**Thesis Draft**

# Provider Strategic Behavior in the Global Budget System:

# A Theoretical Discussion

Bradley Chen

ytchen@hsph.harvard.edu

Harvard University

June 2010

***Abstract***

Given the increasing pressure of health expenditure growth in recent decades, global budget payment has become an important policy option to contain health care cost. This paper presents a theoretical analysis of the incentive structure of the global budget system as well as the strategic behavioral response of the providers. I argue that global budget is a type of common-pool resources (CPR) and therefore presents the providers a social dilemma, in which the individual and group interests conflict. Under a bounded rationality framework, the analysis shows that as the payment system shifts from fee-for-service to global budget, the providers would first engage in a non-cooperative competition where the providers pay more attention to services with higher price-cost ratios. The resulting volume race would lead to a vicious circle of working harder but declining profits. However, in the awareness of the need for collective action, the providers could, with varied levels of success, overcome "the tragedy of the commons" by engaging in cooperative competition, which poses a different set of financial incentives than the earlier phase and could result in structural changes of the health care market in the long-run. Implications for empirical testing are then discussed.

## *I. Introduction*

Many countries around the world, faced with the increasing burden of health care cost inflation, have attempted to devise various measures to control the increase of health expenditures. Among all, global budget is one of the most effective policy interventions for this purpose since by definition global budget system poses a fixed budget cap on the health sector, beyond which the costs incurred by the providers would not be reimbursed.

Taiwan, following countries like Canada and Germany, implemented a global budget system onto clinics and hospitals in 2001 and 2002, respectively, to effectively control the growth rate of health expenditure. The global budget system in Taiwan is basically a point-for-service system with a fixed expenditure cap. Each treatment and intervention has a certain amount of points serving as the indication of the relative prices. The global budget is set through negotiation at the beginning of the accounting year, and the total budget will be divided by the sum of the points of all services by all providers, generating a **conversion factor (**or **point value)**. That is, the relative values of the treatment are the same, but the true prices are determined *ex post* based on the point value. The total national budget is divided into six regional budgets (Taipei, North, Central, South, Kao-ping, and East) by risk-adjusted population. And the providers are competing within their regions and therefore the point value can be different from region to region and is often the case.

The global budget system in Taiwan presents an interesting design to control health expenditure growth while leaving much *professional* autonomy to the providers as to how to provide services. In this system, as discussed later, providers face a social dilemma in which the individual interest is in conflict with the collective one. That is, the pursuit of self-interest by the rational individuals could lead to a collective failure. The invisible hand might not be working as Adam Smith thought.

Despite years of experiences with global budget system, little is known about how in practice the providers react to such policy intervention. Nevertheless, how providers respond to the incentive structure imposed by the global budget has tremendous implications on quality, efficiency, access, and eventually the long-tem objectives of the health systems, including improvement in population health, financial risk protection and satisfaction.

The need for a better understanding of the global budget system and its impact is particularly relevant now. In many nations, including U.S., China, and many other developing countries, not only has the health expenditure been growing at an alarming rate because of the progress in new technology and the increasing demand for health care, the vigorous efforts in recent years to expand health insurance coverage to the previously uninsured, which has almost become a global movement and a heatedly debated issue, will undoubtedly further speed the escalation of expenditures and necessitate cost control measures. This calls for a careful study of the nature and impact of different cost control options, one of which being global budget system, in order to provide solid evidence for policy-making that will affect the health and well-being of millions.

As a starting effort, this essay will carefully examine the global budget system, using the design in the Taiwanese system as an illustrating example and building upon relevant theories in the health economics and political economics literature. The rest of the article is organized as follows. Section II reviews the literature on global budget and other relevant work on the issue of collective action. I argue that global budget in essence is a form of common-pool resources (CPR). Therefore, the discussion of the provider behavior in global budget system can be informed by the literature on the governance of common-pool resources, which is embedded in the larger theory of collective action. Section III explores the issues on rationality assumption in analyzing collective action and argues for a bounded rationality approach that is more compatible with the reality. Section IV presents a basic non-cooperative competition model of provider behavior in global budget and conducts a simulation exercise. Section V summarizes the theoretical findings and implications on the provider behavior as suggested by the model. Section VI investigates the changes in provider behavior and market structure as the dynamics shift from a non-cooperative toward a cooperative one. Section VII concludes.

## *II. Literature Review*

### *II-A. Theoretical Literature on Global Budget*

Only a few studies have presented in-depth theoretical analyses of the global budget scheme. Fan, Chen and Kan (1998) fist approached the theory behind global budget system by comparing the effects of cost control of two alternative methods: *expenditure target* and *expenditure cap*. The former imposes on the providers a target of health expenditure, above which the fees would be significantly reduced (up to 75%), and the latter is a fixed expenditure limit with the retrospective price-setting mechanism as the one in the Taiwanese health sector. They showed that when the number of providers is large, those facing expenditure cap would produce a larger quantity of health services at strong symmetric Nash Equilibrium than they would do under expenditure target. Mougeot and Naegelen (2005) examined expenditure cap policy on hospital sector and argued that the quality of care and social welfare would be lower with expenditure cap than under optimal cost reimbursement. This finding is rather intuitive given their assumption that quality-enhancing efforts bring costs and disutility to the hospital managers and the hospitals examined are either assumed to be monopolies or competing on *relative* quality. Chen (2006), on the other hand, argued that when the competition is strong enough and the hospitals value both profits and managerial slacks, expenditure cap does not necessarily lead to a lower quality of service than the prospective payments do

Benstetter and Wambach (2006), inspired by the German global budget reform on the ambulatory care sector, modeled the strategic behavior of the physicians facing a fixed expenditure cap. They found that when shifting from a fee-for-service system to a global budget with intermediate expenditure cap, the large number of physicians could suffer from coordination problem, resulting in a "treadmill effect", where all physicians work hard and produce so much that the point value becomes lower. Alternatively, if the physicians coordinate, they could work less hard and enjoy a high point value. The authors suggested that an introduction of price floor (i.e. guaranteed point value) or maximal service quota for each physician by the regulators would alleviate the coordination issue. Similar to Fan, Chen and Kan (1998), they assume individual physician's behavior could not affect the price, implying a free competition model.

### *II-B. Common-Pool Resources and Collective Action*

Despite that only a limited number of studies attempted to provide theoretical understanding of the global budget, a careful look at the design will reveal that it forces the profits for individual provider to be linked with the behavior of others. As rational players would always attempt to increase their share of the budget, individual and group interests conflict: the simultaneous increases in service volume by each provider would automatically result in a lower point value and the *de facto* prices of the services. The dilemma situation presented by the global budget system is by no means unique. I argue that the global budget, as a resource system, is in essence a type of common-pool resources (CPR). The nature of the problem in global budget system is very much similar to those in other natural or man-made CPRs, such as fisheries and groundwater basins. Therefore, the political economics literature on CPR provides extremely helpful tools and insights in advancing our understanding of the global budget system. Before turning to the theoretical analysis, we need to first lay out the definition of CPR and its link with global budget.

### 1. Common-Pool Resources (CPR)

Traditionally, in economics, the classification of goods is based on a dichotomy of public and private. Yet, there is a great degree of heterogeneity among goods in each category. Ostrom. Gardner and Walker (1994) proposed a more useful classification based on two attributes: exclusion and subtractability:

- *Exclusion:* this refers to the ability to exclude or limit the potential beneficiaries or users of the goods once they are provided. The exclusion could be done through physical, economic or legal means.
- *Subtractability:* this refers to the degree whether the goods consumed by one user are available to others. For example, food eaten by one cannot be consumed by others, and fish caught by a fisherman are not available to other fishermen. On the other hand, the availability of things like scientific knowledge is not affected by its use by any individual.

Based on these two attributes, goods can be classified into four types (Table 1). Private goods, which are the "typical" goods in the traditional microeconomic studies, have high subtractability and are easy to exclude. Public goods have the properties that are exactly opposite on subtractability and exclusion. Club goods or toll goods are those that it is easy to exclude or limit its use but have relatively lower subtractability. Goods, the consumption of which usually involves a compulsory toll or fee to eliminate free-riding, belong to this category. The subject of this study, common-pool resources, refers to the goods that are high in subtractability but it is relatively difficult to limit or exclude certain people from consuming them. Many natural resources, like forests, fisheries, and groundwater basins, are typical CPR, but it can also be man-made, e.g. an irrigation system for the farmers in a particular area (Ostrom et al. 1999). Unlike pure public goods, the CPR often face issues of over-crowding or overuse, which I will discuss in details later.

The global budget system, based on the classification presented above, is clearly a CPR. Given the fixed budget, expenditures claimed by one hospital are apparently not available to others. That is, the subtractability is high and the increase of service volume by a provider has a negative externality on the profits of other providers in the same system. At the same time, global budget system is also characterized by its difficulty in exclusion. All contracted clinics and hospitals are entitled to make claims to the health expenditure budget. It is almost impossible to exclude anyone with legal and economic means and given the objective of cost containment, the design of the global budget system also wants to include every provider.

*Table 1. Classification of Goods*

| | | Subtractability | |
|---|---|---|---|
| | | *Low* | *High* |
| Exclusion | *Difficult* | Public Goods | Common-Pool Resources |
| | *Easy* | Club Goods | Private Goods |

## 2. Tragedy of the Commons and Non-cooperative Game Theory

A classic and often cited discussion on CPR is Garrett Hardin's "*The Tragedy of the Commons*" (1968). He used the example of herdsmen raising cattle in an open pasture, where a rational herdsman tries to maximize his profits while the negative effects of overgrazing by adding more animals are borne by all. The only sensible action for the herdsman is to keep putting more animals to the pasture. Hardin pointed out "…this is the conclusion reached by each and every rational herdsman sharing a commons. Therein is the tragedy…. Freedom in a commons brings ruin to all." (p. 1244) A similar pessimistic conclusion was reached even earlier by Mancur Olson in his monumental book- "*The Logic of Collective Action"* (1965). He argued that there exists a *collective action problem* where even if everyone in a group has interests in common, they will not necessarily act collectively to achieve them unless there is some additional imposed incentives or policy interventions to overcome free-riding problems. The rationality at individual level, ironically, give rises to group irrationality.

Many researchers attempted to investigate the social dilemma in CPR using game theory, which is particularly useful to think about the strategic aspects of the behavior. The simplest presentation of collective action problem in game theory is the Prisoner's Dilemma

(PD) in non-cooperative games. The analogy of the collective action problem in the CPR situation to PD yields very pessimistic predictions like those by Olson (1965) and Hardin (1968) that if non-excludable public goods are involved, individuals will not act to achieve their common interest.[1] The "invisible hands" would in fact lead to a pareto-inferior outcome than what could have been achieved. Such reasoning dominated the thinking on collective action and common resources for about two decades. It even became popular metaphor used as the basis for policy prescriptions: a centralized control is required to dictate the behavior of the players and perform sanctions as individuals fail to cooperate. For example, Carruthers and Stoner (1981) argued in their analysis of water resource management in developing countries that "common property resources *require* public control if economic efficiency is to result from their development".

To apply such reasoning to the global budget, we can think of the providers in the global budget system as the prisoners in the jail and simplify their actions into only two options: *defect*, to increase their service volume; and *cooperate*, to coordinate and keep down the service volume to a collectively more efficient level. If such game is played only once, the prediction by the Nash equilibrium is unquestionably defection by all players. Since defection is the dominant best strategy, the theoretical outcome is zero cooperation and every provider increases his or her service volume in attempting to maximize profits. This is essentially what has been found by the few theoretical studies in the current global budget literature (Benstetter and Wambach, 2006; Fan et al. 1996). If the PD game is iterated, the number of possible equilibrium strategies proliferates rapidly as the number of iterations increases. Yet, when we limit the equilibrium to those with *subgame perfection* with *backward induction*, the prediction in the iterated PD is still the combination of the one-shot PD, which is defection all the time.

