American Economic Review: Papers & Proceedings 2017, 107(5): 274–277
https://doi.org/10.1257/aer.p20171041

Core Determining Class and Inequality Selection†

By Ye Luo and Hai Wang*

Suppose we observe the outcomes but not the events that might imply the outcomes. In many situations the relations between events and outcomes are indeterministic, i.e., a single event may lead to different outcomes, and an outcome may have several events that may lead to it. Such relations can be characterized by a bipartite graph $G = (\mathcal{U}, \mathcal{Y}, \varphi)$, where $\mathcal{U}$ is a set of unobservable events, $\mathcal{Y}$ is a set of observed outcomes, and φ is a correspondence mapping from $\mathcal{U}$ to $\mathcal{Y}$ such that $\varphi(u) \subset \mathcal{Y}$ is the set of all possible outcomes that could be led by event $u \in \mathcal{U}$.

One important application of the above partite graph structure is game theory. When multiple equilibria exist, we have a partially identified model. Such a partially identified model can be characterized as a bipartite graph $G = (\mathcal{U}, \mathcal{Y}, \varphi)$ as we mentioned above. The key interest is to estimate bounds on the measure of $\mathcal{U}$, denoted as v, given a measure μ on $\mathcal{Y}$. Artstein's (1983) theorem indicates that all the information on v is contained in a collection of inequalities indexed by any subset A of $\mathcal{U}$, i.e.,

$$v(A) \leq \mu(\varphi(A)).$$

Therefore, the number of inequalities equals $2^{d_u} - 2$, where d_u denotes the cardinality of $\mathcal{U}$.

In such a case, when d_u is large, numerous inequalities will be generated from the corresponding bipartite graph. Such a case leads to two problems: first, many inference procedures such as those described in Chernozhukov, Hong, and Tamer (2007) may fail with large number of inequalities; second, the computation of these inference procedures is extremely time consuming and may cost intolerably long time in practice.

Galichon and Henry (2011) defines "Core Determining Class'' as a collection of subsets of $\mathcal{U}$ that contains sharp information. Chesher and Rosen (2012) proposes an algorithm that is able to eliminate some redundant inequalities. Our main contribution can be summarized as the following:

(i) We propose a method to select the exact set of irredundant inequalities. We define such a set of inequalities as "exact Core Determining Class." We prove that the inequalities selected are only dependent on the structure of the graph but independent from the probability measure observed on $\mathcal{Y}$ under mild nondegeneracy conditions, i.e., the noise in the observed measure on $\mathcal{Y}$ does not affect the "exact Core Determining Class" in general.

(ii) For a general linear inequalities selection problem under noise, we propose a selection procedure similar to the Dantzig-selector described in Candes and Tao (2007). We prove that the selection procedure has good statistical properties under some sparse assumptions.

(iii) We apply the selection procedure to construct the set of irredundant inequalities for the bipartite graph with data noise. We prove that the selection procedure has better statistical properties compared to that applied to the general problem due to the structure of the graph.

(iv) We demonstrate the good performance of our selection procedure through several sets of Monte-Carlo experiments: first, the inference based on the selection procedure has desired size; second, in Luo and Wang (2016), we designed an

*Luo: University of Florida (e-mail: kurtluo@gmail.com); Wang: Singapore Management University (e-mail: wanghaimit@gmail.com). We thank the discussant, Adam Rosen, at the session Machine Learning in Econometrics at the AEA annual meeting 2017 in Chicago for invaluable comments.

† Go to https://doi.org/10.1257/aer.p20171041 to visit the article page for additional materials and author disclosure statement(s).

experiment to demonstrate that the procedure has strong power against local alternatives; third, it is relatively computationally efficient.

I. Exact Core Determining Class

The Artstein's theorem stated in Artstein (1983) presents that all information of v in the bipartite graph model $G = (\mathcal{U}, \mathcal{Y}, \varphi)$ is characterized by the set of constraints described below:

LEMMA 1 (Artstein's Theorem): *The following set of inequalities/equalities contains sharp information on v*:

(i) *for any $A \subset \mathcal{U}$*,

$$v(A) := \sum_{u \in A} v(u) \leq \mu(\varphi(A)),$$

where $\mu(\varphi(A)) := \sum_{y \in \varphi(A)} \mu(y)$;

(ii) $\sum_{u \in \mathcal{U}} v(u) = 1$.

According to Artstein's theorem, the minimum set of irredundant inequalities, referred as "exact Core Determining Class" in this paper, can be characterized by the following conditions using linear programming.

