_2^\circ \leq C n^*$, we obtain the desired result. ■

Lemma 2.2. *Suppose that the conditions of Theorem 2.1 hold. Then,*

$$||S_{\tilde{\varphi}v}^{\mathsf{FS}}||^2 \leq \frac{C}{n^*},$$

for some constant C that does not depend on n.

Proof: Recall the definitions of e_{ii}^{FS} and e_{ij}^{FS} in (2.9) and below. Note that

$$(2.17) \qquad ||\Lambda^{\mathsf{FS}}|| \leq \frac{\sigma_\varepsilon^2}{n^*} \sum_{i \in N^*} \sum_{j \in N_{-i}^*: N_{P,2}(i) \cap N_{P,2}(j) \neq \varnothing} |e_{ij}^{\mathsf{FS}}| ||\tilde{\varphi}_i|| ||\tilde{\varphi}_j|| + \frac{1}{n^*} \sum_{i \in N^*} (|e_{ii}^{\mathsf{FS}}| \sigma_\varepsilon^2 + \sigma_\eta^2) ||\tilde{\varphi}_i||^2.$$

By (2.14), we have

$$(2.18) \qquad \max_{i \in N^*} |e_{ii}^{\mathsf{FS}}| \leq C, \text{ and } \max_{i,j \in N^*: i \neq j} |e_{ij}^{\mathsf{FS}}| \leq C,$$

for constant $C > 0$. Thus, we find that $|e_{ij}^{\mathsf{FS}}| \leq C$. Therefore, both terms on the right hand side of (2.17) is bounded by C/n^*. ■

Recall the definition of $\tilde{v}_i^{\mathsf{FS}}$ in (2.13).

Lemma 2.3. *Suppose that the conditions of Theorem 2.1 hold. Then the following holds.*

(i) $\frac{1}{n^*}\sum_{i \in N^*}((\tilde{v}_i^{\mathsf{FS}})^2 - (v_i^{\mathsf{FS}})^2)\tilde{\varphi}_i\tilde{\varphi}_i' = O_P(1/\sqrt{n^*})$.

(ii) $\frac{1}{n^*}\sum_{i \in N^*}\sum_{j \in N_P(i)\cap N^*}(\tilde{v}_i^{\mathsf{FS}}\tilde{v}_j^{\mathsf{FS}} - v_i^{\mathsf{FS}}v_j^{\mathsf{FS}})\tilde{\varphi}_i\tilde{\varphi}_j' = O_P(1/n^*)$.

(iii) $\frac{1}{n^*}\sum_{i \in N^*}((v_i^{\mathsf{FS}})^2 - \mathbf{E}[(v_i^{\mathsf{FS}})^2|\mathcal{F}])\tilde{\varphi}_i\tilde{\varphi}_i' = O_P(1/\sqrt{n^*})$.

(iv) $\frac{1}{n^*}\sum_{i \in N^*}\sum_{j \in N_P(i)\cap N^*}(v_i^{\mathsf{FS}}v_j^{\mathsf{FS}} - \mathbf{E}[v_i^{\mathsf{FS}}v_j^{\mathsf{FS}}|\mathcal{F}])\tilde{\varphi}_i\tilde{\varphi}_j' = O_P(1/\sqrt{n^*})$.

Proof: (i) Note that

$$\left\|\frac{1}{n^*}\sum_{i \in N^*}(\tilde{v}_i^{\mathsf{FS}} - v_i^{\mathsf{FS}})^2\tilde{\varphi}_i\tilde{\varphi}_i'\right\| \le \frac{C}{n^*}\sum_{i \in N^*}(\tilde{v}_i^{\mathsf{FS}} - v_i^{\mathsf{FS}})^2,$$

for some constant $C > 0$. As for the last term, note that for some constant $C > 0$,

$$\frac{1}{n^*}\sum_{i \in N^*}\mathbf{E}\left[(\tilde{v}_i^{\mathsf{FS}} - v_i^{\mathsf{FS}})^2|\mathcal{F}\right] \le \frac{C}{n^*}\mathrm{tr}(\Lambda^{\mathsf{FS}}) = O_P\left(\frac{1}{n^*}\right), \tag{2.19}$$

by Assumption 2.1 and by Lemma 2.2. However, we need to deal with

$$\left|\frac{1}{n^*}\sum_{i \in N^*}((\tilde{v}_i^{\mathsf{FS}})^2 - (v_i^{\mathsf{FS}})^2)\right| \le \sqrt{\frac{1}{n^*}\sum_{i \in N^*}(\tilde{v}_i^{\mathsf{FS}} - v_i^{\mathsf{FS}})^2}\sqrt{\frac{1}{n^*}\sum_{i \in N^*}(\tilde{v}_i^{\mathsf{FS}} + v_i^{\mathsf{FS}})^2}.$$

Note that

$$\begin{aligned}\frac{1}{n^*}\sum_{i \in N^*}(\tilde{v}_i^{\mathsf{FS}} + v_i^{\mathsf{FS}})^2 &\le \frac{2}{n^*}\sum_{i \in N^*}(\tilde{v}_i^{\mathsf{FS}} - v_i^{\mathsf{FS}})^2 + \frac{8}{n^*}\sum_{i \in N^*}(v_i^{\mathsf{FS}})^2 \\ &= O_P\left(\frac{1}{n^*}\right) + \frac{8}{n^*}\sum_{i \in N^*}(v_i^{\mathsf{FS}})^2.\end{aligned}$$

As for the last term,

$$\frac{1}{n^*}\sum_{i \in N^*}\mathbf{E}[(v_i^{\mathsf{FS}})^2|\mathcal{F}] \le \frac{2}{n^*}\sum_{i \in N^*}\mathbf{E}[R_i^{\mathsf{FS}}(\varepsilon)^2|\mathcal{F}] + \frac{2}{n^*}\sum_{i \in N^*}\mathbf{E}[\eta_i^2|\mathcal{F}].$$

The last term is bounded by σ_η^2, and the first term on the right hand side is bounded by

$$\frac{2\sigma_\varepsilon^2}{n^*}\sum_{i \in N^*}e_{ii}^{\mathsf{FS}} \le C,$$

for some constant $C > 0$, by (2.18). Combining this with (2.19), we obtain the desired result.