While the prediction from early non-cooperative game studies on the prospect of collective action seems to be rather pessimistic, some more recent developments offered some optimism. Fudenberg and Maskin (1986) posited that when the number of rounds is uncertain and the players can make a firm commitment of a "grim strategy"— cooperate first but defect for the rest of the rounds if other players defect, it is possible to induce cooperation because of the grave punishment would deter any defection. In a less extreme

[1] The failure of cooperation predicted by game theory was repeated reinforced by laboratory experiments (Gardner and Ostrom, 1991) and empirical studies, e.g. see Moran and Ostrom (2005), Myers and Worm (2003) and Pauly et al. (2002).

version, Robert Axelrod used iterated PD in his famous *The Evolution of Cooperation* (1984) to model the emergence of collective action. He showed in the computer tournament a much more optimistic conclusion: with a significant shadow of the future, the best strategy is unexpectedly a simple 'tit-for-tat', which means cooperation based on reciprocity. This implies that even in an egoist community, cooperation is desirable because of the gains in cooperative actions. Note that these studies provide two accounts of possible mechanisms of how cooperation could be sustained at equilibrium with the preconditions that there exist at least a small yet critical number of people who adopt a strategy that cooperates first and punish if other players defect. However, how do these people come about in the first place? If everyone has similar decision-making processes and objective functions, there is no reason to believe that any one would suddenly decide to change to a strategy that is different from others'. Abreu (1988) noted that in infinitely repeated games, there are multiple possible equilibrium outcomes, even optimal ones. Yet again, the game theory does not provide too much guidance as to which equilibrium to occur and how could individuals move from one optimum to another.

## 3. Structural Determinants of Collective Action

Ostrom (2007) reviewed the literature and summarized the long list of theoretically speculated structural factors that could influence the chance of achieving better outcomes than the non-cooperative equilibrium down to eight major ones- the number of participants involved; whether benefits are subtractive or fully shared; the heterogeneity of participants; face-to-face communication; the shape of the production function; information about the past action; how individuals are linked; and whether individuals can enter or exit voluntarily. Given that benefits of CPR are subtractive and the shape of the production function most concerns with provision problem and is less of an issue in the global budget system, the list of potential factors is down to six. I will briefly discuss them in the following sections.

(i) The Number of Participants

The number of participants in the dilemma situation is one of the first few factors under study in the collective action literature. In fact, Olson (195) devote a lot of efforts analyzing the role of group size in the *Logic*. He noted that large groups will face relatively high cost when attempting to organize for collective action but each member enjoy a smaller share of benefits from such action, while small groups will face relatively low cost, and players in small groups will gain relatively more per capita through successful collective action. Hence, large groups will depart most from the course of action that maximize the group utility unless some player have unequally large share of the benefits so that it would bear the cost of pubic good provision anyway. In contrast, some other theorists have produced predictions opposite to the one from Olson's. Chamberlin (1974) analyzed the effect of group size on the provision levels of public goods and showed that relationship is in fact increasing in many cases. Sandler (1992) also posited that the departure from the group optimum only grows in certain cases as the group size increases. That is, on this question of whether the free-rider problem is more serious among larger groups, the literature is seemingly inconclusive. Nevertheless, the underlying issue is the exact nature of the goods. In many earlier discussions of collective action, the distinction between "pure" public goods and CPR was not made explicit. For example, in Olson's analysis, the benefits to individuals will necessarily decline as the group size is larger, which implies *rivalry*, but this is not the case in Chamberlin (1974). Consequently, if we limit the discussions to CPR, holding all other factors constant, an increase in the number of participants would exacerbate collective action problem and lead to an inferior social outcome.

(ii) Heterogeneity of Participants

Heterogeneity of participants is the other factor that Olson (1965) posited to affect the extent of collective action problem. In his conceptualization, heterogeneity mainly refers to the different relative share of benefits from public goods. He argued that the greater share of the benefits that the collective action would give to a single member, the higher the propensity that this player would be willing to bear the costs and induce the cooperation from "small" players for the collective action to occur. Nevertheless, heterogeneity is not necessarily good for collective action. It could also have another effect that could lead the outcome to the opposite direction. When there is a large variation among the participants in terms of their endowments and payoffs, the transaction cost of coordinating the efforts becomes higher. Hardin (1982) argued that asymmetry of demand could be a serious deterrent to cooperation, especially when some other alternatives or substitutes are available. Libecap and Wiggins (1984) showed in their analysis of the competitive production on common oil pools that heterogeneity of firms could seriously increase the bargaining costs and compromises the success of contractual arrangements of prorationing. Other studies on the impact of inequality in the distribution of wealth also suggest that heterogeneity may make the collective action more difficult because of the reduction of the acceptability of available regulatory schemes (Baland and Platteau, 1999) or the trust and cooperation (Jones, 2004) during the process.

In the global budget system, both of the two effects of heterogeneity are likely. On one hand, some providers who benefit more from the efficient outcome may have higher tendency to bear much of the transaction costs and induce others to come along. On the other, the efforts required for coordination among various types of providers might be quite daunting. Since the common-pool resource in global budget is the money, it is unlikely that providers value the benefits differently. Therefore, the heterogeneity in this setting should lie in the basic attributes of the providers. For hospitals, this means the differences in size of the hospital (endowments), ownership, and whether it is specialized or general hospital (contents of the services and production function).

(iii) Face-to-Face Communication

In typical non-cooperative game theory, there is no communication among the players. Even if communication is possible, the view is that words alone are very weak constraints and do not suffice as credible commitments and make no difference to the likelihood of collective action (Hobbes, 1960). Nevertheless, this is certainly at odds with findings from laboratory experiments, where it has been repeatedly found that communication does matter (Sally, 1995). Communication is also a critical component in the cooperative game theory, which would be discussed later.

(iv) Information about Past Actions

As Axelrod posited, in a repeated social dilemma, if people using cooperative strategy can have enough of encounters with other cooperators, it might be possible to ease the collective action problem. The implication is that if a participant can recognize the type of the player he or she is playing against then a strategy based on reciprocity would be able to gain higher payoffs and survive. One common way to gain such knowledge is to gather information about past actions through some monitoring mechanism. In two-person PD, this is rather straightforward since each one knows their action and can fully infer that of the other. This becomes a bit more complicated in *n*-person situation. Hence, whether monitoring mechanism exists and information about past actions is available would significantly influence whether cooperation can grow (Bendor and Mookherjee, 1987). In a more extreme case, if somehow participants carry with them some symbols, which can be used to recognize their types, Janssen (2008) showed that coupled with the ability to withdraw from the game, coordination is possible even in the one-shot PD.

In a global budget system, the critical information would be the service volume by each provider, which should be available at the insurer for reimbursement purpose. The question is whether such information is made available to the providers so they can recognize the types and any reciprocity or retaliation is possible. Furthermore, if the information on the service volume is not available, it might be possible that providers develop some informal mechanism to observe and learn relevant information.

(v) How Individuals are Linked

Sociologists emphasizing the importance of links and networks among individuals in social dilemmas argued that if individuals can be linked in a way that they draw benefits from particular participants instead of the common pool, the free-rider problem could be overcome (Yamagishi and Cook, 1993). Nevertheless, in the CPR situation, network that could change the structure in this way is improbable. Links and networks could, however, influence the collective action in the global budget, particularly in the Taiwanese system, in another way. In Taiwan, similar to the Japanese system, many hospitals, especially the regional hospitals and medical centers, have strong ties with the medical schools and university hospitals. Department chairmen in the university hospitals oftentimes dictate which hospitals the graduates of the residency programs work upon completion of their training. The networks could promote coordination among hospitals within the same "gang" because of the reduction in transaction costs. Similar phenomena have been well documented in the sociology literature on organizational networks and social networks. For instance, Gulati and Gargiulo (1999) asserted that organizations tend to establish preferential relationships with a high degree of trust with specific partners to reduce risks of opportunism and they often resort to prior alliance to determine future decision of whom they would cooperate and create new alliances with. However, on the other hand, the existence of "gangs" could also be a serious deterrent to cooperation among hospitals belonging to different networks but within the same region, especially when there are remarkable differences in the value systems in different networks. In essence, the networks function as a symbol that identify one's type as discussed previously. To what degree the networks are present and their relative dominance would determine the level of start-up costs that participants have to bear in order to make the collective action happen.

(vi) Entry and Exit

Janssen (2008) showed that the ability to withdraw from a game, coupled with the symbols that revealing the information on trustworthiness, could promote cooperation even in a one-shot PD. Earlier studies by Orbell and Dawes (1991; 1993) and Hauk and Nagel (2001) also argued that the third choice of "opting out" in addition to defection and cooperation could in fact enhance the willingness of cooperation and ease collective action problem. In the global budget, it is much more difficult for any given provider or hospital to exit and terminate the game. Since substantial endowments are required, providers cannot simply move in and out the scheme freely. However, it is still possible that providers have some sort of "outside option" that they can shift more of its business to. For instance, a hospital can decide to rely more on the services that are not covered by the health insurance scheme and hence become less constrained by the global budget payment. Whether a hospital can actually do so and to what degree would depend on what type of hospital it is and the existing resources (e.g. equipments and human resources) it possesses, as well as the related regulations imposed by the government or insurer.

### *III. Prisoners' Dilemma and Incremental Adjustment*

In most global budget systems, including the one in Taiwan, the budget is predominantly historical, which means the global budget, after excluding factors such as aging, population growth and inflation, is similar to the total health expenditure in the fee-for-service period prior to the global budget. With the same levels of spending, why is that the imposition of expenditure cap itself would lead to any behavioral change?

The analogy of Prisoner's Dilemma in the analysis of provider behavior in the global budget system requires that there is an incentive for individuals to increase its service volume. That is, the precondition is that moving away from its current level of service provision must be profitable. Such argument presupposes that the providers are not producing at its Pareto efficient level prior to the global budget. In other words, if we take the total health spending in the fee-for-service (FFS) system to be the global budget, the budget would be lower than it could have been were the providers operating at Pareto efficiency. Such assumption, though oftentimes implicit, can be commonly found in previous research on global budget. For instance, Fan et al. (1990) focused their analysis of physician market in the global budget system in economies where global budget is restraining and smaller than what the actual health spending could have been. Benstetter and Wambach (2006) also argued that "treadmill effect" in the global budget of the German ambulatory sector, in which providers are stuck in a dilemma of working harder but earning less, occurred because of an intermediate budget. Were the budget large enough, the providers could comfortably enjoy the higher prices and profits.

The question is whether it is a reasonable assumption that providers are not at their Pareto optimum under fee for service. This in fact begs a much larger and core issue of rationality. Traditionally, rationality in neoclassical economics means maximization of profits or utilities, taking into account all possible alternatives and their consequences. However, as Herbert Simon (1978) asserted that the complexity and uncertainties, together with the cost of information gathering, make "substantive rationality" in real-life almost impossible. In other words, the rationality is bounded by the constraints of cognitive capacity and the attention of mind itself is the scare resource. How rationality manifests itself is not equating values at the margins but to respond to the perceived local environment and options with actions that are *satisficing*, not maximizing. As Simon put it, "reasonable men reach reasonable conclusions" (1978, p. 14).

