# SUPPLEMENT TO "INFERENCE OF SIGNS OF INTERACTION EFFECTS IN SIMULTANEOUS GAMES WITH INCOMPLETE INFORMATION"

(*Econometrica*, Vol. 80, No. 1, January 2012, 143–172)

BY ÁUREO DE PAULA AND XUN TANG

## APPENDIX A: ADDITIONAL RESULTS

### *Proof of Nonidentification of the Full Structure*

Because the data only provide information on the mixtures of equilibria, there are limits to what can be learned about the structure from the data without additional assumptions. This point is illustrated in this appendix using results from the literature on identifiability (or lack thereof) in mixture models.

Let $\theta$ denote the structure $(u_i, \delta_i)_{i=1}^N$ and $F_{\varepsilon|X}$, and let $\mathcal{L}_{x,\theta}$ denote the choice probabilities profiles corresponding to BNE for a given $x$ and parameter $\theta$. That is, $\mathcal{L}_{x,\theta} \equiv \{p \in [0,1]^N : p \text{ solves (1) for } \theta \text{ and the given } x\}$. We let $\Lambda_{x,\theta}$ be an equilibrium selection mechanism. The following proposition illustrates the limits of what can be learned about the structure from the mixture data without imposing additional assumptions. Let $\#A$ denote the cardinality of set $A$ and define $h : [0,1]^N \longrightarrow [0,1]^N$ as

$$h(p(x); x, \theta) \equiv \left( p_i(x) - F_{\varepsilon_i|X}\left( u_i(x) + \delta_i(x) \sum_{j \neq i} p_j(x) \right) \right)_{i=1,\ldots,N}. \tag{A.1}$$

PROPOSITION A.1: *Assume*

$$\det\left( \frac{\partial h(p(x); x, \theta)}{\partial p(x)} \right) \neq 0.$$

*Then the structure is not identified if* $\#\mathcal{L}_{x,\theta} > \frac{2^N - 2}{N}$.

PROOF: We first show that, for given $x$, the number of equilibria is finite. An equilibrium vector $p(x)$ is a fixed point to the mapping depicted on display (1). Equivalently, we represent it as a solution to the equation

$$h(p(x); x, \theta) = 0.$$

Notice that $\{0,1\} \cap F_{\varepsilon_i|x}(\mathbb{R}) = \emptyset$ for any $i$, given the full support of $\varepsilon_i$. Consequently, for a solution vector, $p_i(x) \notin \{0,1\}$ and $p(x) \in (0,1)^N$. Since

$$\det\left( \frac{\partial h(p(x); x, \theta)}{\partial p(x)} \right) \neq 0,$$

© 2012 The Econometric Society DOI: 10.3982/ECTA9216

the implicit function theorem directly implies that the set of fixed points to (A.1) is discrete (i.e., its elements are isolated points: each element is contained in a neighborhood with no other solutions to the system). Infinitesimal changes in $p(x)$ will imply a displacement of $h(\cdot; x, \theta)$ from zero, so local perturbations in $p(x)$ cannot be solutions to the system of equations. Since $p(x) \in [0,1]^N$, the set of solutions is a bounded subset of $\mathbb{R}^N$. In $\mathbb{R}^N$, every bounded infinite subset has a limit point (i.e., an element for which every neighborhood contains another element in the set) (Theorem 2.42 in Rudin (1976)). Consequently, a discrete set, having no limit points, cannot be both bounded and infinite. Being bounded and discrete, the set of solutions is finite.

In this case, the observed joint distribution of equilibrium actions is a finite mixture. Given Assumption 1, the cumulative distribution function for the observed actions is given by

$$\Phi(y_1, \ldots, y_N; x, \theta) = \sum_{\mathcal{L}_{x,\theta}} \Lambda_{x,\theta}(p^l(x)) \prod_{i \in \{1,\ldots,N\}} (1 - p_i^l(x))^{1-y_i}.$$

For a given $x$, the problem of retrieving this cumulative distribution function and mixing probabilities from observed data was analyzed by Hall, Neeman, Pakyari, and Elmore (2005), who showed that the choice and mixing probabilities ($p_i^l(x)$ and $\Lambda_{x,\theta}$) cannot be obtained from observation of $\Phi(y_1, \ldots, y_N; x, \theta)$ if $\#\mathcal{L}_{x,\theta} > \frac{2^N-2}{N}$. Consequently, it is necessary for identifiability of the relevant probabilities that $\#\mathcal{L}_{x,\theta} \leq \frac{2^N-2}{N}$. Finally, if the equilibrium-specific choice probabilities cannot be identified, the utility function and the distribution of private components cannot be identified either (or else one could obtain the equilibrium-specific choice probabilities and use those to obtain the mixing distribution from the data). *Q.E.D.*