(ii) Define

$$\begin{aligned} A_{n,1} &= \frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}(\tilde{v}_i^{\mathsf{FS}}-v_i^{\mathsf{FS}})(\tilde{v}_j^{\mathsf{FS}}-v_j^{\mathsf{FS}}) \\ A_{n,2} &= \frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}(\tilde{v}_i^{\mathsf{FS}}-v_i^{\mathsf{FS}})v_j^{\mathsf{FS}}, \text{ and} \\ A_{n,3} &= \frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}v_i^{\mathsf{FS}}(\tilde{v}_j^{\mathsf{FS}}-v_j^{\mathsf{FS}}), \end{aligned}$$

and write

$$\frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}(\tilde{v}_i^{\mathsf{FS}}\tilde{v}_j^{\mathsf{FS}}-v_i^{\mathsf{FS}}v_j^{\mathsf{FS}}) = A_{n,1}+A_{n,2}+A_{n,3}.$$

As for the leading term, by Cauchy-Schwarz inequality,

$$|A_{n,1}| = \sqrt{\frac{1}{n^*}\sum_{i\in N^*}(\tilde{v}_i^{\mathsf{FS}}-v_i^{\mathsf{FS}})^2}\sqrt{\frac{1}{n^*}\sum_{i\in N^*}\left(\sum_{j\in N_P(i)\cap N^*}(\tilde{v}_j^{\mathsf{FS}}-v_j^{\mathsf{FS}})\right)^2}.$$

Note that

$$\begin{aligned} &\frac{1}{n^*}\sum_{i\in N^*}\mathbf{E}\left[\left(\sum_{j\in N_P(i)\cap N^*}(\tilde{v}_j^{\mathsf{FS}}-v_j^{\mathsf{FS}})\right)^2|\mathcal{F}\right] \\ \le\;& \frac{1}{n^*}\sum_{i\in N^*}|N_P(i)\cap N^*|\sum_{j\in N_P(i)\cap N^*}\mathbf{E}\left[\left(\tilde{v}_j^{\mathsf{FS}}-v_j^{\mathsf{FS}}\right)^2|\mathcal{F}\right] \\ =\;& \frac{1}{n^*}\sum_{i\in N^*}\left(\sum_{j\in N_P(i)\cap N^*}|N_P(j)\cap N^*|\right)\mathbf{E}\left[\left(\tilde{v}_i^{\mathsf{FS}}-v_i^{\mathsf{FS}}\right)^2|\mathcal{F}\right]. \end{aligned}$$

Hence the last term is bounded by

$$\frac{\max_{i\in N^*}|N_P(i)\cap N^*|^2}{n^*}\sum_{i\in N^*}\mathbf{E}\left[\left(\tilde{v}_i^{\mathsf{FS}}-v_i^{\mathsf{FS}}\right)^2|\mathcal{F}\right] \le O_P\left(\frac{1}{n^*}\right),$$

by (2.19). Thus we conclude that $|A_{n,1}| = O_P(1/n^*)$.

Similarly, using Cauchy-Schwarz inequality and applying the same arguments, we have

$$|A_{n,2}| = O_P\left(\frac{1}{n^*}\right) \text{ and } |A_{n,3}| = O_P\left(\frac{1}{n^*}\right),$$

obtaining the desired result.

(iii) Note that

$$\text{Var}\left(\frac{1}{n^*}\sum_{i\in N^*} R_i^{\mathsf{FS}}(\varepsilon)|\mathcal{F}\right) \leq \frac{1}{(n^*)^2}\sum_{i\in N^*}\mathbf{E}[(R_i^{\mathsf{FS}}(\varepsilon))^2|\mathcal{F}] = O_P((n^*)^{-1}),$$

from the proof of (i).

(iv) The proof is similar to (iii). Hence we omit the details. ■

Lemma 2.4. *Suppose that the conditions of Theorem 2.1 hold. Then,*

$$\hat{\Lambda}^{\mathsf{FS}} - \Lambda^{\mathsf{FS}} = O_P\left(\frac{1}{\sqrt{n^*}}\right).$$

Proof: We write

$$\begin{aligned}
\hat{\Lambda}_1^{\mathsf{FS}} - \Lambda_1^{\mathsf{FS}} &= \frac{1}{n^*}\sum_{i\in N^*}((\tilde{v}_i^{\mathsf{FS}})^2 - \mathbf{E}[(v_i^{\mathsf{FS}})^2|\mathcal{F}])\tilde{\varphi}_i\tilde{\varphi}_i' \text{ and} \\
\hat{\Lambda}_2^{\mathsf{FS}} - \Lambda_2^{\mathsf{FS}} &= \frac{\hat{s}_\varepsilon^{\mathsf{FS}} - s_\varepsilon^{\mathsf{FS}}}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*} q_{\varepsilon,ij}^{\mathsf{FS}}\tilde{\varphi}_i\tilde{\varphi}_j'.
\end{aligned}$$

Thus the desired result follows from Lemma 2.3. ■

Lemma 2.5. *Suppose that the conditions of Theorem 2.1 hold. Then the following holds.*

(i) $\frac{1}{n^*}\sum_{i\in N^*}((\hat{v}_i^{\mathsf{FS}})^2 - (v_i^{\mathsf{FS}})^2)\tilde{\varphi}_i\tilde{\varphi}_i' = O_P(1/n^*)$.

(ii) $\frac{1}{n^*}\sum_{i\in N^*}\sum_{j\in N_P(i)\cap N^*}(\hat{v}_i^{\mathsf{FS}}\hat{v}_j^{\mathsf{FS}} - v_i^{\mathsf{FS}}v_j^{\mathsf{FS}})\tilde{\varphi}_i\tilde{\varphi}_j' = O_P(1/\sqrt{n^*})$.

