Oracle-Agent Problem

10/22/2018

1 Introduction

The UMA (Universal Markets Access) protocol is meant to provide universal access to financial markets so that everyone might have the opportunity to benefit from sharing risks with others. The way that we will achieve this is by building a platform for trading this risk and an *oracle* that will dictate what market prices are for various forms of risk... In order to support having a non-trivial amount of risk on the UMA protocol, we need to ensure that the oracle has the ability to produce accurate price information, and, in particular, it must be resistant to bribery attacks where an individual attempts to move prices in their favor through corrupt means.

This document describes a model which we use to influence decisions about how to ensure that the cost of corruption (CoC) is higher than the profit from corruption (PfC). The system is economically stable as long as the CoC $>$ PfC. All things equal, it is best if the system generates a large spread between CoC and PfC to ensure that attacking the system requires as high a cost as reasonable, however, raising this spread will likely come at the cost of the efficiency of the system so it is important to think about the externalties imposed by the oracle on the trading protocol.

We begin by describing a static environment that allows us to think about the tradeoffs faced when attempting to raise the CoC-PfC spread and ensuring that the system is difficult to corrupt.

2 Static Environment

In the static world, the trading protocol generates $T = 2\tau N$ revenue and has mN margin that is exposed to being seized if the Oracle reports incorrect information[1].

There is a random state of the world $S \in \{0, 1\}$. There are three types of agents:

[1]For the assumptions that generate these outcomes, see Appendix A

1. The oracle: Responsible for reporting $\hat{S}$. The oracle wants to choose $\hat{S} = S$.
2. The malevolent: Can observe S and receives mN if the oracle reports $\hat{S} = 1 - S$.
3. The individuals: Can observe S and are incentivized by the oracle and the malevolent into voting $x = \{0, 1\}$.

Agents choose whether to truthfully report the state. A truthful report entails $x = S$ and untruthful report is $x = 1 - S$. In order to participate in the vote, agents must purchase the right to vote at price p. In order to entice truthful reporting, the oracle is allowed to make payments to the agents, $\varepsilon(x, X)$ where x is the agent's vote and X number of people who voted $x_i = 0$. This payment is in terms of *rights to vote for tomorrow* which are valued at $p' = \begin{cases} p & \text{if } \hat{S} = S \\ 0 & \text{else} \end{cases}$. The malevolent is also allowed to make payments to the agents, $\tilde{\varepsilon}(x, X, S)$, but can condition on the agent's vote, the number of votes for 0, AND the state.

Agents report to maximize:

$$V(x) = \max_{x \in \{0,1\}} p' \varepsilon(x, X) + \tilde{\varepsilon}(x, X, S)$$

The malevolent would like to corrupt the system as cheaply as possible. The cost at which the malevolent can corrupt the system is given by

$$C^M(S, \varepsilon(x, X)) = \min_{\tilde{\varepsilon}(x,X,S)} \int_i \varepsilon(x_i, X, S)$$
$$\text{subject to} \qquad V(1 - S) > V(S) \quad \text{(Lie Compatiblity)}$$

The oracle would like to report the truth at the minimum cost. It's problem is given by

$$V^O = \min_{\varepsilon(x,X)} \int_i \varepsilon(x_i, X) di$$
subject to
$$p' \int_i \varepsilon(x_i, X) di \leq T + \int_{i \in \text{Voters}} p di \quad \text{(Budget Constraint)}$$
$$V(S) \geq V(1 - S) \quad \text{(Incentive Compatible)} \qquad C^M(S) \geq mN \quad \text{Malevolent Incompatible}$$

3 Appendix A

To generate the assumption of $T = 2\tau N$ and an amount of seizable margin of mN we assume:

1. There are N contracts held in the trading protocol
2. These contracts are symmetric in the sense that each counterparty holds m margin
3. The counterparty of each contract is responsible for paying a tax, τ, in order to be provided access to the oracle's information

4 Old

In this document, the goal is to write down a simple environment that yields insights into how the voting payments should be structured to ensure that the oracle reports the correct state of the world (price). In particular, it focuses on the issue of incentivizing agents to report truthfully in the face of another individual attempting to corrupt the system.

There is still some ironing out to do in the environment. As currently presented, it is not quite a fully described environment.