Health sector is perhaps even more complicated than other markets for regular goods. It is composed of at least five inter-connected markets (Hsiao, 1995), presenting enormous difficulties to consider all possible alternatives and consequences. Let alone that the health market is afflicted with various issues of information asymmetry and uncertainties, which make any attempt to reach "global rationality" seems implausible. Empirically, there is a wealth of literature in health service research that documents the inefficiencies of hospitals. We also observed that providers constantly make adjustments to expand their operation in the fee for service systems, which we would not have expected to see had the providers reached their optimum, *ceteris paribus.* Consequently, it should be safe and, in fact, more realistic, to start the analysis with the presupposition that in most health markets, providers on average, rarely have reached their Pareto optimum, but rather produce at a local maximum that gives them *satisficing* profits. And in the presence of shift in the balances of economic consequences, or the awareness of them, providers then make incremental adjustments accordingly (Simon, 1978). In other words, what drives human behavior is not a *substantive* rationality of maximizing profits at all times but a *procedural* rationality based on what players perceived in the local environment.

Under this framework, when the global budget is imposed on the health sector, it would become obvious for the providers that they are competing within the same pool of resources and the consequence of their actions is inevitably tied to the action of others. Therefore, they are motivated to ensure that they can earn at least similar, or preferably higher, profits in the new circumstance. Without effective communication and carefully coordinated and enforced efforts, provided it is possible for the providers to earn more profits with the increase of certain type of services, it would not take long before one can conclude with some simple calculations that it is the best response to increase service volume, regardless what others' actions might be- you increase your profits if others maintain business as usual and keep your share of the budget if others also increase. For any rational man in such situation, to increase service volume would be the rational and reasonable thing to do. This does not require him to consider all the observable and unobservable factors and uncertainties in the market and equate things at the margin, nor does he need to understand Nash Equilibrium to take actions. A more daily life sense of rationality would suffice in such case. Providers think locally, not globally. In fact, it only takes some providers to raise their service volumes and others would soon learn they should follow suit. The non-cooperative competition becomes self-perpetuating and the dilemma of working harder but earn less would be self-fulfilling.

Another advantage of the framework of bounded rationality is that it can also readily explain the dynamic learning process where providers become aware of the new consequences and adjust their responses. Without the learning and adjustments, we would not have observed empirically the exacerbation of declining point values both in Germany and Taiwan. Moreover, the phenomenon of shifting from non-cooperative to cooperative behaviors in the Taiwanese global budget system is also compatible with a satisficing, not maximizing, motivation. My qualitative interviews also show that after a few years of experience with global budget, the hospitals, being upset with the declining point values, realized the volume race would only lead to a death spiral and they need to restrain themselves from the temptation of self-interest and take collective action, albeit with varied degrees of success in different regions. Namely, in the awareness of the lose-lose situation, they adapted by slowing down the escalation of service volumes incrementally, which manifests as the gradual return of point values since 2005. This is similar to what Cyert and March (1992) found about the decision-making in business firms, which changed their business practice only when difficulties arose and profits fell below their expectations. Within the maximization framework, in the absence of some effective external intervention, it would be very difficult to explain why would the providers suddenly deviate from the equilibrium and take a drastically opposite course of action.

One puzzle in the collective action literature in economics is the gap between theoretical predictions and the empirical observations. Most early game theoretical analyses on this subject concluded with a rather pessimistic view on the prospect of collective action. Later studies, such as Robert Axelrod's *The Evolution of Cooperation* (1984), did provide some mechanisms to sustain cooperative equilibrium but, as discussed above, required the preconditions of a significant size of individuals carrying certain cooperative strategies. Moreover, these theoretical investigations are also less helpful in explaining how the players can deviate from a non-cooperative equilibrium to cooperative behavior without becoming what Axelrod called the futile "isolated revolt" (p.150). However, in great contrast, laboratory experiments and empirical observations have consistently shown that people do have the capacity to get themselves out of the "tragedy of the commons" and take collective actions without external interventions (Ostrom, 1990). I argue that the observed emergence of collective action and the shift from a non-cooperative toward cooperative competition is not because people suddenly become more altruistic and act in a less rational way, in a narrow definition of self-interest sense. To a great degree, it is because they did not fully realize or value the consequences of non-cooperative action before it actually played out. When the adverse consequences of non-cooperation become so present and eminent, they would undoubtedly catch the attention of each individual and bear much weight in the decision-making. With the negative experience of fierce competition so dominant in their minds, it is not **rational** anymore to behave in the same way. People are rational. The question is what entered into their calculus given the limited attention and the cognitive capacity of the mind.

This is why information matters. And it is not only about the availability of information, but also how such information is presented and hence how they weight in individuals' decision-making is rather critical. Kahneman and Tversky (1973) demonstrated in many instances people pay more attention to recent information without fully consider information on the prior probability. More recently, Branas-Garza, Fatas and Guillen (2006) describe how the initial announcement of possible use of grim-strategy could lead to self-fulfilling prophecy and improve contributions in the repeated public good games. The possibility of someone using grim-strategy is always present, but the act of announcement helps to bring it into everyone's judgment.

Of course, a distinction should be made between the motivation for collective action and the process of achieving one. Even if the players are motivated to pursue a win-win situation with collective actions, there are still factors that could affect the likelihood of success in coordinating such efforts. Trust, reciprocity, and the structural factors affecting the likelihood of collective action in social dilemma discussed in Ostrom (2007) would be very relevant here. In fact, as pointed out by Elinor Ostrom (2007), the structural variables should actually "make [no] difference in the probability of successful collective action if we continue to treat the *model* of rationality…to be a universal *theory* of human behavior" (p.3). Only under a bounded rationality framework, would it start to make sense to talk about how the prospect of individuals avoiding or lifting themselves out of the undesirable equilibrium might differ.

In the next section, I would present a formal model of provider behavior under global budget system and derive the Nash equilibrium strategy as the reimbursement scheme shifts from fee-for-service to global budget. As I argued above, in real life, providers hardly operate at the perfectly efficient level predicted by the much-simplified models. Therefore, I do not intend to argue that providers would produce at the *exact* level as the Nash equilibrium. In stead, the Cournot model below is used to examine the incentive for the providers and the resulting pattern of behavioral change when they are mostly driven by self-interest and engage in non-cooperative competition. I am less concerned with the optimal quantity than the functional nature and quality of their shift in action.

I would also note that the behavioral change identified using the non-cooperative competition model is likely to be short-term and not permanent. As discussed, when the providers begin to suffer from the adverse outcome from the non-cooperation and volume race, there could be an emergence of cooperative behavior. A cooperative game entails a drastically different dynamic and incentive for the providers and would result in another shift in behavior pattern, and possibly, a more profound change in the health market structure, which would be discussed in later part of the thesis.

## *IV. Model*

In this section, I will present a formal model of provider behavior under global budget payment. The much-simplified Cournot model is used to derive the Nash equilibrium strategy and predict the direction of behavioral change as the reimbursement scheme shifts from fee-for-service to global budget. The comparative statics and the simulation results would then be used to generate hypotheses and implications for later empirical studies

### 1. Models

(i) Basic model

Assume *n* homogenous providers in the market with a global budget *B*. Providers compete on quantities, deciding on the total number of points, *k*, they want to provide, which determined the relative share of the budget each provider would receive at the end of the period. Two types of services are provided, with the relative prices decided by the "points" allocated to these service, $p_1$ and $p_2$. Each provider independently decides the quantities of each services, $q_1$ and $q_2$. Further assume that there is no interaction in the production of the two, i.e. no complementarity in production.

The two-product setting was chosen to extend the theoretical models in the literature which mostly base on homogenous single product, so as to reflect the real settings where providers usually provide a mix of different products, e.g. inpatient and outpatient services for the hospitals; or labor-intensive consultations and high-tech lab examinations for the clinics. The absolute and relative demands for the different services are driven by the providers in this model.[2]

The cost function for individual provider takes the following form:

$$C(q_1, q_2)=c_1q_1+c_2q_2+c_{11}q_1^2+c_{22}q_2^2 \qquad C'>0,\ C''>0$$

Then the optimization problem is essentially an exercise of cost-minimization:

$$\min C(q_1, q_2)=c_1q_1+c_2q_2+c_{11}q_1^2+c_{22}q_2^2 \qquad s.t.\quad k = p_1 q_1+ p_2 q_2$$

[2] The assumption of provider-driven demands is not unrealistic in the Taiwanese health system. Firstly, information asymmetry and provider-induced demands have been well documented in health care in a wide range of different settings. Secondly, the universal coverage with relatively low premiums and a comprehensive benefit package means that the patients are very much protected against financial costs and hence the providers have high latitude of inducement.

Taking the first-order condition, one can derive $q_1$ and $q_2$ as a function of $p_1$, $p_2$ and $k$. And the cost function, as a function of $q_1$ and $q_2$, can also be expressed as a function of $p_1$, $p_2$ and $k$. For details of the derivation, please see Appendix I.1.

Now, the profit for each provider ($\pi$) would depend on its total output, $k$

$\pi(k_x) = \bar{p}k_x - c(k_x)$, where $\bar{p}$ is the price,

Under the previous fee-for-service scheme, the real prices of the services were decided by the externally set fee schedule, independent of the actual service volume provided by the actors in the health sector. Maximizing the profit function with respect to $k$ yields $p-c'(k)=0$. Using the implicit function theorem, it follows that $k'(p)=1/c''(k)$, which is positive. Therefore, under the fee-for-service scheme with simplistic assumption of no entry and exit, providers respond to any increase in price by an increase in the service provision volume. This is the *profit maximization argument* in the literature that hypothesizes physicians would equate price with marginal costs.

Under global budget scheme, in contrast, the real prices are no longer exogenously set but determined by the point value, which is equal to dividing the total budget by the total number of points. And the profits now equals point value times the service volume provided minus cost of producing that volume:

$$\pi(k_x) = \frac{B}{\sum_{i=1}^{n} k_i} k_x - c(k_x)$$

The competition here is in essence a Cournot competition where firms compete on, instead of price, the output quantities and set them independently and simultaneously. Now, I consider the Nash equilibrium (NE) outcome, in which each provider chooses their best production strategy given the behavior of others. Due to the homogeneity of the providers in the model assumption, each player should adopt the same strategy given others'. We can derive the output volume ($k^*$) at NE (please see Appendix I.2) and study its comparative statics.

$$k^* = \frac{-n^2(c_1c_{22}p_1 + c_2c_{11}p_2) + \sqrt{n^4(c_1c_{22}p_1 + c_2c_{11}p_2)^2 + 8B(n-1)n^2c_{11}c_{22}(c_{22}p_1^2 + c_{11}p_2^2)}}{4n^2c_{11}c_{22}}$$

If we take partial derivative of equilibrium outputs ($k^*$) over number of providers ($n$) in the market, we have:

$$\frac{\partial k^*}{\partial n} = \frac{B(2-n)(c_{22}p_1^2 + c_{11}p_2^2)}{\sqrt{n^4(c_1c_{22}p_1 + c_2c_{11}p_2)^2 + 8B(n-1)n^2c_{11}c_{22}(c_{22}p_1^2 + c_{11}p_2^2)}} < 0$$

This shows an intuitive result that total service volume (in terms of points) at equilibrium for each provider would decrease as number of players in the market increases. This is because each of them would have a smaller share of the market and would have to lower the service volume as otherwise the point value (hence the *de facto* "price" of the provided service) would become too low to compensate for the marginal cost of production. In addition, $k^*$ might also differ as points assigned to the products ($p_1$ and $p_2$), i.e. the relative values of two services, change. As shown in Appendix I.2, the nature of association between $k^*$ and the relative value of two services is undetermined, depending on other factors, including coefficients in the cost function, the overall budget, and the number of providers. Alternatively, I will conduct simulation exercise to show their relationship in later parts of this section.