DEFINITION 1 (Set $\mathcal{S}^*$): *The Exact Core Determining Class $\mathcal{S}^*$ is the collection of all subsets $A \subset \mathcal{U}$ and $A \neq \mathcal{U}$, such that*

$$v^{M^*}(A) > \mu(\varphi(A)),$$

where

$$v^{M^*}(A) := \max\{v(A) \mid v(A') \leq \mu(\varphi(A')), \forall A' \subset \mathcal{U}, A' \neq A; v(\mathcal{U}) = 1\}.$$

We define a collection of subsets of $\mathcal{U}$ using the following combinatorial rules.

DEFINITION 2 (Set $\mathcal{S}_u$): $\mathcal{S}_u' \subset 2^{\mathcal{U}}$ *is the collection of all nonempty subsets $A \subset \mathcal{U}$ and $A \neq U$, such that*:

(i) *A is self-connected, i.e., $\forall A_1, A_2 \subset A$ such that $A_1, A_2 \neq \emptyset$ and $A_1 \cup A_2 = A$, it holds that $\varphi(A_1) \cap \varphi(A_2) \neq \emptyset$;*

(ii) *there exists no $u \in \mathcal{U}$, such that $u \notin A$ and $\varphi(u) \subset \varphi(A)$.*

Similarly, we can define $\mathcal{S}_y$ as follows.

DEFINITION 3 (Set $\mathcal{S}_y$): $\mathcal{S}_y' \subset 2^{\mathcal{Y}}$ *is the collection of all subsets $B \subset \mathcal{Y}$ and $B \neq \mathcal{Y}$, such that*:

(i) *B is self-connected, i.e., $\forall B_1, B_2 \subset B$, such that $B_1, B_2 \neq \emptyset$ and $B_1 \cup B_2 = B$, it holds that $\varphi^{-1}(B_1) \cap \varphi^{-1}(B_2) \neq \emptyset$;*

(ii) *there exists no $y \in \mathcal{Y}$, such that $y \notin B$ and $\varphi^{-1}(y) \subset \varphi^{-1}(B)$.*

We can define $\mathcal{S}_y^{-1} := \{A \mid A \neq \mathcal{U},$ there exists $B \in \mathcal{S}_y$, such that $A = \varphi^{-1}(B)^c\}$.

THEOREM 1: *Assume that G is self-connected. If the measure μ on $\mathcal{Y}$ is nondegenerate, i.e., $\mu(y)$ is non-zero for all $y \in \mathcal{Y}$, then $S^* = \mathcal{S}_u \cap \mathcal{S}_y^{-1}$.*

Theorem 1 shows that in general the "exact Core Determining Class" does not depend on the numerical value of μ, but only depends on the structure of the graph G, since both S_u and S_y^{-1} depend on the structure of the graph G only. Luo and Wang (2016) proposes a fast combinatorial algorithm to construct S^* by exploring the structure of the graph G . In example 3 of Luo and Wang (2016), the algorithm helps to eliminate 98.56 percent of the inequalities for a 15×25 bipartite graph. The bipartite graph is available as Figure 1 in the online Appendix. In this example, 471 out of 32,766 inequalities are selected as the "exact Core Determining Class," while other inequalities are recognized as redundant inequalities.

II. A General Linear Inequality Selection Procedure

A. *General Framework*

We consider a more general framework where many linear inequalities exist. Consider a set of inequalities such that $Mv \leq b$ with M being an $m \times d$ matrix and b being an $m \times 1$ vector, $v \geq 0$.

Suppose the set of irredundant inequalities is indexed by $\mathcal{I}_0$. For any $\mathcal{I} \subset \{1, 2, \ldots, m\}$, we define $M_{\mathcal{I}}$ and $b_{\mathcal{I}}$ as the corresponding subvector of M and b. Define $Q = \{v|Mv \leq b\}$ and $Q_{\mathcal{I}} := \{v\,|\,M_{\mathcal{I}}v \leq b_{\mathcal{I}}\}$.

Farkas Lemma implies that there exists an $m \times m$ matrix $\tilde{\Pi}$ such that

$$(1) \qquad \tilde{\Pi} M \geq M,\ \tilde{\Pi} \geq 0,$$

$$(2) \qquad \tilde{\Pi} b \leq b.$$

In practice, assume that we observe $\hat{b}$ as a consistent estimator of b. In order to select the set of irredundant inequalities, we consider a procedure similar to the "Dantzig Selector'':

Procedure $\hat{\mathcal{R}}$,

$$\min_{\Pi} \sum_{k=1}^{m} ||\,\Pi_{*k}||_{\infty}$$

subject to

$$(3) \qquad \Pi M \geq M, \Pi \geq 0,$$

$$(4) \quad \Pi(\hat{b} - \Lambda_n) + 2\text{diag}(\Pi)\,\Lambda_n \ \leq\ \hat{b} + \Lambda_n,$$

where Π_{*k} is the k^{th} column of Π, $\Lambda_n := (\lambda_{n,1}, \ldots, \lambda_{n,m})$ is a vector of relaxation parameters, and objective function $||\cdot||_{\infty}$ is the sup-norm over a vector.