Proof: The proof is the same as that of Lemma D.8 of Canen, Schwartz, and Song (2019). ■

Proof of Theorem 2.1: The proof is precisely the same as that of Theorem 3.2 of Canen, Schwartz, and Song (2019) except that we use the above auxiliary lemmas instead. Details are omitted. ■

2.4. Monte Carlo Simulations for Games with the First Order Sophisticated Players

2.4.1. **Simulation Design.** In this section, we investigate the finite sample properties of the inference for first-order sophisticated types across various configurations of the payoff graphs G_P. We generate graphs for the two specifications models and check our inference under different parameters, described in the following paragraph. Specification 1 uses an Erdös-Rényi (random graph formation) payoff graph and Specification 2 uses Barabási-Albert (preferential attachment) graphs seeded with an Erdös-Rényi graph of the smallest integer larger than $5\sqrt{n}$. Some summary statistics of the graphs used for the Monte Carlo study is given in Table 1.

For the simulations, we also set the following:

$$\tau_i = X_i'\rho_0 + \varepsilon_i,$$

where $\rho_0 = (2, 4, 1, 3, 4)'$ and $X_i = (X_{i,1}, \overline{X}_{i,2})'$, and

$$\overline{X}_{i,2} = \frac{1}{n_P(i)} \sum_{j \in N_P(i)} X_{j,2}.$$

We set and a to be a column of ones so that $a_0'\rho = 14$. The variables ε and η are drawn i.i.d. from $N(0,1)$. The first column of $X_{i,1}$ is a column of ones, while remaining columns of $X_{i,1}$ are drawn independently from $N(1,1)$. The columns of $X_{i,2}$ are drawn independently from $N(3,1)$.

For instruments, we consider the following nonlinear transformations of X_1 and X_2:

$$\varphi_i = [\tilde{Z}_{i,1}, X_{i,1}^2, \overline{X}_{i,2}^2, \overline{X}_{i,2}^3]',$$

where we define

$$\tilde{Z}_{i,1} \equiv \frac{1}{n_P(i)} \sum_{j \in N_{P,2}(i)} \lambda_{ij} X_{j,1}.$$

We generate Y_i from the best response function as in (2.2). We used the Monte Carlo simulation number equal to 5000.

2.4.2. **Results.** The results are found in Tables 2-5. In Tables 2 and 3 we report the finite sample coverage probabilities of the confidence intervals for β_0 and for $a'\rho_0$ respectively. For β_0, the coverage probabilities perform very well, whereas for $a'\rho_0$, they are conservative. Overall, for the range of the sample sizes $500 - 1000$, the finite sample properties of the inference procedure seem reasonable.

In Tables 4-5, we report the average length of the confidence intervals. Clearly, as we increase the sample size from 500 to 1000, the length of the confidence intervals tends to shrink substantially. This suggests that accummulation of data leads to increased information and improved accuracy in inference.

TABLE 1. The Average and Maximum Degrees of Graphs in the Simulations

		Specification 1			Specification 2		
n		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
500	d_{mx}	17	21	30	5	8	11
	d_{av}	1.7600	3.2980	4.8340	0.9520	1.9360	2.9600
1000	d_{mx}	18	29	34	6	7	9
	d_{av}	1.8460	3.5240	5.2050	0.9960	1.9620	3.0020

Notes: d_{av} and d_{mx} represent the average and maximum degrees of the networks respectively; that is, $d_{av} \equiv \frac{1}{n}\sum_{i\in N} n_P(i)$ and $d_{mx} \equiv \max_{i\in N} n_P(i)$.

TABLE 2. The Empirical Coverage Probability of Confidence Intervals for β_0 from First-Order Sophisticated Types

		Specification 1			Specification 2		
β_0		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	0.9634	0.9566	0.9562	0.9576	0.9528	0.9586
	$n=1000$	0.9542	0.9552	0.9566	0.9566	0.9526	0.9558
-0.3	$n=500$	0.9586	0.9542	0.9536	0.9526	0.9534	0.9570
	$n=1000$	0.9490	0.9518	0.9546	0.9508	0.9506	0.9542
0	$n=500$	0.9576	0.9532	0.9548	0.9478	0.9530	0.9552
	$n=1000$	0.9502	0.9522	0.9530	0.9462	0.9474	0.9516
0.3	$n=500$	0.9652	0.9606	0.9582	0.9470	0.9514	0.9554
	$n=1000$	0.9548	0.9520	0.9548	0.9454	0.9494	0.9530
0.5	$n=500$	0.9710	0.9658	0.9614	0.9502	0.9528	0.9584
	$n=1000$	0.9600	0.9570	0.9578	0.9474	0.9494	0.9566

Notes: This table shows the empirical coverage probabilities $R = 5000$ of the confidence intervals for β_0 under two models of graph formation. The nominal size is $\alpha = 0.05$. As expected, the coverage probabilities are close to the nominal size.

3. Model Selection between Games Γ_0 and Γ_1

It is a matter of econometric model specification to choose between Γ_0 with simple-type agents or Γ_1 with the first-order sophisticated type agents as an empirical model. Both models are distinct and nonnested. Here we provide an empirical procedure to select among the two models.[3]

First, we write $v_i(\beta_0)$, $v_i^{\mathsf{FS}}(\beta_0)$, and $\tilde{\varphi}_i(\beta_0)$ in place of v_i, v_i^{FS}, and $\tilde{\varphi}_i$ to make their dependence on β_0 explicit. Let B be a set contained in $(-1,1)$ and assumed to contain

[3]Note that the two models Γ_0 and Γ_1 may not be exclusive of each other because there can be a data generating process such that the payoff graph G_P is a cluster structure where each cluster is a complete graph (so that $N_{P,2}(i) = N_P(i)$), or $\beta_0 = 0$.