Once the total service output at equilibrium ($k^*$) is set, the providers needs to allocate the output quota to each service. Points allocated to service/product 1 ($T_1$) equals:

$$p_1^* q_1^* = T_1 = \frac{2kc_{22}p_1^2 + c_2 p_1^2 p_2) - c_1 p_1 p_2^{\ 2}}{2(c_{11}p_2^2 + c_{22}p_1^2)}$$

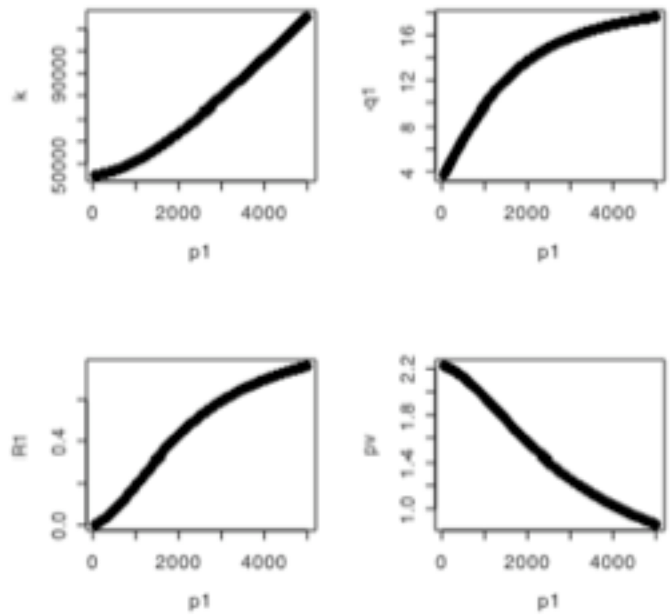
$$\frac{\partial T_1}{\partial p_1} = \frac{p_2^2[2c_{11}p_1(2kc_{22} + c_2 p_2) + c_1(c_{22}p_1^2 - c_{11}p_2^2)]}{2(c_{11}p_2^2 + c_{22}p_1^2)^2}$$

Service output for product 1 at equilibrium, *ceteris paribus*, would be affected by the relative value of the two services ($p_1/p_2$). In fact, the direction of this relationship would be decided by the term- ($c_{22}p_1^2$-$c_{11}p_2^2$) in the above equation. This implies that if the point-cost ration of product 1 is relative higher, providers would shift the provision toward product 1 as $p_1$ and hence the "profitability" for product 1 increases. This is intuitively reasonable since if health providers are profit-maximizers, they can achieve better allocative efficiency by producing more profitable services.

*<Simulation>*

Based on a hospital total cost function reported in Wagstaff and Barnum (1992), I arbitrarily chose the following values for the parameters in the cost function to conduct a simulation exercise: $c_1$=-1000, $c_2$=-6200, $c_{11}$=150, and $c_{22}$=960. Moreover, I assume service

2 to be more profitable and set the service points for the two products at $p_1$=500 and $p_2$=5000. The total budget (B) is set at 10 million ($1x10^7$).



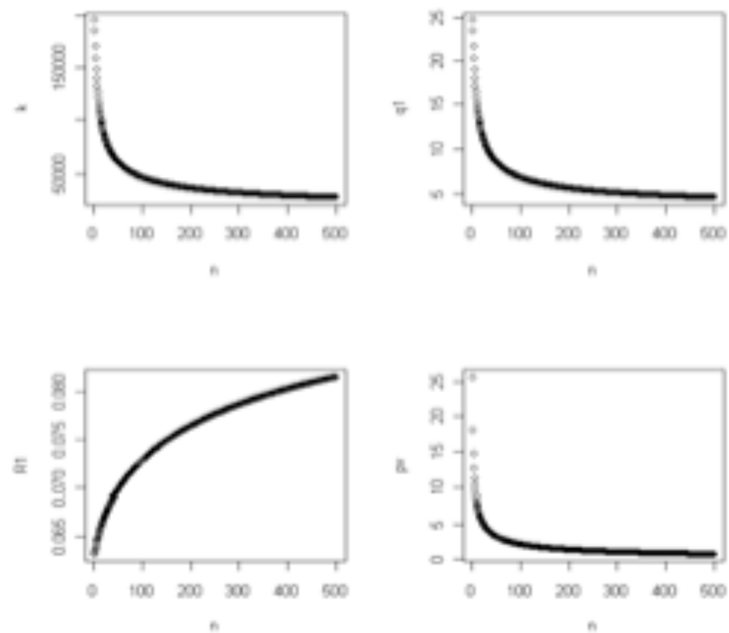

(1) Change in points for product 1 ($p_1$: 50 ~ 5000). Group size (n)=100

The simulation shows that as the point for the product increases, the total service outputs at equilibrium (*k**) increases, at an increasing rate; quantities for service 1 increases (at a decreasing rate); more points are allocated to the service 1 (R1) as its profitability increases, and the overall point value (PV) drops as a consequence of higher service volume.

What is shown here is basically in line with the profit-maximization behavior of the providers: when the profitability of certain product increases, providers would seek higher profits by producing more of that service. More importantly, the simple simulation exercise also illustrates the *collective action problem* that lies at the core of the problem of global budget scheme. When the providers respond to the higher profitability by increased outputs, *collectively*, the higher total service volume would in fact drive down the point value and the *de facto* prices. That is, providers are confronted with a dilemma where the pursuit of individual interest as a perfectly rational individual would ironically compromise the interest of the whole group and everyone within it.

As Olsonian theory of collective action puts a great emphasis on the group size, it would also be of much interest to examine the role of group size, in the following exercise.

(2) Change in group size (n: 2 ~ 500)




As the group size (n) increases, the total output at equilibrium of each individual provider ($k^*$) and quantities of products drop rapidly. This is because, as shown previously, the market share for each individual decreases due to the marginal cost of production. The allocation of points to each product does not have significant differences since the group size should not have an important bearing on the allocation efficiency. Here, the most interesting part is that the point value drops drastically to less than one-fifth when the number increases to 100. This is consistent with the Olsonian theory that larger group would have a more serious collective action problem because in a large group each player does not expect their behavior would have a significant impact on the price of the products and thus focus on increasing their share of the budget. Consequently, the market becomes more and more competitive and price is driven down. As the Cournot Theorem states, in the absence of fixed cost, as the number of firms in the market becomes larger, the prices would approach marginal cost.

(ii) Model with Joint Production Complementarity

In reality, there is often complementarity among a firm's production of different products because of the joint inputs. In our case of health service provision, this is also quite true most of the time. For example, in a hospital, the employed physicians are very likely to provide services both in the inpatient ward as well as ambulatory service; the diagnostic imaging equipments are also used to serve patients in different departments of the hospitals.

To allow for the possibility of complementarity in joint production, I include a joint product coefficient, $c_{12}$, in the cost function;

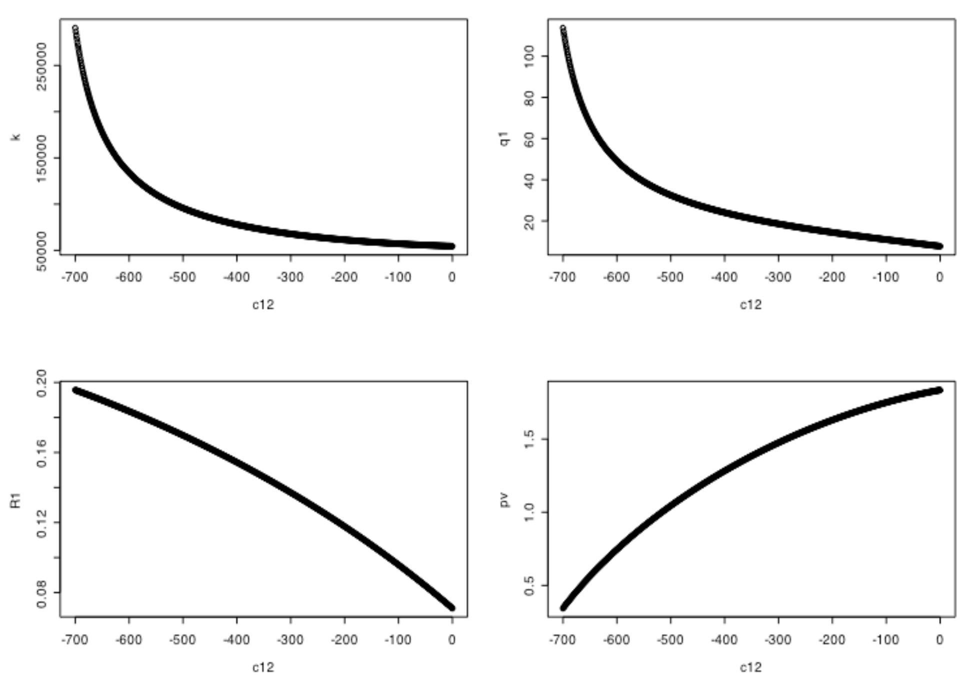
$$C(q_1, q_2)=c_1q_1+c_2q_2 + c_{12}\, q_1q_2 +c_{11}q_1^2+c_{22}q_2^2 \qquad C'>0,\ C''>0$$

The derivation of equilibrium outputs for individual provider ($k^*$) is similar to the above (please see Appendix II for details of derivation), and given the difficulty to determine the nature of comparative statics, simulation exercise is again used.

*<Simulation>*

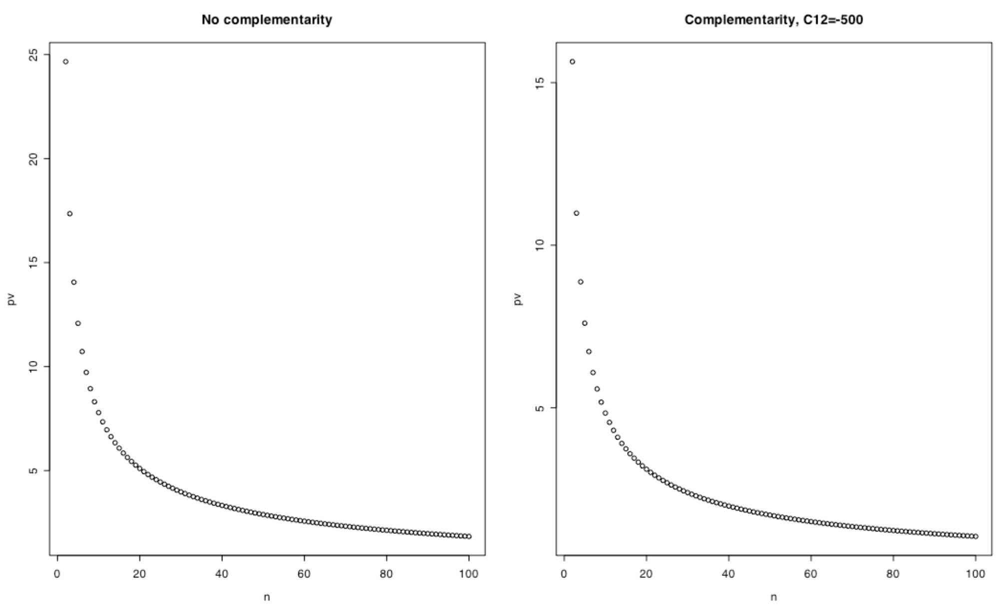
Except the coefficient for production complementarity, all other parameters are kept at the same value.

