Robust Mechanism Design

Dirk Bergemann and Stephen Morris
Laffont Conference

June 21, 2005

Wilson doctrine

"Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent that it assumes other features to be common knowledge, such as one agent's probability assessment about another's preferences or information.

I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality." Wilson (1987)

Aspects of Robustness

1. Robustness to beliefs and higher order beliefs

2. Robustness to specific problem - "detail-free"

3. Robustness to description of environment: collusion, shill bidders, etc....

Relaxing Common Knowledge is equivalent to adding types

- Harsanyi (1967/68), Mertens & Zamir (1985) established environments with incomplete information can be modeled as a Bayesian game: wlog assume

  1. common knowledge among players of each player's type spaces
  2. each type's beliefs over types of other players

- Economic analysis assumes smaller type spaces than universal type space, *yet maintains common knowledge of (1) and (2)*

- Are results of economic analysis robust to inclusion of rich type spaces with little common knowledge? (Neeman (2003))

Example: Allocating a Single Good with Interdependent Valuations

- Quasi-linear environment with I agents

- Each agents has a "payoff type" $\theta_i \in [0, 1]$

- Each agent i's valuation of an object is $\theta_i + \gamma \sum_{j \neq i} \theta_j$

- Efficient allocation: object goes to individual with highest θ_i

Question 1: Robust Mechanism Design

- Fix a payoff environment (e.g., the single good example) and a social choice correspondence (e.g., efficient allocations): *does there exist a mechanism such that whatever the agents' beliefs and higher order beliefs about other agents' types, there is an equilibrium consistent with the social choice correspondence?*

- Answer (for separable environments): if and only if there is an ex post incentive compatible (EPIC) allocation

- Ex post incentive compatible allocation: each individual has an incentive to report his payoff type truthfully *whatever his beliefs about others' payoff types*.

- Efficient allocation is EPIC in example (Cremer and Mclean (1985), Dasgupta and Maskin (2001)) if $\gamma \leq 1$ and winner pays

$$\max_{j \neq i} \theta_j + \gamma \sum_{j \neq i} \theta_j .$$

- EPIC is clearly sufficient for robust mechanism design

- Examples show that without separability condition, EPIC is not necessary

Question 2: Ex Post Implementation

- Fix a payoff environment and a social choice correspondence: *does there exist a mechanism such that every ex post equilibrium delivers outcomes consistent with the social choice correspondence?*

- Answer: "ex post monotonicity" condition is necessary and almost sufficient

- Ex post monotonicity is satisfied by efficient allocation in single good example if there are at least three players

Question 3: Robust Implementation

- Fix a payoff environment and a social choice correspondence: *does there exist a mechanism such that such that whatever the agents' beliefs and higher order beliefs about other agents' types, every equilibrium is consistent with the social choice correspondence?*

- Answer: "robust monotonicity" condition is necessary and almost sufficient

- Robust monotonicity is satisfied in single good example for a nearly efficient allocation if and only if $\gamma \leq \frac{1}{I-1}$.

Idea of Impossibility Result

- Suppose $\gamma\,(I-1)>1$
- Consider a type of agent i with "payoff type" $\theta_i \in [0,1]$
- Suppose his expected value of $\frac{1}{I-1}\sum\limits_{j\neq i}\theta_j$ is $\overline{\theta}$, with $\overline{\theta}=\frac{1}{2}+\frac{1}{\gamma(I-1)}\left(\frac{1}{2}-\theta_i\right)$
- Then his value is

$$\begin{aligned} v_i &= \theta_i+\gamma\,(I-1)\left(\frac{1}{2}+\frac{1}{\gamma\,(I-1)}\left(\frac{1}{2}-\theta_i\right)\right) \\ &= \frac{1}{2}\,(1+\gamma\,(I-1)) \end{aligned}$$