TABLE 3. The Empirical Coverage Probability of Confidence Interval for $a'\rho_0$ for First-Order Sophisticated Types

		Specification 1			Specification 2		
β_0		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	0.9906	0.9928	0.9900	0.9896	0.9896	0.9910
	$n=1000$	0.9856	0.9872	0.9872	0.9836	0.9846	0.9896
-0.3	$n=500$	0.9874	0.9878	0.9880	0.9830	0.9862	0.9874
	$n=1000$	0.9802	0.9816	0.9862	0.9740	0.9786	0.9852
0	$n=500$	0.9814	0.9848	0.9854	0.9760	0.9808	0.9848
	$n=1000$	0.9714	0.9806	0.9812	0.9568	0.9720	0.9772
0.3	$n=500$	0.9840	0.9856	0.9872	0.9644	0.9796	0.9828
	$n=1000$	0.9710	0.9810	0.9796	0.9488	0.9616	0.9766
0.5	$n=500$	0.9842	0.9880	0.9886	0.9496	0.9750	0.9836
	$n=1000$	0.9650	0.9784	0.9820	0.9456	0.9526	0.9750

Notes: This table shows the empirical coverage probabilities $R = 5000$ of the confidence intervals for $a'\rho_0$ under two models of graph formation. The nominal size is $\alpha = 0.05$. The procedure is conservative, as expected from the Bonferroni procedure.

TABLE 4. The Average Length of Confidence Intervals for β_0 for First-Order Sophisticated Types

		Specification 1			Specification 2		
β_0		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	0.1171	0.2867	0.3663	0.0627	0.1040	0.2124
	$n=1000$	0.0829	0.2271	0.3321	0.0366	0.0614	0.1333
-0.3	$n=500$	0.0924	0.1442	0.1980	0.0658	0.0860	0.1409
	$n=1000$	0.0653	0.0965	0.1299	0.0384	0.0513	0.0831
0	$n=500$	0.0810	0.1041	0.1277	0.0662	0.0726	0.1040
	$n=1000$	0.0545	0.0653	0.0781	0.0386	0.0420	0.0587
0.3	$n=500$	0.0761	0.0972	0.1165	0.0696	0.0767	0.1030
	$n=1000$	0.0504	0.0599	0.0704	0.0415	0.0456	0.0603
0.5	$n=500$	0.0698	0.0773	0.1008	0.0223	0.0386	0.0728
	$n=1000$	0.0198	0.0479	0.0643	0.0068	0.0168	0.0468

Notes: This table shows the average length of confidence intervals for β_0 for two models of graph formation $R = 5000$. The nominal size is $\alpha = 0.05$. As expected the average length of the confidence interval falls with n.

the true parameter β_0, and define

$$T_{\mathsf{ST}} = \inf_{\beta \in B} T(\beta), \text{ and } T_{\mathsf{FS}} = \inf_{\beta \in B} T^{\mathsf{FS}}(\beta),$$

where $T(\beta_0)$ is as defined in (3.8) of Canen, Schwartz, and Song (2019) and $T^{\mathsf{FS}}(\beta_0)$ is similarly defined after (2.3). Then, we consider the set

$$\hat{S} = \{s \in \{\mathsf{ST}, \mathsf{FS}\} : T_s \leq c_{1-\alpha/2}\},$$

TABLE 5. Average Length of of Confidence Intervals for $a'\rho_0$ for First-Order Sophisticated Types

		Specification 1			Specification 2		
β_0		$m=1$	$m=2$	$m=3$	$\lambda=1$	$\lambda=2$	$\lambda=3$
-0.5	$n=500$	1.5489	1.4912	1.6440	0.6448	1.0615	2.1044
	$n=1000$	1.2581	1.1732	1.2974	0.4196	0.6760	1.2460
-0.3	$n=500$	1.9736	2.3093	2.6460	0.8698	1.4064	2.9878
	$n=1000$	1.5288	1.6772	1.8918	0.5571	0.8749	1.6194
0	$n=500$	2.3591	2.7335	3.1419	1.1100	1.6186	3.2359
	$n=1000$	1.6733	1.8155	2.0145	0.6959	0.9779	1.6898
0.3	$n=500$	2.2385	2.6666	3.1282	1.3922	1.8623	3.2965
	$n=1000$	1.6314	1.8048	2.0041	0.8720	1.1455	1.8393
0.5	$n=500$	3.1651	2.1409	2.5747	0.7793	1.2397	2.3673
	$n=1000$	1.0193	1.5059	1.7956	0.3911	0.7031	1.5259

Notes: The true $a'\rho_0$ is equal to 14. The length of confidence intervals tends to be small and substantially shortened as the size of the network increases.

where $c_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-percentile of the distribution of χ^2_{M-d}. Among the two models based on games Γ_0 and Γ_1, the set $\hat{S}$ is the set of the models that are not rejected at $100(1-\alpha)\%$ in the sense to be explained below.

Define

$$\begin{aligned} m_{\mathsf{ST}} &= \inf_{\beta\in B}\left|\frac{1}{n^*}\sum_{i\in N^*}\mathbf{E}\left[v_i(\beta)|\mathcal{F}\right]\tilde{\varphi}_i(\beta)\right|, \text{ and} \\ m_{\mathsf{FS}} &= \inf_{\beta\in B}\left|\frac{1}{n^*}\sum_{i\in N^*}\mathbf{E}\left[v_i^{\mathsf{FS}}(\beta)|\mathcal{F}\right]\tilde{\varphi}_i(\beta)\right|. \end{aligned}$$