(1) Change in Degree of Complementarity ($c_{12}$: 0 ~ -700)



Basically, the simulation results are consistent with my expectation. When the productions of two services are independent (i.e. no complementarity, $c_{12}$=0), the providers would shift more of its production toward the more profitable one, in this case, $q_2$. However, when there exists a complementarity between the productions of two products, for example, in the case of some common inputs, it makes sense for the providers to take advantage of it and produce both service 1 and 2. Consequently, we can see from the figure that the higher the complementarity (the more negative of the $c_{12}$ term), the more quantities the provider would produce service 1 and the higher the total service volume. At the same time, the point value falls as a result of the increased production.

(2) Complementarity and Group size




We've learnt that in both of the two previous scenarios, point value is negative associated with the group size. The above figure compares this relationship when there is production complementarity between two services and when there is not. It shows that as the group size increases, the point values are lower and drop faster below 1 when there is production complementarity. That is, the collective action problem is more severe when the provisions of different services are not independent.

(iii) Comparison Between FFS and Global Budget

I have shown, in the previous section, the strategic behavior of the providers under the global budget scheme. Yet, another interesting and critical question is what would be the

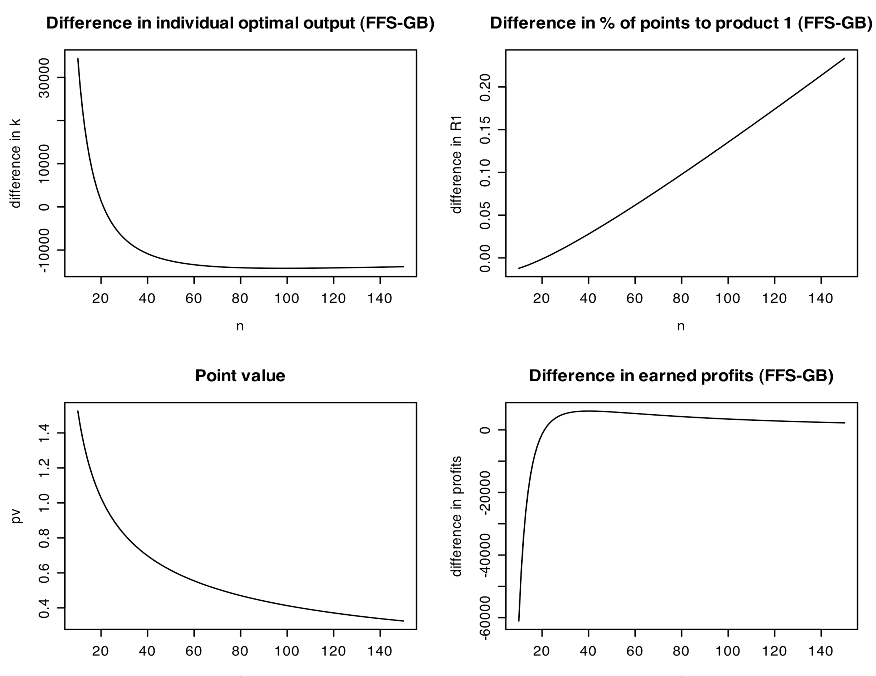
***change*** in providers' service provision when the system shifts from fee-for-service (FFS) to global budget.

To examine the change in the strategic behaviors, I conduct the following simulation exercise to compare FFS and GB schemes:

Assume that under FFS the health sectors spends total amount of B. Further assume that providers are homogenous and have equal market size. Given that providers are fully reimbursed for the service provided, this virtually means point value is fixed at 1. Therefore, the output for each ($k$) is B/n, where n is the number of providers.

Moving from FFS system to a global budget scheme, I take budget B to be the fixed budget in the global budget system. When Taiwan implemented global budget system, the initial budget at year 1 was largely based on the health expenditures in the previous years when the providers were paid on the FFS basis. Then, optimal outputs, point values, and profits are derived in the same way as earlier exercise.

Simulation first starts with the assumption of no complementarity in joint production. Values of the parameters in simulation are set as the following: $c_1$=-1000, $c_2$=-500, $c_{11}$=100, $c_{22}$=300, $p_1$=500 and $p_2$=5000. The number of hospitals is modeled between 10 and 150 since the number of hospitals in each region in Taiwan lies in this range.




The figure above shows that, assuming no entry or exit of providers in the market, moving from FFS to global budget system would result in some interest phenomena in a Cournot competition. First of all, as expected, point value falls when group size increases as a consequence of increasing total output of the health sector. What is worth noticing is that when the number of providers is sufficiently small, at equilibrium, each provider would actually provide at a lower service volume than what it would have been in FFS setting in pursuit of the maximal profits. Point value would be above one, and the profits earned by each provider are large in GB than in FFS scheme. Yet, when the number of providers increases, individual output becomes larger under global budget, point value is down further, and the profits earned become fewer than what could have been earned in the FFS system.

One potential reason that could account for this observation is that when the group size is small enough, each player acknowledges that its behavior would have a discernible impact on the point value and profit margins, and the race on quantities would immediately lead to a vicious circle that makes everyone worst-off. Consequently, the best strategy is to keep the volume down, drive up the point value and make good profits for everyone.

Another observation made here is that shifting to the global budget scheme, the larger the group size and the more serious the collective action problem, the more the providers tend to move their production allocation away from the less profitable service.

Then, similar exercise was conducted assuming the existence of complementarity between the two products. The results show similar findings as in the case of no complementarity: point value drops; compared with FFS, equilibrium output is higher; and profits are lower as the group size increases. The main difference here is that in the case of complementarity, providers tend to produce even more. The point values are consistently below one. This suggests that even when the group size is small and the impact of the players' behavior on the point value should be easily recognized, the tendency to overproduce given the joint inputs would still dominate.

## 2. Results

A summary of the main findings in the modeling and simulation exercise:

- Under the global budget system, the equilibrium output by each provider would decrease as the number of providers increases in the market. However, in aggregate, there would be an increase in total service volume, which drives down the point value and *de facto* prices of the services. This is in line with the Olson's hypothesis that large group would suffer more from the collective action problem.

- Similar to the response to higher prices in the FFS system, the providers would attempt to achieve better allocative efficiency by producing more when the profitability of the services, indicated by the point-cost ratio, increases.
- When there is complementarity in the joint production of services, providers would have even higher tendency to over-produce and collective action problem would be worse in such scenario.
- When the group size is sufficiently small, the providers tend to be more cautious with the production volume because the impact of their behavior is well perceived by the players themselves. In some case, they could even lower their service volume in pursuit of maximized profits.

### 3. Implications and Hypothesis for Empirical Testing

The results of the above exercise have several important implications regarding the change in providers' practice pattern. First of all, moving from FFS to global budget, there is an incentive for the providers to increase their service volume measured by the amount of points. Hence the first hypothesis is:

- *H1: Shifting to global budget scheme would lead to an increase in the intensity of care*

Given the financial incentives, providers would attempt to allocate their service provision toward services with higher point-cost ratios and the costs are likely to include not only capital costs but also labor and time costs of the providers. Empirically, this means we need to test the following hypothesis:

- *H2: When shifting from FFS to global budget, providers would provide proportionally more profitable, less labor intensive services.*

In most cases, the more profitable services are usually the more pricy diagnostics or interventions that usually involve more advanced technology. Therefore, the implementation of global budget could also have a role in the **technology adoption**. For example, some imaging services, such as magnetic resonance imaging, are profitable services themselves and also facilitate other profitable interventions, such as orthopedic surgeries. Therefore, another hypothesis would be:

- *H3: The implementation of global budget would facilitate the adoption of technology and its diffusion in health care.*

In the above analysis, only the quantity is explicitly modeled, yet the change in the composition of the service outputs could also have some implications for the quality of care. We know from the simulation results that, moving from the FFS to global budget system, the production allocation would shift toward profitable services and, similarly, profitable patients. Therefore, one way the implementation of global budget system could impact health care quality is by providers' choice over which services to produce and what kind of patients to treat. That is, the providers could engage in *patient selection*, paying more attention in serving the "profitable" patients (*creaming*) and avoiding the costly ones (*skimping and dumping*). This means that the costly patients, such as those with multiple co-morbidities, might have more difficulty of access to adequate health care in global budget system than in fee-for-service scheme. In Taiwan, an important source of inpatient admission is emergency room. Thus, a testable hypothesis would be for instance:

- *H4: Under global budget, upon visiting to the emergency room, patients with multiple co-morbidities would have difficulty of being admitted and stay longer in the emergency room compared to those with uncomplicated conditions.*

One of the tasks of the modeling exercise is to investigate the relationship between the group size and point value, as an indication of the severity of collective action problem. It is consistently shown that when the number of providers increases, *everything else the same*, the point value will drops and profit margins for the providers go down drastically. An immediate implication from this reasoning is that there could be a movement toward **consolidation** of the market under global budget.

- *H5: The implementation of global budget would facilitate consolidation of the health care market, as reflected in an increase in concentration ratio or Herfindahl index.*

- *H6: There would be higher entry barrier for the providers to enter into the market in the global budget system.*

### *V. Behavioral Change in the Early Phase*

From the modeling and simulation exercise, we learned that in a non-cooperative game, without effective communication and enforceable agreements, the providers have an incentive to maximize their individual profits by increasing their service volume, measured by the total points of service provided by each provider. The only exception is that when the number of providers is sufficiently small, as the consequence of the action by the provider can be readily expected by itself, the strategic response would lead to a lower service volume in global budget than in the fee-for-service system. Therefore, the first behavioral response we should observe is that, with large enough number of providers, the collective action problem would lead to an increase in the treatment intensity when the payment scheme changes from fee-for-service to global budget.

The increase in the treatment intensity in the global budget system would manifest in three ways. First of all, provided that there should not be too much of a change in the number of providers, as the service volume escalates at a faster rate than the increase of the budget, the conversion factor, or point value, would expectedly be on a decreasing trend. This was already observed in the hospital sector as a whole for the early phase of the global budget system, from 2002 to 2004. Theoretically, similar phenomenon should also be observed with the regional point values. Secondly, on the supply side, everything else the same, the implementation of global budget would result in higher service volume by each provider, compared to the level of production under fee-for-service.[3] Lastly, increased service volume by providers would also show up on the demand side. There would be higher frequency or episodes of treatments, higher amount/intensity of services provided to a given individual once he or she is treated by a provider, or both.

The above argument and hypotheses are made under the assumption that there shall not be other demand-side factors affecting the incentive of the patients and utilization of health services is largely affected by the supply side. On the former requirement, it has to be made sure that there is no significant change in enrollees' benefit packages. For the Taiwanese National Health Insurance (NHI), except an increase in the co-payment of less than 5 USD for the hospital ambulatory services in 2005, there has not been any change in the benefit package since the implementation of global budget. The second assumption means that

[3] This of course requires that there is large enough number of providers in the region. For the hospital sector, except the east region, which has only 17 hospitals, All other regions have close to or more than 70 hospitals.

providers have the capability to induce demand. Undoubtedly, it is likely that providers are constrained by legal rules or medical ethics and cannot freely induce *any* and *every thing*. Nevertheless, the capacity to freely induce demand without any binding regulation is not a necessary condition of my argument. Inducing demand does not equate defiance of standards of medical practice. There are too many circumstances where the uncertainties call for the professional judgment of the physicians and present a gray area where standards and protocols are difficult to apply. Therefore, as long as the providers have considerably large latitude to affect the consumption of health services, which is commonly observed empirically in both developed and developing countries, we shall see higher treatment intensity in the global budget scheme.