The condition that $\det(\frac{\partial h(p(x);x,\theta)}{\partial p(x)}) \neq 0$ is likely to be satisfied. With two players, for example, this determinant equals

$$1 - \delta_1(x)\delta_2(x) f_{\varepsilon_1|X}(u_1(x) + \delta_1(x) p_2(x)) f_{\varepsilon_2|X}(u_2(x) + \delta_2(x) p_1(x)).$$

Also when there are two players, the bound on the number of equilibria implies that, without further assumptions, the existence of more than one equilibrium precludes identification.

## APPENDIX B: HETEROGENEOUS PAYOFF IMPACTS

Here we define $U_{1i}(X, \varepsilon_i) \equiv u_i(X) + \delta_i(X) f_i(X, D_{-i}) - \varepsilon_i$. For a fixed $x$, any function $f_i(x, D_{-i})$ can take at most $2^{N-1}$ values corresponding to the possible $D_{-i}$ vectors: $\{f_i(x, \pi) : \pi \in \{0,1\}^{N-1}\}$. We can then write $f_i(x, D_{-i}) = \sum_{\pi \in \{0,1\}^{N-1}} f_i(x, \pi) \prod_{j \neq i} 1\{D_j = \pi_j\} = \sum_{\pi \in \{0,1\}^{N-1}} f_i(x, \pi) \prod_{j \neq i} D_j^{\pi_j} (1 - D_j)^{1-\pi_j}$, where $\pi_j$ denotes the $j$th component of $\pi$ and $1\{\cdot\}$ is the indicator function.

For example, if $N = 3$ and $f_i(x, D_{-i}) = \max_{j \neq i}(D_j)$, we have that $f_1(x, D_{-1}) = 1 \times D_2 D_3 + 1 \times D_2(1 - D_3) + 1 \times (1 - D_2)D_3 + 0 \times (1 - D_2)(1 - D_3) = D_2(1 - D_3) + D_3(1 - D_2) + D_2 D_3$ (and analogously for $i = 2, 3$). By Assumption 1, in a single equilibrium indexed by $l$,

$$\begin{aligned}
&\mathbb{E}\big[f_i\big(X, (S_j^l(X, \varepsilon_j))_{j \neq i}\big) | X = x, \varepsilon_i\big] \\
&\quad = \mathbb{E}\big[f_i\big(X, (S_j^l(X, \varepsilon_j))_{j \neq i}\big) | X = x\big] \\
&\quad = \sum_{\pi \in \{0,1\}^{N-1}} f_i(x, \pi) P^l(D_{-i} = \pi | x) \\
&\quad = \sum_{\pi \in \{0,1\}^{N-1}} \left[ f_i(x, \pi) \prod_{j \neq i} p_j^l(x)^{\pi_j} (1 - p_j^l(x))^{1 - \pi_j} \right] \\
&\quad \equiv \phi_{f_i}(x, p_{-i}^l(x)),
\end{aligned}$$

where $p_j^l(x) \equiv \mathbb{E}(S_j^l(X, \varepsilon_j) | X = x)$ as before, $p_{-i}^l(x) \equiv (p_j^l(x))_{j \neq i}$, and $P^l(\omega | x)$ denotes the probability that the event $\omega$ happens conditional on $x$, as implied in the equilibrium $p^l$. Notice also that the mapping $\phi_{f_i} : \Omega_X \times [0, 1]^{N-1} \to \mathbb{R}$ is a simple extension of $f_i$ to $\Omega_X \times [0, 1]^{N-1}$. It is known as long as $f_i$ is known. The equations that characterize a single equilibrium $p^l$ in (1) now become

$$p_i^l(x) = F_{\varepsilon_i | X = x}\big(u_i(x) + \delta_i(x) \phi_{f_i}(x, p_{-i}^l(x))\big) \quad \text{for all } i = 1, \ldots, N.$$