Let $S_0 = \{s \in \{\mathsf{ST},\mathsf{FS}\} : m_s = 0\}$. Hence S_0 denotes the collection of true models (as distinguished by the moment condition $m_s = 0, s \in \{\mathsf{ST},\mathsf{FS}\}$.) Let $\mathcal{P}_n$ be the collection of the joint distributions of all the observables in the data. For each $\delta > 0$, let us define

$$\begin{aligned} \mathcal{P}_{n,1}(\delta) &= \{P\in\mathcal{P}_n : m_{\mathsf{ST}} = 0, m_{\mathsf{FS}} > \delta\} \\ \mathcal{P}_{n,2}(\delta) &= \{P\in\mathcal{P}_n : m_{\mathsf{ST}} > \delta, m_{\mathsf{FS}} = 0\} \\ \mathcal{P}_{n,3}(\delta) &= \{P\in\mathcal{P}_n : m_{\mathsf{ST}} = 0, m_{\mathsf{FS}} = 0\}, \text{ and} \\ \mathcal{P}_{n,4}(\delta) &= \{P\in\mathcal{P}_n : m_{\mathsf{ST}} > \delta, m_{\mathsf{FS}} > \delta\}, \end{aligned}$$

and define

$$\mathcal{P}_n(\delta) = \bigcup_{k=1}^{4}\mathcal{P}_{n,k}(\delta).$$

Then, the selection rule $\hat{S}$ can be justified as follows: for each $\delta > 0$,

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n(\delta)} P\{S_0 = \hat{S}\} \geq 1-\alpha.$$

We can also make the selection rule a consistent selection rule, by choosing $\alpha = \alpha_n$ to be a sequence so that $c_{1-\alpha_n/2} \to \infty$ but slowly at a proper rate.

The procedure can be modified to perform model selection with other combinations of the models as long as a testing procedure for moment conditions from each model is available. For example, suppose that $m_{\mathsf{EQ}} = 0$ is a moment condition for a complete information game model with equilibrium strategies and a consistent testing procedure (at level α) for this moment condition is given by $1\{T_{\mathsf{EQ}} > c^{\mathsf{EQ}}_{1-\alpha}\}$ for some test statistic T_{EQ} and critical value $c^{\mathsf{EQ}}_{1-\alpha}$. Then, one can replace T_{FS} and $c_{1-\alpha/2}$ by T_{EQ} and $c^{\mathsf{EQ}}_{1-\alpha/2}$ in the previous procedure to select a set of models from $\{\mathsf{ST}, \mathsf{EQ}\}$ that are not rejected at $100(1-\alpha)\%$.

Let us provide conditions and a brief proof for the asymptotic justification of the model selection procedure. For brevity, we will provide high level conditions and discussions on how they can be verified using low level conditions.

Let us first define for each $\delta > 0$:

$$\begin{aligned}
p_{n,1}(\delta) &= \inf_{P\in P_{n,1}(\delta)} P\left\{T_{\mathsf{ST}} \leq c_{1-\alpha/2}, T_{\mathsf{FS}} > c_{1-\alpha/2}\right\} \\
p_{n,2}(\delta) &= \inf_{P\in P_{n,2}(\delta)} P\left\{T_{\mathsf{ST}} > c_{1-\alpha/2}, T_{\mathsf{FS}} \leq c_{1-\alpha/2}\right\} \\
p_{n,3}(\delta) &= \inf_{P\in P_{n,3}(\delta)} P\left\{T_{\mathsf{ST}} \leq c_{1-\alpha/2}, T_{\mathsf{FS}} \leq c_{1-\alpha/2}\right\}, \text{ and} \\
p_{n,4}(\delta) &= \inf_{P\in P_{n,4}(\delta)} P\left\{T_{\mathsf{ST}} > c_{1-\alpha/2}, T_{\mathsf{FS}} > c_{1-\alpha/2}\right\}.
\end{aligned}$$

Then, we make the following assumption:

$$\begin{aligned}
\min\left\{\liminf_{n\to\infty} p_{n,1}(\delta), \liminf_{n\to\infty} p_{n,2}(\delta)\right\} &\geq 1-\alpha/2, \\
\liminf_{n\to\infty} p_{n,2}(\delta) &\geq 1-\alpha, \\
p_{n,2}(\delta) &\to 1, \text{ as } n\to\infty.
\end{aligned} \tag{3.1}$$

The assumptions in (3.1) follow if the tests $1\{T^{\mathsf{ST}}(\beta_0) > c_{1-\alpha/2}\}$ and $1\{T^{\mathsf{FS}}(\beta_0) > c_{1-\alpha/2}\}$ are asymptotically valid uniformly over the probabilities that satisfy the respective moment conditions $m_s = 0$, and if the tests are consistent under fixed alternatives (i.e., $m_s > \delta$). The uniform validity of the tests can be proved by invoking the uniform boundedness of certain moments and eigenvalues of the variance matrices, and by using Berry-Esseen Lemma. As such arguments are standard, details are omitted here.

Under the assumptions in (3.1), it is not hard to see that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n(\delta)} P\{S_0 = \hat{S}\} \geq 1-\alpha.$$

Indeed, noting that $\mathcal{P}_n(\delta)$ is partitioned into $\mathcal{P}_{n,k}(\delta)$, $k = 1, ..., 4$, we can write

$$\begin{aligned}\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n(\delta)} P\{S_0 = \hat{S}\} &= \liminf_{n\to\infty} \min_{1\leq k\leq 4} \inf_{P\in\mathcal{P}_{n,k}(\delta)} P\{S_0 = \hat{S}\} \\ &= \min_{1\leq k\leq 4} \liminf_{n\to\infty} \inf_{P\in\mathcal{P}_{n,k}(\delta)} P\{S_0 = \hat{S}\}.\end{aligned}$$

The assumptions in (3.1) tell us that the last term is bounded from below by $1-\alpha$.

4. Testing for Information Sharing on Unobservables

4.1. The Model with Simple Type Players

One may want to see how much empirical relevance there is for incorporating information sharing on unobservables. Here we explain how one can performa a formal test of information sharing for the case of $\beta_0 \neq 0$. Observe that when $\beta_0 = 0$, presence of information sharing on unobservables is not testable. When $\beta_0 = 0$, it follows that

$$s_i^{[0]}(\mathcal{I}_{i,0}) = X_i'\rho_0 + v_i,$$

where $v_i = \varepsilon_i + \eta_i$. In this case, it is not possible to distinguish between contributions from ε_i and η_i.