The increased service volume is consistent with what we have inferred from the qualitative reasoning using the bounded rationality framework and is not particularly surprising. What is more interesting and has a greater implication for the patients is how the providers achieve the higher volume and the change in treatment pattern, if any. And the second main finding from the theoretical exercise is that not only do providers provide more services, they do so by focusing more on the production of the more profitable ones. When providers are paid fee-for-service, they would also attempt to maximize profits through profitable services as well, but there are several distinct features of the incentive in the global budget system.

First, what constitutes higher profitability is different under the two schemes. For instance, say a health service with the cost $c$, and price $p$. Under fee-for-service, what the providers are most concerned about as profit-maximizers is the profit, $\delta$, which is the difference between $p$ and $c$. Providers would be inclined to provide services with larger $\delta$. On the other hand, in the global budget system, what the providers pay attention to, as indicator of profitability, is the price-cost ($p/c$) ratio, or more accurately, the point-cost ratio, rather than the difference between the two. This is suggested by the formal modeling exercise in which the providers are expected to shift more of its production toward one service versus the other if it has a higher price-cost ratio. The intuition behind this phenomenon is that the production of health services, because of the fixed expenditure cap of the global budget, now has an opportunity cost of not providing other alternatives that we don't see in the fee-for-service system. Consequently, in the global budget system, for every unit of production cost, the providers would seek to provide the services with higher profits, hence the attention to p/c ratio.

Another way to look at this problem is that in the non-cooperative competition, the actual price is only know *ex post* because of the point value introduced to automatically discount the price when the volume increases. Such discount factor functions essentially like a fee cut. McGuire (2000) modified the McGuire and Pauly (1991) model and showed that the inducement by the profit-maximizing physicians should satisfy:

$$m_1 x'_1 = m_2 x'_2 = -U_I / U_Y$$

Where $m$ is the profit margin, $x'$ is derivative of service quantity respect to inducement, $U_I$ derivative of utility respect to inducement and $U_Y$ derivative of utility respect to income. What this means is that the response of the physicians to fee cuts would depend on the *income effect*, due to the increase in $U_Y$, which would increase the production of both services; and a *substitution effect*, due to the drop in $m$, which would reduce the production of the service with cut-fee and increase the other. Products with lower p/c ratios would have more of their profits cut proportionally as prices are discounted by the point value. Hence, the substitution effect would be more dominant among products with lower p/c ratios and the providers would shift toward high p/c ratio services.

The distinction between the p/c difference and ratio is seemingly trivial but rather critical. Products with larger p/c difference and ratio could overlap but it is not necessarily always the case. For instance, assume the hospitals provide two types of services: *product x*, with price $1000 and cost $900; *product y*, with price $50 and cost $20. Under fee-for-service, it makes perfect sense the providers to focus on x, which would give the hospital a profit of $100, while product y only has a profit of $30. In contrast, under global budget, assume now the point value is 0.9, product y gives higher profits for the hospitals than x because the actual profits for producing x now turn out to be zero (1000x0.9-900=0) but $25 for y (50x0.9-20=25). A higher price-cost ratio could ensure that the hospitals can still earn significant profits even when the prices are discounted because of the *ex post* determination of the point value.

The implication of the preference for high p/c products in the non-cooperative competition of the global budget is that empirically we should observe an increase in the services with high profit margins indicated by the p/c ratios with the implementation of global budget. In addition, such change should be most prominent among services with relative low price-cost differences but higher price-cost ratios. On the contrary, services with relative high price-cost differences but low price-cost ratios might experience some decline in quantity. In a word, the global budget could potentially reverse some of the bias toward high-cost procedures commonly seen in the FFS systems.

### *VI. The Behavior and Structural Impact of Cooperation*

When the providers start to recognize their loss from the non-cooperative competition and the need to find a collective solution, we would start to see the market becoming more like a cooperative game.

John Nash (1953) defined cooperative games as situations where "the two individuals are supposed to be able to discuss the situation and agree on a rational joint plan of action, an agreement that should be assumed to be *enforceable*." (emphasis added in italic, p.128). The two critical ideas are *communication* and *enforceable* agreement, both of which are not allowed in traditional non-cooperative game settings. Some theorists have put a particular emphasis on the question of enforceable agreements: "Even if [players] are allowed to talk and to negotiate an agreement, this fact will be of no real help if the agreement has little chance of being kept. An ability to negotiate agreements is useful only if the rules of the game make such agreement binding and enforceable" (Harsanyi and Selten, 1988, p.3). The implication following this reasoning is communication should not make any difference when people cannot make credible *ex ante* commitment if there is not an external third-party to enforce the implementation. Nevertheless, empirical evidence has shown individuals facing commons-type dilemma sometimes could develop credible ex ante commitments without the external authority (Ostrom, Walker and Gardner, 1992). Moreover, past research in experimental economics shows that communication, particularly in a face-to-face fashion, does substantially improve the joint outcomes (Sally, 1995). It facilitates the coordination of strategies in both repeated social dilemmas and even one-shot games (Ostrom, Walker and Gardner, 1992). This conflict between theoretical predictions and empirical findings, again, comes from the assumption of substantive rationality. If the players were already making decisions based on the consequences of all possible alternatives, why would communication, whose purpose is exchange of information, make any difference? I would argue, information matters, not in changing people's minds to be cooperative, but in the success of implementing a collective action. When the players reckon that it is actually rational to cooperate with one another, becoming "willing-cooperators", information is useful to eliminate some of the uncertainties around the coordination of efforts and reinforce the trust within the group. External rules and third party are not always necessary because players already have the incentive to cooperate, and the external rewards could sometimes even be harmful as they can have negative effects on the intrinsic motivation (Ostrom, 2000).

In fact, some more recent studies, such as Minhardt (1999), attempted to analyze social dilemma with common-pool resources with cooperative game theory have proved the convexity of such games and that they do not have an empty core, which is the set of all efficient payoff vectors that no coalition suffers a loss. For a non-empty core, the grand coalition can distribute the highest profits to its members. Therefore, in common-pool resources games, there is an incentive for large-scale cooperation and merging economic activities into a cartel or monopoly (Ostmann and Meinhardt, 2007).

If providers do engage in cooperative behavior in global budget, how would that manifest? What does cooperation among the providers imply for the health market? For illustrative purpose, consider the case where providers form one grand coalition. Essentially this means a pure monopoly—there is only one grand coalition of health providers that produce health services and no effective substitutes exist.

Monopolies, in seeking for maximized profits, will take advantage of the market power and select a higher price with lower quantity of production than compared to competitive market. In a global budget system, since the amount of points also determines the point value and thus the *de facto* prices, the monopoly would be expected to lower down the service volume in terms of the total points than what it would be in a non-cooperative game situation. The exact total service volume produced by the monopoly, in this case, health providers as a group, would be determined by equaling marginal cost with marginal revenue. Each individual provider then, through bargaining and agreement, is allocated a "quota" of points as the service volume constraint. However, in practice, the service output is likely to be higher than what it would be with a pure monopoly where the service producers can freely sets the price discrimination and volume desired. As the service volume decreases, the point value, which is calculated every quarter, would rise and signals to the Bureau of National Health Insurance that providers have limit their service in exchange for a significant amount of profits. This would hurt the providers' ability to negotiate for a larger global budget for the next accounting year (personal communication, Vice-director, Bureau of National Health Insurance, 2010).

For the individual provider, each on is given an explicit or implicit quota of service volume measured by total amount of points. Assuming profit-maximization, with fixed revenue, what drives the behavior is likely to be cost-minimization. This in essence is similar to an individual budget for the hospital. From theory and empirical experiences, we all know that individual budget would lead to a drop in treatment intensity. For example, the expenses for each episode could be lower and the duration of stay of the hospital admission might decrease. Moreover, there could also be an increase in the waiting lines for particularly procedures and services. Secondly, on the patient selection, one of the potential quality hazards from non-cooperative competition in the global budget system is the provider could disfavor more costly patients, such as those with multiple chronic diseases. Whether it is less or more serious in cooperative game setting is ambiguous. On one hand, since what drives provider behavior is the price-cost difference, if large profit is possible by prescribing more services for these patients, the providers still have to incentive to take these patients. On the other hand, expensive services for the costly patients can crowd out others and decrease physicians' latitude in practice, and therefore can be un-welcomed. Consequently, whether there would exist selection against these patients would depend on which effect dominates. One factor that could affect this dynamic is the type and size of the hospitals. That is, for instance, in a large teaching medical center, the crowding out effect would tend to be smaller, in comparison with small local hospitals that have a relative smaller volume limit. Hence, one hypothesis that one could further test empirically with data is whether there is a negative relationship between the scale of the hospital and the degree of patient selection.

At the inter-provider level, there are a few things one could reasonably expect to see. Firstly, the cooperative behavior necessitates the establishment of institutions to govern the process, including determining the optimal total service volume, allocating the "quota" fairly to the individual player, monitoring the output, and if necessary, sanctioning the "defectors". By institution, I mean the "written and unwritten rules, norms, and constraints" that are created and devised to reduce uncertainty and govern the environment (Menard and Shirley, 2008). The sanctioning mechanism, as shown in the laboratory experiments, are particularly important in sustaining the cooperation in repeated games, through deterring defection as well as encourage cooperation by minimizing uncertainty of free-riding. The sanctioning mechanisms can take several forms and do not have to be formal rules. Depending on the circumstances, providers may resort to the formal rules of the professional association if there is a tradition of strong presence of the association and stable trusting relationship. However, the unwritten norms could even be more effective. In medicine, particularly in Taiwan where the medical profession is a very closed circle and there exists a strong sense of loyalty because of its Japanese legacy from the colonial period, the reputation among the peers is highly valued. In such case, "gossiping" could become the most effective deterrent for defection. Whatever the form, these institutions are likely to come from the health providers themselves, but it could include some external rules imposed by the government or external agencies. Nevertheless, as shown in previous empirical experiences, the external rule might possibly compromise the level of cooperation and thus the capacity of self-governance (Ostrom, 2000). Whether this is also true in health market remains to be tested empirically.

Secondly, the cooperation and coalition can also take the form of alliance or even increasing merging activities. This is especially plausible when the providers see a need to reduce uncertainties and costs in implementing collective action. Ronald Coarse (1937) argued that the existence of firm is a result of trying to reduce the transaction costs associated with the interaction with the external markets. By the same token, it is reasonable to expect that a potential mechanism to eliminate some complexities of coordinating with a large number of actors is to bring them into the internal structure. [4]

Thirdly, there could also be a trend toward specialization, especially for the hospitals. With the guiding principle of restraining volume increase, the hospitals could continue business as usual and simply control the amount of service points in general; or alternatively, they can develop specialization and reputation for different services. Specialization serves the providers well in two aspects. One, the reputation and, as a result, increased patient number for particularly conditions, would allow the providers to enjoy economy of scales and bring down the costs. Furthermore, specialization will help the providers gain prestige

[4] A related but distinct phenomenon one might observe is the merging activities across regional as an attempt by the provider to increase budgets without intensifying the competition within its region.

and at the same time operate in their own "territory" without stepping on others toes. This could further ease the sense of competition and tension, which could have significant negative impact on the core relationship of trust and reciprocity essential for cooperation. In the Appendix III, I also explore the relationship between cooperation and specialization with formal modeling, which shows that the larger number of different services hospital provide, the more difficult to sustain collusion and in situation where there is a high degree of specialization, the cartel is more stable.