- Any mechanism has fully pooling equilibrium where all types behave identically

Idea of Possibility Result

- Chung and Ely (2001): iterated deletion of weakly dominated strategies

- Nearly Efficient allocation achieved in direct allocation under iterated deletion of strictly dominated strategies

- efficient allocation....

$$x_i^*(\theta) = \begin{cases} \frac{1}{\#\{j:\theta_j \geq \theta_k \text{ for all } k\}}, \text{ if } \theta_i \geq \theta_k \text{ for all } k, \\ 0, \text{ otherwise.} \end{cases}$$

- a symmetric ε-efficient allocation rule is the following:

$$x_i^{**}(\theta) = \varepsilon \frac{\theta_i}{I} + (1 - \varepsilon) x_i^{*}(\theta).$$

Payoff Relevant Environment

- agent $i \in \{1, 2, ..., I\}$
- i's payoff relevant type $\theta_i \in \Theta_i$
- payoff relevant type profile $\theta \in \Theta = \Theta_1 \times ... \times \Theta_I$
- social outcome $a \in A$
- utility function $u_i : A \times \Theta \to \mathbb{R}$
- social choice correspondence $F : \Theta \to 2^A$

Mechanism and Ex Post Implementation

- A mechanism $\mathcal{M} = \left((M_i)_{i=1}^{I}, g \right)$, $g : M_1 \times .... \times M_I \to A$

- A pure strategy in payoff types game: $s_i : \Theta_i \to M_i$

- Strategy profile s is an *ex post equilibrium* if

$$u_i \left(g \left(s \left(\theta \right) \right), \theta \right) \geq u_i \left(g \left(m_i, s_{-i} \left(\theta_{-i} \right) \right), \theta \right)$$

for all i, θ and m_i.

Type Space

$$\mathcal{T} = \left\{ T_i, \widehat{\theta}_i, \widehat{\pi}_i \right\}_{i=1}^{I}$$

- i's *type* is $t_i \in T_i$
- $\widehat{\theta}_i(t_i)$ is i's payoff relevant type of t_i

$$\widehat{\theta}_i : T_i \to \Theta_i$$

- $\widehat{\pi}_i(t_i)$ is i's belief type of t_i

$$\widehat{\pi}_i : T_i \to \Delta(T_{-i})$$

Mechanism

- A mechanism $\mathcal{M}$ and a type space $\mathcal{T}$ is an incomplete information game
- A pure strategy $s_i : T_i \to M_i$
- Strategy profile s is a (Bayesian) equilibrium if

$$\sum_{t_{-i}} \widehat{\pi}_i (t_i) [t_{-i}] \, u_i \left(g(s(t)), \widehat{\theta}(t) \right)$$

$$\geq \sum_{t_{-i}} \widehat{\pi}_i (t_i) [t_{-i}] \, u_i \left(g(m_i, s_{-i}(t_{-i})), \widehat{\theta}(t) \right)$$

for all i, t_i.

Ex Post Incentive Compatibility

DEFINITION: A social choice function $f : \Theta \to A$ is ex post incentive compatible if, for all i and $\theta \in \Theta$,

$$u_i\left(f\left(\theta\right),\theta\right) \geq u_i\left(f\left(\theta_i',\theta_{-i}\right),\theta\right)$$

for all $\theta_i' \in \Theta_i$.

Robust Mechanism Design

The environment is *separable* if

$$\begin{aligned} A &= Y \times X_1 \times ... \times X_I \\ u_i\left(\left(y,x\right),\theta\right) &= v_i\left(y,x_i,\theta\right) \\ \text{and } F\left(\theta\right) &= f_0\left(\theta\right) \times F_1\left(\theta\right) \times ... \times F_1\left(\theta\right) \end{aligned}$$

where f_0 is a function.

THEOREM. If the environment is separable, there exists a mechanism $\mathcal{M}$ such that for every type space $\mathcal{T}$, there is an equilibrium s of $(\mathcal{M},\mathcal{T})$ consistent with F (i.e., $g\left(s\left(t\right)\right) \in F\left(\widehat{\theta}\left(t\right)\right)$ for all t) if and only if there exists $f \in F$ that is ex post incentive compatible.