Consider the following hypotheses:

$$H_0 : \sigma_\varepsilon^2 = 0, \text{ and } H_1 : \sigma_\varepsilon^2 > 0,$$

where we recall the definition $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_i^2|\mathcal{F})$. The null hypothesis tells us that there is no information sharing on unobservables. Let $\hat{v}_i(\beta)$, $a_\varepsilon(\beta)$ and $b_\varepsilon(\beta)$ be the same as $\hat{v}_i$, a_ε and b_ε (defined in Appendix C of the Supplemental Note at the end of Canen, Schwartz, and Song (2019)) only with β_0 replaced by generic β. From here on we assume that $\beta_0 \neq 0$.

The main idea for testing the hypothesis is that when $\sigma_\varepsilon^2 > 0$, this implies cross-sectional dependence of residuals v_i. For testing, we need to compute the sample version of the covariance between v_i and v_j for G_P-neighbors i and j. However, Condition C alone does not guarantee that for each $i \in N^*$, we will be able to compute $\hat{v}_j$ for some $j \in N_P(i)$, because there may not exist such j for some $i \in N^*$ at all. Thus let us introduce an additional data requirement as follows:

Condition D: For each $i \in N^*$, the econometrician observes a nonempty subset $\widetilde{N}(i) \subset N_P(i)$ (possibly a singleton) of agents where for each $j \in \widetilde{N}(i)$, the econometrician observes Y_j, $|N_P(j) \cap N_P(k)|$, $n_P(k)$ and X_k for all $k \in N_P(j)$.

Condition D is satisfied if there are many agents in the data set where each agent has at least one G_P-neighbor j for which the econometrician observes the outcome Y_j, the number of their G_P-neighbors, the observed characteristics of their G_P-neighbors, and the number of the agents who are both their G_P-neighbors and the neighbors of their G_P-neighbors. The asymptotic validity of inference is not affected if the researcher chooses a nonempty subset $\widetilde{N}(i)$ in Condition D as a singleton subset, say, $j(i) \subset N_P(i)$, $j(i) \in N$, such that we observe $Y_{j(i)}$, $|N_P(j(i)) \cap N_P(k)|$, $n_P(k)$ and X_k for all $k \in N_P(j(i))$ are available in the data, so far as the choice is not based on Y_i's but on X only. While this data requirement can still be restrictive in some cases where one obtains a partial observation of G_P, it is still weaker than the usual assumption that the econometrician observes G_P fully together with $(Y_i, X_i')_{i \in N}$.

Now let us reformulate the null and the alternative hypotheses as follows:

$$H_0 \ : \ \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \mathbf{E}[v_i v_j | \mathcal{F}] = 0, \text{ and} \tag{4.1}$$

$$H_1 \ : \ \frac{1}{n^*} \sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \mathbf{E}[v_i v_j | \mathcal{F}] \neq 0. \tag{4.2}$$

For testing, we propose the following method. Let $C^{\beta}_{1-(\alpha/2)}$ be the $(1-(\alpha/2))$-level confidence interval for β. We consider the following test statistics:

$$\widehat{IU} = \inf_{\beta \in C^{\beta}_{1-(\alpha/2)}} \frac{1}{2\hat{S}^4(\beta) n^*} \left(\sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \hat{v}_i(\beta) \hat{v}_j(\beta) \right)^2,$$

where

$$\hat{S}^2(\beta) = \frac{\widetilde{d}_{av}^{1/2}}{n^*} \sum_{i \in N^*} \hat{v}_i^2(\beta), \text{ and } \widetilde{d}_{av} = \frac{1}{n^*} \sum_{i \in N^*} |\widetilde{N}(i)|.$$

When the confidence set includes zero, the power of the test becomes asymptotically trivial, as expected from the previous remark that information sharing on unobservables is not testable when $\beta_0 = 0$.

As for the critical value, we take the $(1-(\alpha/2))$-percentile from the χ^2 distribution with degree of freedom 1, which we denote by $c_{1-(\alpha/2)}$. Then the level α-test based on the test statistic $\widehat{IU}$ rejects the null hypothesis if and only if $\widehat{IU} > c_{1-(\alpha/2)}$.

Theorem 4.1. *Suppose that the conditions of Theorem 2.1 and Assumptions 3.1-3.5 of Canen, Schwartz, and Song (2019) hold. Then, under the null hypothesis in (4.1),*

$$\lim_{n^*\to\infty} P\left\{\widehat{IU} > c_{1-\alpha/2}\right\} \le \alpha,$$

as $n^* \to \infty$.

Proof: First, note that

$$\frac{1}{\sqrt{n^*}} \sum_{i\in N^*} \sum_{j\in \widetilde{N}(i)} (\hat{v}_i \hat{v}_j - v_i v_j) = O_P(1/\sqrt{n^*}),$$

by following precisely the same proof as that of Lemma D.6(ii) in Canen, Schwartz, and Song (2019). (Recall that $\widetilde{N}(i)$ is defined in Condition D above.) Now, we let

$$\sigma^2 = \text{Var}\left(\frac{1}{\sqrt{n^*}} \sum_{i\in N^*} \sum_{j\in \widetilde{N}(i)} \eta_i \eta_j | \mathcal{F}\right)$$

and write

$$\frac{1}{\sigma\sqrt{n^*}} \sum_{i\in N^*} \sum_{j\in \widetilde{N}(i)} v_i v_j = \frac{1}{\sqrt{n^*}} \sum_{i\in N^*} r_i,$$