Then the results in Proposition 1 now apply with $\gamma_i^l(x) \equiv \phi_{f_i}(x, p_{-i}^l(x))$. Note that, by the law of total probability,

$$\gamma_i^*(x) \equiv \sum_{\pi \in \{0,1\}^{N-1}} [f_i(x, \pi) P^*(D_{-i} = \pi | x)], \tag{B.1}$$

where $P^*(\omega | x)$ denotes the probability that $\omega$ occurs conditional on $x$, observed from the data. Furthermore,

$$\tilde{\gamma}_i^*(x) \equiv \sum_{\pi \in \{0,1\}^{N-1}} \big[f_i(x, \pi) P^*\{(D_{-i}, D_i) = (\pi, 1) | x\}\big] \tag{B.2}$$

given the fact that $P^l\{(D_{-i}, D_i) = (\pi, 1) | x\} = p_i^l(x) \prod_{j \neq i} p_j^l(x)^{\pi_j} (1 - p_j^l(x))^{1 - \pi_j}$ under Assumption 1 in the paper, the law of total probability, and the definition of $P^*(\cdot | x)$. Hence $p_i^*(x)$ and $\gamma_i^*(x)$ as defined in (B.1), and $\tilde{\gamma}_i^*(x)$ as defined in (B.2) can all be expressed in terms of observable distributions. Thus the sign of $\delta_i(x)$ is identified and multiple BNE can be detected as in Proposition 1.

## APPENDIX C: A WALD TEST FOR MULTIPLE BNE

By the delta method,

$$G^{1/2}(\mathbf{T}_G - \boldsymbol{\Delta}) \xrightarrow{d} N(\mathbf{0}_N, \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}') \quad \text{as} \quad G \to \infty,$$

where $\boldsymbol{\Delta} \equiv (\Delta_i)_{i=1}^N$. The Jacobian $\mathbf{V}$ is an $N$-by-$\tilde{N}$ matrix, with its $i$th row $V_i$ defined by the following table (where $\mu_{(m)}$ and $V_{i,(m)}$ denote the $m$th coordinates of two $\tilde{N}$ vectors $\boldsymbol{\mu}$ and $V_i$, respectively, and $j, k \neq i$):

| $\mu_{(m)}$ | $\mu_0$ | $\mu_i$ | $\mu_j$ | $\mu_{ij}$ or $\mu_{ji}$ | $\mu_{jk}$ |
|---|---|---|---|---|---|
| $V_{i,(m)}$ | $\sum_{j\neq i}(-\frac{\mu_{ij}}{\mu_0^2} + \frac{2\mu_i\mu_j}{\mu_0^3})$ | $-\sum_{j\neq i}\frac{\mu_j}{\mu_0^2}$ | $-\frac{\mu_i}{\mu_0^2}$ | $\frac{1}{\mu_0}$ | 0 |

Let $\hat{\boldsymbol{\Sigma}}$ and $\hat{\mathbf{V}}$ be estimates for $\boldsymbol{\Sigma}$ and $\mathbf{V}$, respectively, constructed by replacing $\mu_0$ and $\mu_I$ with nonparametric estimates

$$\hat{\mu}_0 = G^{-1}\sum_g 1(X_g = x), \quad \hat{\mu}_I = G^{-1}\sum_g \Big[\prod_i D_{i,g} 1(X_g = x)\Big].$$

PROPOSITION C.1: *Suppose the data have $G$ independent games with the same underlying structure, and both $V$ and $\Sigma$ are full rank. Then*

$$G(\mathbf{T}_G - \boldsymbol{\Delta})'(\hat{\mathbf{V}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{V}}')^{-1}(\mathbf{T}_G - \boldsymbol{\Delta}) \xrightarrow{d} \chi^2_{df=N} \quad \textit{as} \quad G \to \infty.$$

Under the null, $\boldsymbol{\Delta} = \mathbf{0}_N$ and the chi-squared distribution can be used to obtain critical values for the test statistic $G\mathbf{T}_G'(\hat{\mathbf{V}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{V}}')^{-1}\mathbf{T}_G$. Because $N \geq 3$ and conditional choice probabilities are bounded away from 0 and 1 (due to the rich support condition in Assumption 1), the full-rank conditions above are not restrictive.

## APPENDIX D: ALGORITHMS FOR STEPWISE PROCEDURE

The following algorithm summarizes the stepwise multiple-testing procedure we adopt from Romano and Wolf (2005).

ALGORITHM D.1—Basic Nonstudentized Step-Down Procedure:

*Step 1.* Relabel the hypotheses in descending order of the test statistics $T_{G,i}$. Let $H^0_{i_k}$ denote the individual null hypothesis whose test statistic is the $k$th largest.

*Step 2.* Set $k = 1$ and $R_1 = 0$.

*Step 3.* For $R_k + 1 \leq s \leq N$, if $T_{G,(s)} - \hat{c}_k > 0$, then reject the individual null $H^0_{i_s}$.