Lastly, there is also a strong incentive for the health providers to increase the entry barrier into the market. Newcomers in the market pose threat and uncertainty to the existing coalition. The coalition has to either spend extra costs to negotiate the new providers into coalition, or sanction them if they are not willing to cooperation. Moreover, as the budget is fixed, any additional member, if the coalition is sustained, results in a smaller share of the pie for every one. In my previous discussion on non-cooperative competition, one of the implications derived from the modeling exercise is also an increased entry barrier into the market, so certainly, cooperative or not, compared with the fee-for-service system, the global budget system would lead to a higher entry barrier.[5] At the same time, given the existence of coalition and the apparent threat posed by the newcomers, the entry barrier when there exists a grand coalition of providers is likely to be even higher.

## *VII. Conclusions*

To sum up, I have argued in this essay that global budget, as a resource system, is a type of common-pool resources. Therefore, the understanding of provider behavior in the global budget system can be informed by previous studies in the political economics literature on collective action. Earlier research on the social dilemma of collective action drew on theory of non-cooperative competition, as well as the analogy of Prisoner's Dilemma, and presented rather pessimistic conclusions that individual rationality would give rise to group irrationality. Nevertheless, the theoretical findings from non-cooperative game theory are inconsistent with what has been found in many laboratory and empirical studies, where individual are capable of overcoming the tragedy of the commons through collective

[5] This of course again assumes that the providers can induce demand or there is unmet demand in the market prior to the entry of new providers. Otherwise, if the market is "saturated", there will also exist entry barrier for any additional health providers even in the fee-for-service system.

action, even without external interventions. Moreover, the rationality assumption fails to explain why the providers, were they already producing at Pareto efficiency in the fee-for-service scheme, would change their behavior with an expenditure cap at a similar level of spending. Nor could it give us insights on how the providers could move from one equilibrium to another or shift from non-cooperative behavior to cooperation. I argue that health providers operate at local maximum that gives them *satisficing* profits and their behavior is driven by *procedural rationality,* which is based on their perceived local environment. That is, providers change their behavior in the presence of shift in the balances of economic consequences, or the awareness of them.

Consequently, when the global budget is introduced, providers would become aware that their consequences of action are now inter-connected and it is rational to increase their service volume in the pursuit of individual profit. The simulation using a non-cooperative competition model, which is used to investigate the direction of change in behavior rather than its exact magnitude, also shows an increase of treatment intensity when the reimbursement moves from fee-for-service to global budget. Furthermore, it also suggests that providers achieve the higher service volume by shifting more of their production toward services with higher price-cost ratios. The attention to service volume, however, would lead to a vicious circle of declining point value and ultimately lead to an awareness of the need for collective action. With varied level of success, the providers could move toward a cooperative competition, which poses a completely different set of incentive for the providers and trigger another shift in behavior and possibly structural change of the market, including emergence of alliance formation and merging activities.

The theoretical discussion and simulation exercise presented above has the same limitations as many economic models. The providers were still assumed to be profit-driven, though it is satisficing rather than maximization. However, empirical research has noted that doctors do care about professional ethics and include patient welfare in their objective function (Mattke 2001). How much providers tradeoff between profits and patient welfare, if any conflict arises, would vary from place to place and is an empirical question unanswered. The theoretical model here does not at all try to settle this debate but to serve as a benchmark scenario to which empirical studies can be compared. In addition, as more and more of the financial risk is being shared by the providers, the role of economic incentive is likely to become more important, not less.

Another limitation of the above discussion is that it cannot incorporate and exhaust all possible contextual factors. For example, no one would disagree that empirical the attitude of the leaders, interaction and relationship between various stakeholders and culture play significant role in decision-making and behavior. Nevertheless, these are all factors that are hard to quantify and include in economic models. Professions and organizations all have their own distinctive personality and culture with a long history that could be traced back to how they began in the first place. This issue would have to be addressed by qualitative studies in empirical research.

The theoretical discussion above has generated several hypotheses that could be put to test. The National Health Insurance in Taiwan established an information system where providers have to submit claims electronically for reimbursement purpose. Future research could take advantage of such comprehensive electronic data to conduct empirical testing and further validate the theoretical predictions presented in this essay.

## Reference

Abreu, D. 1988. "On the Theory of Infinitely Repeated Games with Discounting." *Econometrica* 56(2): 383-396

Axelrod, R.M. 1984. *The Evolution of Cooperation*. New York: Basic Books,

Baland, J-M. and Platteau, J-P. 1999. "The Ambiguous Impact of Inequality on Local Resource Management." *World Development* 27(5): 773-788

Bendor, J. and Mookherjee, D. 1987. "Institutional Structure and the Logic of Ongoing Collective Action." *The American Political Science Review* 81(1): 129-154.

Benstetter, F., and A. Wambach. 2006. "The treadmill effect in a fixed budget system." *J Health Econ* 25 (1): 146-69.

Brañas-Garza, P., Fatas, E. and Guillén, P. 2006. "Inducing a Self-Fulfilling Prophecy in Public Goods Games." No 06/01, ThE Papers, Department of Economic Theory and Economic History of the University of Granada., Available at http://econpapers.repec.org/RePEc:gra:wpaper:06/01. Accessed: May 21, 2010

Carruthers, I., and R. Stoner. 1981. "Economic Aspects and Policy Issues in Groundwater Development." Washington D.C.: World Bank.

Chamberlin, J. 1974. "Provision of Collective Goods As a Function of Group Size." *The American Political Science Review* 68 (2):707-16.

Chen, W-Y. 2006. *National Health Insurance in Taiwan: Welfare Analysis and Hospital Competition*. Oregon State University, Ann Arbor.

Coarse, R.H. 1937. "The nature of the firm." *Economica* 4:386-405

Cyert, MC. And March, J.G. 1992. *Behavioral Theory of the Firm*. Wiley-Blackwell Press.

Fan, C-P, Chen, K-P and Kan, K. 1998. "The design of payment systems for physicians under global budget- an experimental study." *Journal of Economic Behavior & Organization* 34:295-311.

Fudenberg, D. and Maskin, E., 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54(3), 533-554.

Gardner, R. and Ostrom, E. 1991. "Rule and Games." *Public Choice* 70:121-149

Gulati, R. and Gargiulo, M. 1999. "Where Do Interorganizational Networks Come From?" *American Journal of Sociology* 104(5): 1439-1493

Hardin, G. 1968. "The Tragedy of the Commons." *Science* 162 (5364):1243-8.

Hardin, P.R. 1982. *Collective Action,* RFF Press.

Harsanyi, J.C. and Selton, R., 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: The MIT Press.

Hauk, E. and Nagel, R. 2001. "Choice of Partners in Multiple Two-Person Prisoner's Dilemma Games: An Experimental Study." *The Journal of Conflict Resolution* 45(6): 770-793.

Hobbes, T. 1960. *Leviathan*. Oxford: Basil Blackwell.

Hsiao, W. 1995. "Abnormal economics in the health sector." *Health Policy* 32(1-3): 125-139

Janssen, M.A. 2008. "Evolution of cooperation in a one-shot Prisoner's Dilemma based on recognition of trustworthy and untrustworthy agents." *Journal of Economic Behavior & Organization* 65(3-4): 458-471.

Jones, B.D. 2004. “Wealth-based Trust and the Development of Collective Action.” *World Development* 32(4), 691-711

Kahneman, D. and Tversky, A. 1973. “On the psychology of prediction.” *Psychological Review* 80(4): 237-251

Libecap, G.D. and Wiggins, S.N., 1984. “Contractual Responses to the Common Pool: Prorationing of Crude Oil Production.” *The American Economic Review* 74(1), 87-98.

Mattke, S. 2001. *The Effect of Financial Incentives and Institutional Arrangements on Provider Behavior*. Department of Health Policy and Management, Harvard University, Boston.

McGuire, T.G. 2000. “Physician Agency.” In: Newhouse, J.P., Culyer, A.J. (Eds), *Handbook of Health Economics*. North Holland, Amsterdam.

McGuire, T.G. and Pauly, M.V. 1991. “Physician response to fee changes with multiple payers.” *Journal of Health Economics* 10(4): 385-410

Meinhardt, H.I. 1999. “Common Pool Games are convex games.” *Journal of Public Economic Theory* 2: 247-270

Menard, C. and Shirley, M.M. 2008. *Handbook of New Institutional Economics.* Dordrecht, Netherlands: Springer

Moran, E., and Ostrom, E. 2005. *Seeing the Forest and the Trees: Human-Environment Interactions in forest Ecosystems.* Cambridge, MA: MIT Press.

Mougeot, M. and F. Naegelen. 2005. "Hospital price regulation and expenditure cap policy." *J Health Econ* 24 (1):55-72.

Myers, R.A., and Worm, B. 2003. “Rapid worldwide depletion of predatory fish communities.” *Nature* 423: 280-283

Nash, J. 1953. “Two-Person Cooperative Games.” *Econometrica* 21(1): 128-140.

Olson, M. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, M.A.: Harvard University Press.

Orbell, J. and Dawes, R.M. 1991. A "Cognitive Miser" Theory of Cooperators' Advantage. The *American Political Science Review* 85(2): 515-528.

Orbell, J.M. and Dawes, R.M. 1993. “Social Welfare, Cooperators' Advantage, and the Option of Not Playing the Game.” *American Sociological Review* 58(6): 787-800.

Ostmann, A. and Meinhardt, H. 2007. “Non-binding agreements and fairness in commons dilemma games.” *Central European Journal of Operations Research* 15(1): 63-96.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* New York: Cambridge University Press.

Ostrom, E. 2000. “Crowding out Citizenship.” *Scandinavian Political Studies* 23(1), 3-16.

Ostrom, E. 2007. “Collective Action and Local Development Processes.” *Sociologica*. 3. doi: 10.2383/25950

Ostrom, E., Burger, J., Field, C.B., Norgaard, R.B. and Policansky, D. 1999. “Revisiting the Commons: Local Lessons, Global Challenges.” *Science* 284(5412): 278-282

Ostrom, E., Gardner, R. and Walker, J. 1994. *Rules, Games, and Common-Pool Resources,* University of Michigan Press.

Ostrom, E., Walker, J. and Gardner, R. 1992. “Covenants With and Without a Sword: Self-Governance is Possible.” *The American Political Science Review* 86(2): 404-417.

Pauly, D., Christensen, V., Guenette, S., Pitcher, T.J., Sumalia, R., Walters, C.J., Watson, R. and Zeller, D. 2002. “Towards sustainability in world fisheries.” *Nature* 418: 689-695

Sally, D. 1995. “Conversation and Cooperation in Social Dilemmas: A Meta-analysis of Experiments from 1958-1992.” *Rationality and Society* 7:58-92.

Sandler, T. 1992. *Collective Action: Theory and Applications.* Ann Arbor: University of Michigan Press.

Simon, H.A. 1978. “Rationality as Process and as Product of Thought.” *The American Economic Review* 68(2): 1-16

Yamagishi, T. and Cook, K.S. 1993. “Generalized Exchange and Social Dilemmas.” *Social Psychology Quarterly* 56(4): 235-248.