where

$$r_i = \frac{1}{\sigma} \sum_{j\in \widetilde{N}(i)} \eta_i \eta_j,$$

because $v_i = \eta_i$ under the null hypothesis. Note that $\mathbf{E}[r_i|\mathcal{F}] = 0$. Let G_P^* be a graph on N^* such that i and j are adjacent if and only if $j \in \widetilde{N}(i)$ or $i \in \widetilde{N}(j)$. Then $\{r_i\}_{i\in N^*}$ has G_P^* as a dependency graph conditional on $\mathcal{F}$. Now we show the following:

$$(n^*)^{-1/4}\sqrt{\mu_3^3} + (n^*)^{-1/2}\mu_4^2 \to_P 0, \quad (4.3)$$

where for $p \ge 1$,

$$\mu_p = \max_{i\in N^*} \left(\mathbf{E}[|r_i|^p|\mathcal{F}]\right)^{1/p}.$$

Then by Theorem 2.3 of Penrose (2003), we obtain that

$$\frac{1}{\sigma\sqrt{n^*}} \sum_{i\in N^*} \sum_{j\in \widetilde{N}(i)} v_i v_j \to_d N(0,1),$$

as $n^* \to \infty$. First, note that

$$\begin{aligned}\sigma^2 &= \mathbf{E}\left(\left(\frac{1}{\sqrt{n^*}}\sum_{i\in N^*}\sum_{j\in\widetilde{N}(i)}\eta_i\eta_j\right)^2|\mathcal{F}\right)\\ &= \frac{1}{\sqrt{n^*}}\sum_{i_1\in N^*}\sum_{j_1\in\widetilde{N}(i_1)}\sum_{i_2\in N^*}\sum_{j_2\in\widetilde{N}(i_2)}\mathbf{E}\left[\eta_{i_1}\eta_{j_1}\eta_{i_2}\eta_{j_2}|\mathcal{F}\right].\end{aligned}$$

Note that in the quadruple sum, $i_1 \neq j_1$ and $i_2 \neq j_2$. There are only two ways the last conditional expectation is not zero: either $i_1 = i_2$ and $j_1 = j_2$ or $j_1 = i_2$ and $i_1 = j_2$, because η_i's are independent across i's and its conditional expectation given $\mathcal{F}$ is zero. Hence the last term is equal to

$$\frac{2\sigma_\eta^4}{n^*}\sum_{i\in N^*}|\widetilde{N}(i)| = 2\sigma_\eta^4\widetilde{d}_{av} \tag{4.4}$$

Hence for any $p \geq 2$,

$$\begin{aligned}\mu_p^p = \frac{1}{\sigma^p}\max_{i\in N^*}\mathbf{E}\left[\left|\sum_{j\in\widetilde{N}(i)}\eta_i\eta_j\right|^p|\mathcal{F}\right] &\leq \frac{\max_{i,j\in N^*}\mathbf{E}[|\eta_i\eta_j|^p|\mathcal{F}]}{\sigma^p}\\ &\leq \frac{\max_{i,j\in N^*}\mathbf{E}[|\eta_i\eta_j|^p|\mathcal{F}]}{2^p\sigma_\eta^{2p}\widetilde{d}_{av}^p}.\end{aligned}$$

Note that $\widetilde{d}_{av} \geq 1$ because $\widetilde{N}(i) \neq \varnothing$ for all $i \in N^*$. Thus (4.3) follows. Now, by Lemma D.6 of Canen, Schwartz, and Song (2019), and in the light of the expression (4.4), it is not hard to see that

$$2\hat{S}^4(\beta_0) = \sigma^2 + o_P(1).$$

The desired result follows from this and the Bonferroni procedure. ■

4.2. The Model with First Order Sophisticated Players

Let us develop a test for information sharing on unobservables when the game is populated by the first order sophicated players. When $\beta_0 = 0$, it follows that

$$s_i^{[1]}(\mathcal{I}_{i,1}) = X_i'\rho_0 + v_i^{\mathsf{FS}},$$

where $v_i^{\mathsf{FS}} = \varepsilon_i + \eta_i$. Therefore, just as in the case of a simple type model, it is not possible to distinguish between contributions from ε_i and η_i. Thus let us assume that $\beta_0 \neq 0$. The presence of cross-sectional correlation of residuals v_i^{FS} serves as a testable implications from information sharing on unobservables. As in the case of a model with agents of simple type, we need to strengthen Condition D as follows:

Condition D1: For each $i \in N^*$, the econometrician observes a nonempty subset $\widetilde{N}(i) \subset N_P(i)$ (possibly a singleton) of agents where for each $j \in \tilde{N}(i)$, the econometrician observes Y_j, $|N_P(j) \cap N_P(k)|$, $n_P(k)$ and X_k for all $k \in N_{P,2}(j)$.

Similarly as before, we consider the following test statistics:

$$\widehat{IU}^{\mathsf{FS}} = \inf_{\beta \in C^{\beta}_{1-(\alpha/2)}} \frac{1}{2(\hat{S}^{\mathsf{FS}}(\beta))^4 n^*} \left(\sum_{i \in N^*} \sum_{j \in \widetilde{N}(i)} \hat{v}_i^{\mathsf{FS}}(\beta) \hat{v}_j^{\mathsf{FS}}(\beta) \right)^2,$$

where

$$(\hat{S}^{\mathsf{FS}}(\beta))^2 = \frac{\widetilde{d}_{av}^{1/2}}{n^*} \sum_{i \in N^*} \hat{v}_i^2(\beta).$$

As before, we reject the null hypothesis of no information sharing on unobservables if and only if $\widehat{IU}^{\mathsf{FS}} > c_{1-(\alpha/2)}$, where $c_{1-(\alpha/2)}$ is the $(1-(\alpha/2))$-percentile of χ^2_1. Asymptotic validity of this procedure can be shown in a similar manner as for the case of simple types.

5. Empirical Results Based on a Game with the First Order Sophisticated Players

In this section, we report the empirical results based on the game Γ_1 populated by the first order sophisticated players. The results are found in Table 6.