5 Environment

Consider a static world in which there are an exogenous number of individuals who hold margin in bi-lateral contracts. These contracts rely on a system to provide information for which each contract counter-party pays τ. There are N symmetric contracts where each counter-party holds m margin. Thus there is $2mN$ margin being held in this system.

$$
\begin{aligned}
m &\equiv \text{ margin each counterparty puts in each contract} \\
N &\equiv \text{ number of contracts in system} \\
2mN &\equiv \text{ total margin in entire system} \\
\tau &\equiv \text{ payment made by each counter-party to Oracle} \\
T &\equiv \text{ total payments made to Oracle from contract market } = 2\tau N
\end{aligned}
$$

There is a random state of the world $S \in \{0, 1\}$. There is an agent, "the Oracle", who is in charge of ensuring that the system provides correct information about S. There is also a malevolent agent who would like the system to provide false information about S. The malevolent agent is a counter-party in each of the contracts, and, if the wrong state gets reported, can collect the other party's margin. This means the profit from corrupting the system is PFC $= mN$. Note: This is a worst case analysis. This is the largest incentive that one could have for corrupting this system.

$$
\begin{aligned}
S \equiv & \text{ state of the world} \\
\text{PFC} \equiv & \text{ Profit from corruption } = mN \\
\text{CoC} \equiv & \text{ Cost of corruption}
\end{aligned}
$$

The Oracle sells the right to vote on what S is to a measure one of agents who can purchase this right to vote at price p. Both the Oracle and malevolent agent can provide conditional payments to the voters in order to entice them to tell the truth (or lie). The Oracle, can only condition the payments it makes on an individual's action and the actions made by all other individuals, i.e. $\varepsilon(x_i, x^{-i})$. However, the malevolent agent can condition on an individual's action, the actions made by other individuals, and S, i.e. $\hat{\varepsilon}(x_i, x^{-i}, S)$. Note, the oracle will pay individuals in "rights to vote for tomorrow". These rights to vote for tomorrow have value $p' = p$ if the Oracle announces the correct state and $p' = 0$ if the Oracle is corrupted.

$$
\begin{aligned}
p \equiv & \text{ Cost of purchasing a right to vote} \\
p' \equiv & \text{ Value of the right to vote for tomorrow} \\
\varepsilon(x, x^{-i}) \equiv & \text{ The payment the Oracle makes to voter} \\
\hat{\varepsilon}(x, x^{-i}, S) \equiv & \text{ The payment the malevolent agent makes to voter} \\
\text{PFC} \equiv & \text{ Profit from corruption } = mN \\
\text{CoC} \equiv & \text{ Cost of corruption } = \int_i \hat{\varepsilon}(x, x^{-i}, S)
\end{aligned}
$$

Each of the voting individual knows the state S and maximizes their utility given by:

$$
V = \max_{\text{No Vote,Vote}} \{0, \max_{x \in \{0,1\}} E\left[p' \varepsilon(x, x^{-i}) + \hat{\varepsilon}(x, x^{-i}, S)\right] - p\}
$$

The Oracle chooses $\varepsilon(x, x^{-i})$ to solve

$$
\begin{aligned}
V^O = \min_{\varepsilon(x, x^{-i})} & \int_i \varepsilon(x_i, x^{-i}) di \\
& \text{subject to} \\
& p' \int_i \varepsilon(x_i, x^{-i}) di \leq T + \int_{i \in \text{Voters}} p di \quad \text{(Budget Constraint)} \\
& V(S) \geq V(1 - S) \quad \text{(Incentive Compatible)}
\end{aligned}
$$

The malevolent agent chooses $\hat{\varepsilon}(x, x^{-i}, S)$ in (an attempt?) to solve

$$V^M = \min_{\hat{\varepsilon}(x,x^{-i},S)} \int_i \hat{\varepsilon}(x_i, x^{-i}, S)di$$

subject to

$$\int_i \hat{\varepsilon}(x_i, x^{-i}, S)di \leq mN \quad \text{(Budget Constraint)}$$

$$V(1-S) > V(S) \quad \text{(Lie Compatiblity)}$$

Obviously both of these problems can't be solved. They need to be connected such that the Oracle is minimizing the cost of maintaining the Oracle truthful subject to the fact that the malevolent agent is attempting to corrupt the system... If the malevolent agent weren't there then the Oracle would simply set $\varepsilon(x, x^{-i}) = 1$ and people would report the truth because they would be indifferent. What makes the problem interesting is that the malevolent agent is "tempting" the agents to report falsely.

5.1 Solving the Model

We want to solve two problems:

1. Find optimal payment scheme $\varepsilon(x, x^{-i})$... As written this isn't well specified... Need to sort out what exactly this looks like.
2. Next step would be to take a parameterized approximation to $\varepsilon(x, x^{-i})$ and solve for the Ramsey solution in a dynamic versoin of this world

5.2 Next Steps

Based on the learnings from solving the static world exercise, we (hopefully) can use the same functional form for the ideal (Mirrlees) payout function to extend this model to the dynamic world and solve for the simpler Ramsey problem of the optimal flat tax rate for the dynamic system.