ROBUST MECHANISM DESIGN

Referee Report by Yang, Yong Hyeon

Mechanism design literature looks at when desirable results are implementable. This requires that every agent chooses one's best strategy given that the other agents choose some strategies, and that those strategies result in desirable outcomes. This is a question of incentive compatibility. One obvious statement is the following. Whenever ex post implementation is possible, it is also possible to interim implement the result (Proposition 1). A question is when the converse is true. This paper provides several results regarding this question. Clearly, this leads to equivalence theorem, which means that under such environments, ex post implementation is possible if and only if interim implementation is possible.

There are several environments that we can consider. One is that no agent knows anything about the other agents' types. Moreover, a social planner does not know anything, either. Nobody can maximize expected payoff based on some prior. In this case, the only possible mechanism is ex post implementation. It is out of the range of this paper. Bergemann and Morris investigate only the case where every agent has prior beliefs over the others' types. The authors divide the cases further. To make clear, it is necessary to tell payoff type from type. Payoff type is agent's characteristic which controls one's utility. Type consists of one's payoff type and prior beliefs over the others' types.

The most standard case is that every agent has the same prior beliefs over payoff types of all agents (common prior payoff type space). This is usually assumed in most of mechanism design literature. The authors have a question to this. Possibly, each agent has different beliefs over payoff types of all agents (payoff type space). This is captured as a type. In some environments, each has the same prior beliefs over types of all agents (comon prior type space), and in other environments, one has different prior beliefs over types of all agents (type space). Which space we consider makes a clear difference as shown in the provided examples.

Example 1 and 2 show that social choice may not be ex post implementable, while it can be interim implemented on all type spaces. Example 3 shows that it may not be interim implementable on all type spaces, while it is interim implementable on all payoff type spaces. But if we restrict our concern only to separable environments, interim implementation on all common prior payoff type spaces imply ex post implementation (Proposition 2). A social choice environment is said to be separable when the outcome has a common component and a private component for each agent, and agent's utility depends only on a common component and one's own private component.

There are several minor results. If social choice is interim implementable only on every full support common prior payoff type space, we can implement the result ex post under some conditions (Proposition 3−4). The authors look at the budget balanced quasilinear case, too, but the result is not positive: In very restrictive cases only, interim and ex post are equivalent (Proposition 5−6). Their contribution is only that they characterize such cases. But the result is too restrictive to be applied in other studies.

The biggest flaw of this paper lies in Proposition 2. Most importantly, it considers only separable environments. The proposition is not if and only if one. When environments are separable, the statement is true, but the converse does not hold. The authors do not completely answer the question they ask themselves, that is, when interim implementation is equivalent to ex post implementation. They suggest part of the answer. In the following sections, they do not extend the discussion in this direction. In this sense, this paper is restrictive.

In this regard, we can pose a question as to how broad separable environments are. Are they usually used in many cases? The authors mention two leading examples. One is when the social choice always picks one outcome, and the other is when the environment is quasilinear. The former sounds good in theory, but in practice it might not be the case. There can be considerably many candidates of solutions, and it is very difficult to choose only one of them. The latter is widely used. But it restricts the environment severely. A practical problem is that transfers among agents do not usually occur, while theory uses quasilinear environment under the implicit assumption that transfers make all the agents better off by choosing the most efficient outcome.

A more primitive critique is that there might be many cases under which environments are not separable. If there are many agents, for example, if each agent does not care about others' allocation like under perfect competition, it is reasonable that enviroments are separable. But if there are not too many agents, it is possible that each agent cares about other agents' private allocation. Especially, many of game theoretical literature investigates nonseparable environments. The result of this paper does not provide any vision in this case.

Nonetheless, it has positive implication that when environments are separable, we only need to look at the most standard case where agents have the same prior beliefs over payoff types. Actually this gives rise to a new question. We only look at the case where agents have some beliefs. So we do not have to consider if it is ex post implementable. Probably we would be able to find some environments under which interim implementability over all kinds of type spaces are equivalent, whether it is ex post implementable or not. Clearly, when environments are separable, we know the result. But we can go further to characterize all such environments. This would be another extension of this paper.