Compared to the results with simple types, the confidence sets for β in the game with first order sophisticated types are wider. For all specifications, the confidence sets for β for the FOS types includes (most or all of) the confidence set for β for the simple type. In general, the average marginal effects are similar across both models.[4] We note that the instruments used below are the same as those for simple players: polynomials of $X_{i,1}$ and a set of instruments that captures the cross-sectional dependence along the payoff graph ($\tilde{Z}_i = n_P(i)^{-1} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,1}$).

[4]A caveat is that the numerical implementation for the specifications in Columns (3) and (4) in Table 6 appear more sensitive than the others for β close to -1, relying on how the grid is set for those values. Throughout the empirical results in the paper, we present results for a grid of $\beta \in [-0.75, 0.75]$. As can be seen in Table 6, Columns (3)-(4) include a disjoint subset at the smallest values of the grid. This interval does not show up in any of the other specifications (simple type or first order sophisticated), disappears in other specifications similar to Columns (3)-(4) (e.g. when we restrict the set of covariates for land and river quality) and is not present when we consider a smaller grid. We attribute this to (i) lack of variation as $\beta \to -1$ due to more extensive cross-sectional dependence (recall that we need $|\beta|$ away from 1 by Assumption 2.1) (ii) less variation coming from the instruments $\tilde{Z}_i$ when β is smaller. The positive subset in the confidence interval for β is stable across specifications, and we focus on it for the discussion of our results.

TABLE 6. State Presence and Networks Effects across Colombian Municipalities, First Order Sophisticated Types

	Outcome: The Number of State Employees			
	Baseline (1)	Distance to Highway (2)	Land Quality (3)	Rivers (4)
β_0	$[0.17, 0.35]$	$[0.17, 0.35]$	$[-0.75, -0.69] \cup [0.17, 0.45]$	$[-0.75, -0.56] \cup [0.08, 0.46]$
dy_i/d(colonial state officials)	$[-0.055, 0.004]$	$[-0.046, 0.001]$	$[-0.045, 0.005]$	$[-0.032, 0.001]$
Average dy_i/d(colonial state agencies)	$[-1.222, 3.667]$	$[-1.118, 2.611]$	$[-0.926, 3.047]$	$[-3.953, 2.564]$
Average dy_i/d(distance to Royal Roads)	$[-0.010, 0.009]$	$[-0.008, 0.010]$	$[-0.015, 0.021]$	$[-0.013, 0.022]$
n	1018	1018	1003	1003

Notes: Confidence sets for β are presented in the table, obtained from inverting the test statistic $T^{\mathsf{FS}}(\beta)$ in (2.7), with confidence level of 95%. The critical values in the first row come from the asymptotic statistic. Downweighting is used. The average marginal effects for historical variables upon state capacity are also shown. The marginal effect of Colonial State Officials is equal to its γ coefficient. The marginal effect for Distance to Royal Roads for municipality i equals $\gamma_{Royal\ Roads} + 2\gamma_{Royal\ Roads^2}(Royal\ Roads)_i$, where $\gamma_{Royal\ Roads}$ is the γ coefficient of its linear term, and $\gamma_{Royal\ Roads^2}$ is the coefficient of its quadratic term, as this variable enters X_1 as a quadratic form. The analogous expression holds for the variable Colonial State Agencies. We show the average marginal effect for these two variables. We then present the confidence set for these marginal effects, computed by the inference procedure on $a'\gamma$ developed in Section 3. All specifications include controls of latitude, longitude, surface area, elevation, rainfall, as well as Department and Department capital dummies. Instruments are constructed from payoff neighbors' sum of the G_P neighbors values of the historical variables Total Crown Employees, Colonial State Agencies, Colonial State Agencies squared, population in 1843, distance to Royal Roads, distance to Royal Roads squared, together with the non-linear function $\tilde{Z}_i = n_P(i)^{-1} \sum_{j \in N_P(i)} \lambda_{ij} X_{j,1}$. Column (2) includes distance to current highway in X_1, Column (3) expands the specification of Column (2) by also including controls for land quality (share in each quality level). Column (4) controls for rivers in the municipality and land quality, in addition to those controls from Column (1). One can see that the results are very stable across specifications.

Given the results for the empirical model based on the game Γ_1, we then conduct the model selection procedure developed previously. This selects among the simple type and first order sophisticated type models. Table 7 presents the results. As the results show, the data did not reject either of the models ST and FS at 5%.

TABLE 7. Model Selection, Simple Type or First Order Sophisticated

	Specification			
	(1)	(2)	(3)	(4)
Test Statistics and p-values (in parentheses)				
T_{ST}	3.361 (0.762)	4.705 (0.582)	0.495 (0.998)	4.756 (0.575)
T_{FS}	4.260 (0.642)	4.897 (0.557)	1.018 (0.985)	5.010 (0.543)
Models Not Rejected				
$\hat{S}$	$\{\mathsf{ST}, \mathsf{FS}\}$	$\{\mathsf{ST}, \mathsf{FS}\}$	$\{\mathsf{ST}, \mathsf{FS}\}$	$\{\mathsf{ST}, \mathsf{FS}\}$

Notes: The table shows the results of the Model Selection test, developed in Section 3. Here ST refers to the simple type model, FS to the First Order Sophisticated. The critical value for the test ($c_{1-\alpha/2}$), with 6 degrees of freedom ($M - d$) and level $\alpha = 0.05$, is 14.449. The specifications in each column are the same as those in Table 6. The first panel shows the values of the statistics, with the p-values in parentheses. The bottom panel shows the set $\hat{S}$ of models that are not rejected by the test.

References

CANEN, N., J. SCHWARTZ, AND K. SONG (2019): “Estimating Local Interactions Among Many Agents Who Observe Their Neighbors,” *arXiv:1704.02999v4 [stat.ME]*.

PENROSE, M. (2003): *Random Geometric Graphs*. Oxford University Press, New York, USA.

SHORACK, G. R. (2000): *Probability for Statistics*. Springer, New York.