# Appendix

### Appendix I.1 Derivation of the Basic Model

The optimization problem is:

$$\min C(q_1, q_2)=c_1q_1+c_2q_2+c_{11}q_1^2+c_{22}q_2^2 \qquad s.t. \quad k = p_1 q_1+ p_2 q_2$$

Set $L= c_1q_1+c_2q_2 +c_{11}q_1^2+c_{22}q_2^2-\lambda( p_1 q_1+ p_2 q_2-k)$

F.O.C

$$\frac{\partial L}{\partial q_1} = c_1 + 2c_{11}q_1 - \lambda p_1 = 0 \ldots\ldots\ldots\ldots\ldots(1)$$

$$\frac{\partial L}{\partial q_2} = c_2 + 2c_{22}q_2 - \lambda p_2 = 0 \ldots\ldots\ldots\ldots(2)$$

$$\frac{\partial L}{\partial \lambda} = p_1q_1 + p_2q_2 - k = 0 \ldots\ldots\ldots\ldots\ldots(3)$$

From (1), $\frac{c_1 + 2c_{11}q_1}{c_2 + 2c_{22}q_2} = \frac{p_1}{p_2}$ (2): ; and (3) $k = p_1 q_1+ p_2 q_2$

$$\rightarrow q_1 = \frac{1}{2}\frac{[c_2p_1p_2 - c_1p_2^2 + 2kc_{22}p_1]}{(c_{11}p_2^2 + c_{22}p_1^2)}$$

$$\rightarrow q_2 = \frac{1}{2}\frac{[c_1p_1p_2 - c_2p_1^2 + 2kc_{11}p_2]}{c_{11}p_2^2 + c_{22}p_1^2}$$

$$C(q_1,q_2) = f(p_1,p_2,k) =$$
$$\frac{c_{22}(c_2p_1^{\;2} - 2kc_{11}p_2 - c_1p_1p_2)^2}{4(c_{22}p_1^2 + c_{11}p_2^2)^2} + \frac{c_{11}(c_1p_2^{\;2} - 2kc_{22}p_1 - c_2p_1p_2)^2}{4(c_{22}p_1^2 + c_{11}p_2^2)^2} -$$
$$\frac{c_2(c_2p_1^{\;2} - 2kc_{11}p_2 - c_1p_1p_2)}{2(c_{22}p_1^2 + c_{11}p_2^2)} - \frac{c_1(c_1p_2^{\;2} - 2kc_{22}p_1 - c_2p_1p_2)}{2(c_{22}p_1^2 + c_{11}p_2^2)}$$

### Appendix I.2 Derivation of Nash Equilibrium Output

$$\pi'(k_x) = \frac{B}{\sum_{i=1}^{n} k_i} - \frac{Bk_x}{(\sum_{i=1}^{n} k_i)^2} - c'(k^*) = \frac{B}{nk} - \frac{Bk}{(nk)^2} - c'(k^*) = \frac{B}{nk}(1-\frac{1}{n}) - c'(k^*)$$

$$c'(k^*) = \frac{c_1c_{22}p_1 + c_{11}(2k^*c_{22} + c_2p_2)}{c_{22}p_1^{\ 2} + c_{11}p_2^{\ 2}} = \frac{B(n-1)}{n^2k^*}$$

$$\rightarrow k^* = \frac{-n^2(c_1c_{22}p_1 + c_2c_{11}p_2) + \sqrt{n^4(c_1c_{22}p_1 + c_2c_{11}p_2)^2 + 8B(n-1)n^2c_{11}c_{22}(c_{22}p_1^2 + c_{11}p_2^2)}}{4n^2c_{11}c_{22}}$$

$$\Rightarrow \frac{\partial k^*}{\partial n} = \frac{B(2-n)(c_{22}p_1^2 + c_{11}p_2^2)}{\sqrt{n^4(c_1c_{22}p_1 + c_2c_{11}p_2)^2 + 8B(n-1)n^2c_{11}c_{22}(c_{22}p_1^2 + c_{11}p_2^2)}} < 0$$

$$\Rightarrow \frac{\partial k^*}{\partial p_1} = \frac{-c_1 + \dfrac{n^2c_1^2c_{22}p_1 + 8B(n-1)c_{11}c_{22}p_1 + n^2c_1c_2c_{11}p_2}{\sqrt{n^4(c_1c_{22}p_1 + c_2c_{11}p_2)^2 + 8B(n-1)n^2c_{11}c_{22}(c_{22}p_1^2 + c_{11}p_2^2)}}}{4c_{11}}$$

**Appendix II. Model with Complementarity**

The cost function is now:

$$C(q_1, q_2)=c_1q_1+c_2q_2 + c_{12}\, q_1q_2 + c_{11}q_1^2+c_{22}q_2^2 \qquad C'>0,\ C''>0$$

Then the optimization problem is essentially:

$$\min C(q_1, q_2)=c_1q_1+c_2q_2+ c_{12}\, q_1q_2 + c_{11}q_1^2+c_{22}q_2^2 \qquad s.t.\quad k = p_1\, q_1+ p_2\, q_2$$

Set $L= c_1q_1+c_2q_2 + c_{12}\, q_1q_2 + c_{11}q_1^2+c_{22}q_2^2-\lambda( p_1\, q_1+ p_2\, q_2-k)$

F.O.C

$$\frac{\partial L}{\partial q_1} = c_1 + c_{12}q_2 + 2c_{11}q_1 - \lambda p_1 = 0 \ldots\ldots\ldots\ldots\ldots(1)$$

$$\frac{\partial L}{\partial q_2} = c_2 + c_{12}q_1 + 2c_{22}q_2 - \lambda p_2 = 0 \ldots\ldots\ldots\ldots(2)$$

$$\frac{\partial L}{\partial \lambda} = p_1q_1 + p_2q_2 - k = 0 \ldots\ldots\ldots\ldots\ldots(3)$$

From (1), (2): $\dfrac{c_1 + c_{12}q_2 + 2c_{11}q_1}{c_2 + c_{12}q_1 + 2c_{22}q_2} = \dfrac{p_1}{p_2}$

And (3) $k = p_1\, q_1+ p_2\, q_2$

$$\rightarrow q_1 = \frac{1}{2}\frac{[c_2p_1p_2 - c_1p_2^2 + 2kc_{22}p_1 - kc_{12}p_2]}{(c_{11}p_2^2 - c_{12}p_1p_2 + c_{22}p_1^2)}$$

$$\rightarrow q_2 = \frac{1}{2}\frac{[c_1p_1p_2 - c_2p_1^2 + 2kc_{11}p_2 - kc_{12}p_1]}{c_{11}p_2^2 - c_{12}p_1p_2 + c_{22}p_1^2}$$

$$C(q_1,q_2) = f(p_1,p_2,k) =$$
$$-\frac{k^2c_{12}^2 - 4kc_1c_{22}p_1 + c_2^2p_1^2 - 2c_1c_2p_1p_2 + c_1^2p_2^2 + 2kc_{12}(c_2p_1 + c_1p_2) - 4kc_{11}(kc_{22} + c_2p_2)}{4(c_{22}p_1^2 - c_{12}p_1p_2 + c_{11}p_2^2)}$$

Then, set $\pi'(k_x)=0$, $\pi''(k_x)<0$

$$\pi'(k_x) = \frac{B}{\sum_{i=1}^{n} k_i} - \frac{Bk_x}{(\sum_{i=1}^{n} k_i)^2} - c'(k^*)$$

Now, consider the Nash equilibrium:

$$\pi'(k_x) = \frac{B}{\sum_{i=1}^{n} k_i} - \frac{Bk_x}{(\sum_{i=1}^{n} k_i)^2} - c'(k^*) = \frac{B}{nk} - \frac{Bk}{(nk)^2} - c'(k^*) = \frac{B}{nk}(1-\frac{1}{n}) - c'(k^*)$$

$$c'(k^*) = \frac{2(c_1c_{22}p_1 + c_{11}(2k^*c_{22} + c_2p_2)) - c_{12}(c_2p_1 + c_1p_2) - k^*c_{12}^2}{2(c_{22}p_1^2 - c_{12}p_1p_2 + c_{11}p_2^2)} = \frac{B(n-1)}{n^2k^*}$$

$$\rightarrow k^* = \frac{n^2(c_2(c_{12}p_1 - 2c_{11}p_2) + c_1(c_{12}p_2 - 2c_{22}p_1))}{2n^2(4c_{11}c_{22} - c_{12}^2)} +$$

$$+\frac{\sqrt{n^4(c_2(c_{12}p_1 - 2c_{11}p_2) + c_1(c_{12}p_2 - 2c_{22}p_1))^2 - 8B(n-1)n^2(c_{12}^2 - 4c_{11}c_{22})(c_{22}p_1^2 + c_{11}p_2^2 - c_{12}p_1p_2)}}{2n^2(4c_{11}c_{22} - c_{12}^2)}$$

### Appendix III. Specialization and Cartel Sustainability

To assess the relationship between specialization and cartel stability, I assume the hospital face the following profit function;

$$\pi(x) = px - c(x) = px - mkx$$

The cost function is affected by m, which is the number of service varieties produced by a given hospital. What this suggests is that to produce a high number of services types are costly and there is a high scale economy of producing specialized services.

The hospital strategies in the infinitely repeated game are 'grim strategies', where each hospital sets the collusive output each period as long as no defection has occurred in the past, otherwise it sets the static NE output forever. Following the literature, the collusion is sustainable as a SPNE as long as:

$$\delta \geq \delta^* = \frac{\pi^D - \pi^C}{\pi^D - \pi^N}$$

Where D superscript denotes defection, C as collusion, and N as non-cooperative outcome.

In the collusion situation, the profit for each hospital would have to be B/n, B being the total budget and n the number of hospitals in the region. In the non-cooperative situation.

$$\pi_N(x) = px - c(x) = px - mkx = \frac{B}{\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}} x - mkx$$

Using first order condition, it can be derived that NE output and hospital profit:

$$\mathbf{x_N} = \frac{B(mn-\mathbf{1})}{km^3n^2}, \boldsymbol{\pi}_\mathbf{N} = \frac{B}{mn^2}$$

In the defection phase, all other hospitals maintain at the collusion output:

$$\mathbf{x_C} = \frac{B}{mn}$$

The defector has the following profit function:

$$\pi_D(x) = px - c(x) = px - mkx = \frac{B}{mx + \frac{m(n-1)B}{mn}} x - mkx$$

Again, using the first order condition, the profit function at defection phase is:

$$\pi_D = B + \frac{Bkm(-1+n)}{n} + \frac{2\sqrt{B^2km^3(-1+n)n^3}}{mn^2}$$

So the critical factor:

$$\delta^* = \frac{B + \frac{Bkm(-1+n)}{n} + \frac{2\sqrt{B^2km^3(-1+n)n^3}}{mn^2} - \frac{B}{n}}{B + \frac{Bkm(-1+n)}{n} + \frac{2\sqrt{B^2km^3(-1+n)n^3}}{mn^2} - \frac{B}{mn^2}} = \frac{Bm(1+km)(-1+n)n + 2\sqrt{B^2km^3(-1+n)n^3}}{2\sqrt{B^2km^3(-1+n)n^3} + B(-1+mn(km(-1+n)+n))}$$

Taking derivative over m:

$$\frac{\partial\delta^*}{\partial \mathbf{m}} = \frac{B^2(-\mathbf{1}+n)n\left(Bkm^2n^2(-\mathbf{3}+mn) + \sqrt{(B^2km^3(-\mathbf{1}+n)n^3)}(-\mathbf{1}+km(-\mathbf{2}+mn))\right)}{\sqrt{B^2km^3(-\mathbf{1}+n)n^3}\left(\mathbf{2}\sqrt{B^2km^3(-\mathbf{1}+n)n^3} + B(-\mathbf{1}+mn(km(-\mathbf{1}+n)+n))\right)^2}$$

Which is positive for mn>3. Therefore:

Given the number of hospitals in the region, N, the larger the number of service types, the harder it is to sustain collusion under global budget. That is, situation where there is a high degree of specialization, the collusion is more stable.