Example

- $\Theta_1 = \{\theta_1, \theta_1'\}$
- $\Theta_2 = \{\theta_2, \theta_2'\}$
- $A = \{a, b, c\}$
- Payoffs

a	θ_2	θ_2'
θ_1	$1, 0$	$-1, 2$
θ_1'	$0, 0$	$0, 0$

b	θ_2	θ_2'
θ_1	$-1, 2$	$1, 0$
θ_1'	$0, 0$	$0, 0$

c	θ_2	θ_2'
θ_1	$0, 0$	$0, 0$
θ_1'	$1, 1$	$1, 1$

Example

Social Choice Correspondence

F	θ_2	θ_2'
θ_1	$\{a, b\}$	$\{a, b\}$
θ_1'	$\{c\}$	$\{c\}$

Example

Agent 1's ex post incentive constraints require

F	θ_2	θ_2'
θ_1	$\{a\}$	$\{b\}$
θ_1'	$\{c\}$	$\{c\}$

This violates agent 2's ex post incentive constraints....

Three Questions

- Robust Mechanism Design: Does there exist a mechanism $\mathcal{M}$ such that for every type space $\mathcal{T}$, there is an equilibrium s of $(\mathcal{M}, \mathcal{T})$ consistent with F, i.e.,
$$g\left(s\left(t\right)\right) \in F\left(\widehat{\theta}\left(t\right)\right)$$
for all t.

- Ex Post Implementation: Does there exist a mechanism $\mathcal{M}$ such that there exists an ex post equilibrium s and every ex post equilibrium is consistent with F,
$$g\left(s\left(\theta\right)\right) \in F\left(\theta\right)$$
for all θ.

Three Questions

- Robust Implementation: Does there exist a mechanism $\mathcal{M}$ such that for every type space $\mathcal{T}$, there exists an equilibrium of $(\mathcal{M}, \mathcal{T})$ and every equilibrium is consistent with F.

Dominant Strategy versus Bayesian Implementation

- Private Values:

$$u_i\left(a,\theta\right) \equiv u_i\left(a,\theta_i\right)$$

- With private values, ex post incentive compatibility is the same as dominant strategies incentive compatibility...

- Ledyard (1978, 1979), Dasgupta, Hammond and Maskin (1979): Fix the direct mechanism on a fixed type space; Bayesian implementation for all priors is equivalent to dominant strategy implementation.

- Criticism: Why fix direct mechanism? Allowing agents to report prior as well, should be able to do better....

- Our answer:
  - in a separable environment, more complicated reports do not help
  - in a non-separable environment, more complicated reports may help

Public Goods

- Green and Laffont (1977): Dominant strategy implementation with budget balance impossible

- d'Apremont and Gerard-Varet (1979), d'Apremont, Cremer and Gerard-Varet (2002): Bayesian implementation possible (for independent types and "most" correlated type spaces)

- No inconsistency as budget balance makes the environment non-separable.

- However, d'Apremont and Gerard-Varet constructed examples showing impossibility of Bayesian implementation on "non-generic" type spaces.

- Genericity of types: Heifetz and Neeman (2004), Dekel, Fudenberg and Morris (2005).

Detail Free and Natural Language

- should be able to express incentive compatibility conditions in natural language that does not refer to artificial "type space"
- $I = 2, \gamma = \frac{1}{2}$
- $v_i = \theta_i + \frac{1}{2}\theta_j$
- so $2v_i - v_j = \frac{3}{2}\theta_i$
- content: common knowledge that each agent i knows $\frac{3}{2}\theta_i$

Collusion

- Laffont and Martimort (1997, 2000)
- Collusion: in situation modelled as private values, in fact agents maximize joint utility, introducing interdependence
- Shill bidders: in situation modelled as private values, in fact agents are copies of each other maximizing joint utility