To guarantee good properties of $\hat{\mathcal{R}}$, we require Λ_n to be chosen such that with probability $1-\alpha$, $\lambda_{n,j} \geq |\hat{b}_j - b_j|$ uniformly for all $j = 1, 2, \ldots, m$; α should be chosen as a small number, e.g., 0.05. In practice, this can be done by utilizing results in Peña, Lai, and Shao (2009) or Chernozhukov, Chetverikov, and Kato (2013). The benefit of choosing the objective function as the sup-norm is that the optimization of the procedure $\hat{\mathcal{R}}$ becomes a linear programming problem, and linear programming is widely known to be fast in computational time.

Denote the solution of $\hat{\mathcal{R}}$ as $\hat{\Pi}^{L_1}$. Such a solution can be called the $L-1$ selector. The selected set of inequalities is defined by $\hat{\mathcal{I}}_{\eta} := \{k: ||\,\hat{\Pi}^{L_1}_{*k}|| \geq \eta\}$, where $0 < \eta < 1$ is a tuning parameter. Luo and Wang (2016) shows that under certain sparsity assumptions, $\hat{\mathcal{I}}_{\eta} \supset \mathcal{I}_0$ with probability going to 1. Luo and Wang (2016) also discusses a conservative inference procedure that is easy to implement in practice. Under their assumptions, Luo and Wang (2016) shows that $\hat{Q}_{\hat{\mathcal{I}}_{\eta}} \oplus \Lambda_n := \{v|v \geq 0, M_j v \leq \hat{b}_j + \lambda_{n,j},\ \text{for all } j \in \hat{\mathcal{I}}_{\eta}\}$ covers the identified region $Q := \{v|v \geq 0, Mv \leq b\}$ with probability at least $1-\alpha$.

For comparison, we also consider a procedure similar to $\hat{\mathcal{R}}$, but using sign$(||\cdot||_{\infty})$ as the objective function instead of $||\cdot||_{\infty}$. We denote the solution of such a procedure as the $L-0$ selector. The optimization of the $L-0$ selector requires integer programming, and therefore its computation in general much slower than the $L-1$ selector.

B. *Hybrid Methods in Core Determining Class Problem*

We propose a hybrid method to eliminate redundant inequalities in the Core Determining Class problem: First, we use the combinatorial algorithm described in Section I. Second, we apply the procedure $\hat{\mathcal{R}}$ to the inequalities selected from the first step. All the results are available in the Table 1 and Table 2 of the online Appendix.

In this experiment, we consider using the hybrid method for the bipartite graph with size 15×25 as we mentioned earlier. We examined different cutoff levels $\eta = 0, 0.1, 0.2$ and sample size $n = 500$ and 2,000. The number of inequalities are cut down substantially, with an average of $26.73 - 187.42$, depending on the value of η and the sample size. We compare our $L-1$ with the $L-0$ selector. While the $L-0$ selector is choosing 79 inequalities, our $L-1$ selector with cutoff level $\eta = 0.05$ and $\eta = 0.10$ choose 78 out of 79 inequalities that are selected by the $L-0$ selector. The time cost of the $L-1$ selector is 1.45 minutes, while the time cost of the $L-0$ selector is 2,195 minutes. Both $L-1$ and $L-0$ selectors are calculated with single thread on Intel Xeon Quad Core E5430 2.66GHz with 8GB memory.

REFERENCES

Artstein, Zvi. 1983. "Distributions of Random Sets and Random Selections." *Israel Journal of Mathematics* 46 (4): 313–24.

Candes, Emmanuel, and Terence Tao. 2007. "The Dantzig Selector: Statistical Estimation when p is much larger than n." *Annals of Statistics* 6: 2313–51.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. 2013. "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors." *Annals of Statistics* 41 (6): 2786–2819.

Chernozhukov, Victor, Han Hong, and Elie Tamer. 2007. "Estimation and Confidence Regions for Parameter Sets in Econometric Models." *Econometrica* 75 (5): 1243–84.

Chesher, Andrew, and Adam M. Rosen. 2012. "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification." Centre for Microdata Methods and Practice Working Paper CWP21/12.

Galichon, Alfred, and Marc Henry. 2011. "Set Identification in Models with Multiple Equilibria." *Review of Economic Studies* 78 (4): 1264–98.

Luo, Ye, and Hai Wang. 2016. "Core-Determining-Class: Construction, Approximation, and Inference." http://people.clas.ufl.edu/yeluo/files/V2.7pdf.

Peña, Victor H., Tze Leung Lai, and Qi-Man Shao. 2009. *Self-Normalized Processes: Limit Theory and Statistical Applications*. New York: Springer.