*Step 4.* If no (further) null hypotheses are rejected, then stop. Otherwise let $R_{k+1}$ denote the total number of hypotheses rejected so far (i.e., $R_k$ plus the number of hypotheses rejected in the $k$th step) and set $k = k + 1$. Then return to Step 3 above.

We consider two alternative methodologies for the computation of $\hat{c}_k$: bootstrap and using the asymptotic distribution of the test statistic. The two are summarized in the following two algorithms.

ALGORITHM D.2—Computing $\hat{c}_k$ Using Bootstrap:

*Step 1.* Let $i_k$ and $R_k$ be defined as in Algorithm D.1 above.

*Step 2.* Generate $B$ bootstrap data sets.

*Step 3.* From each bootstrap data set (indexed by $b$), compute the vector of test statistics $(T^{*,b}_{G,1}, \ldots, T^{*,b}_{G,N})$.

*Step 4.* For $1 \leq b \leq B$, compute $\max^{*,b}_{G,k} = \max_{R_k+1 \leq s \leq N}(T^{*,b}_{G,i_s} - T_{G,i_s})$.

*Step 5.* Then compute $\hat{c}_k$ as the $(1-\alpha)$th empirical quantile of the $B$ values $\{\max^{*,b}_{G,k}\}_{b \leq B}$.

ALGORITHM D.3—Computing $\hat{c}_k$ Using Parametric Simulations:

*Step 1.* Estimate the covariance matrix of the vector of test statistics that corresponds to hypotheses which are not rejected after the first $k-1$ steps, that is, $(T_{G,(R_k+1)}, T_{G,(R_k+2)}, \ldots, T_{G,(N)})$. Denote the estimate by $\hat{\Sigma}_k$.

*Step 2.* Simulate a data set of $M$ observations $\{v_m\}_{m=1}^{M}$ from the $(N - R_k)$-dimensional multivariate normal distribution with parameters $(0_{N-R_k}, \hat{\Sigma}_k)$, where $0_k$ is a $k$ vector of zeros.

*Step 3.* Then $\hat{c}_k$ is computed as the $(1-\alpha)$th empirical quantile of the maximum coordinates of $v_m$ in the simulated data. ($M$ can be large relative to the number of bootstrap samples $B$ in Algorithm D.2 above.)

The studentized stepwise procedure is summarized in the following algorithm. As before, $R_k$ denotes the total number of hypotheses not rejected in the first $k-1$ steps.

ALGORITHM D.4—Studentized Step-Down Procedure:

*Step 1.* Relabel the individual hypotheses in descending order of studentized test statistics $Z_{G,i} \equiv T_{G,i}/\hat{\sigma}_{G,i}$, where $\hat{\sigma}_{G,i}$ are estimates for standard deviation of $T_{G,i}$.

*Step 2.* Set $k = 1$ and $R_1 = 0$.

*Step 3.* For $R_k + 1 \leq s \leq S$, if $Z_{G,i_s} > \hat{d}_j$, then reject the individual null $H^0_{i_s}$.

*Step 4.* If no further individual null hypotheses are rejected, stop. Otherwise let $R_{k+1}$ denote the total number of hypotheses rejected so far and set $k = k + 1$. Then return to Step 3 above.

The critical values for the studentized stepwise method $\hat{d}_k$ are computed by an algorithm similar to Algorithm D.1, where standard errors $(\hat{\sigma}^{*,b}_{G,1}, \ldots, \hat{\sigma}^{*,b}_{G,N})$ are also computed in Step 3 and $\max^{*,b}_{G,k} \equiv \max_{R_k+1\leq s\leq N}(T^{*,b}_{G,i_s} - T_{G,i_s})/\hat{\sigma}^{*,b}_{G,i_s}$ in Step 4.

## REFERENCES

HALL, P., A. NEEMAN, R. PAKYARI, AND R. ELMORE (2005): "Nonparametric Inference in Multivariate Mixtures," *Biometrika*, 92 (3), 667–678. [2]

ROMANO, J. P., AND M. WOLF (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73 (4), 1237–1282. [4]

RUDIN, W. (1976): *Principles of Mathematical Analysis* (Third Ed.). New York: McGraw-Hill. [2]

*Dept. of Economics, University College London, Gower Street, London, WC1E 6BT, United Kingdom, CeMMAP, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, United Kingdom, and Dept. of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.; aureo@sas.upenn.edu*

*and*

*Dept. of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.; xuntang@sas.upenn.edu.*

*Manuscript received April, 2010; final revision received June, 2011.*