Hospital Non-price Competition under Global Budgeting

Wen-Yi Chen, Assistant Professor, Department of Leisure Business Management, Nan Kai Institute of Technology
Yu-Hui Lin, Assistant Professor, Department of Business Administration, Nan Kai Institute of Technology

ABSTRACT

We present a hospital non-price competition model to investigate two important concerns when a global budgeting system is applied to the expenditure cap policy and the expenditure target policy. Our results indicate that (1) the treadmill effect, (describing a phenomenon whereby health care providers provide more health care services with a lower reimbursed price under the expenditure cap policy) would be a quality-enhancing and efficiency-improving outcome; (2) both quality of care and managerial efficiency are higher under the expenditure cap policy than those under the expenditure target policy.
Keyword: *Treadmill effect; Quality competition; Expenditure cap policy; Expenditure target budget, Global Budgeting*
JEL classification: I11; I18

INTRODUCTION

Many countries with public financing health care system have adopted either the expenditure target policy (Canada for example) or the expenditure cap policy (Taiwan for example) for controlling medical cost under a global budgeting system. The expenditure target policy assigns a quota of health care services for which providers are reimbursed at a fixed price. When the volume of health care services exceeds the quota, the reimbursed price will be reduced for a certain proportion. The expenditure cap policy, however, refers to a fixed total budget for all health care services during a given period (usually one year) with hospitals being paid on the fee-for-service basis.

Fan et al. provide a theoretical analysis for the physician's behavior under these two policies ([2]). They show that given the same spending, a higher volume of health care services are provided under the expenditure cap policy than the expenditure target policy. Fan et al. consider the expenditure cap policy a better reimbursement policy than the expenditure target policy in terms of more quantity of health care services provided by physicians ([2]). Benstetter and Wambach, however, criticize the expenditure cap policy for a possible lower quality of care generated by the so-called treadmill effect, describing a phenomenon whereby physicians will provide more health care services with a lower reimbursed price under the expenditure cap policy ([1]).

In both studies mentioned earlier, the physician's objective is to maximize its profit, implying that cost minimization is assumed for their analyses. Nevertheless, this assumption has been questioned in the analyses of hospital behavior since non-for-profit hospitals usually dominate the health market ([5]). Even for-profit hospitals, the assumption of cost minimization may not be appropriate in modeling hospital behavior because internal organizations in hospitals are so complicated that each internal group (such as physicians, administrators, and owners) may have distinct and conflicting objectives ([3]).

Because of the existence of the technical efficiency in hospital operation, the purpose of this paper is to investigate hospital non-price competition under a global budgeting system in which both quality of care and managerial efficiency under the expenditure target policy are compared to that under the expenditure cap policy. To this end, we utilize a hospital non-price competition model, developed by [4], to explore two important concerns when a global budgeting system is applied to the expenditure cap policy and the expenditure target policy: (1) whether the treadmill effect generated by the expenditure cap policy is an adverse outcome that decreases both quality of care and managerial efficiency; (2) whether the expenditure cap policy is a better reimbursement policy than the expenditure target policy in terms of quality of care and managerial efficiency.

The rest of this paper is organized as follows: Section 2 models the treadmill effect under the expenditure cap policy, Section 3 presents the expenditure target policy, and the final section summarizes our results.

Expenditure Cap Policy

(1) Demand

Consider a health care system with three agents, the government, patients, and hospitals. The patients are usually fully insured in a health care system regulated by the expenditure cap policy, so price has a minor influence on health care demand. Hospitals tend to compete with quality of care in the way of Cournot competition. Due to the hospital's capacity, asymmetry of information about quality of care and a possible limitation of patient's mobility, hospitals offering the highest quality of care cannot serve the whole market and each hospital obtains a market share determined by its perceived quality relative to the quality of competing hospitals. Using the functional form from the study of non-price competition among hospitals developed by [4], the patient's demand for the hospital *i* is characterized by equation (1)-(2):

$$q_i(x_1, x_2, ...x_i, ..., x_n) = \frac{x_i^{\alpha}}{\sum_{j=1}^{n} x_j^{\alpha}} * D_T \qquad (1)$$

$$D_T = A * \left(\sum_{j=1}^{n} x_j^{\alpha} \Big/ n \right)^{\beta} \text{ is the overall demand for health care} \qquad (2)$$

The notations, q_i and D_T, represent the demand for the hospital *i*, and the overall demand for health care, respectively. x_i is the perceived quality of care from the hospital *i*. By a slight abuse of notation, we also use x_i to denote the quality of care at the hospital *i*. $A \in R^+$ is the total population. $\beta \in (0,1)$ is the elasticity of the overall demand with respect to the average perceived quality of care in the health care market. $\alpha \in (0, 1/\beta)$ is the degree of patients' mobility among hospitals according to their perceived quality of care. $\alpha \to 0$ means that hospitals are farther apart by either geographical area or information on quality of care. In this case, the increase of admissions in a hospital, due to a higher quality of care relative to its rivals, is small. In addition, $\alpha\beta$ is assumed to be bounded by an interval between zero and one (i.e., $\alpha\beta \in (0,1)$). $n \in N$ is the number of hospitals in the market. We consider the mobility of patients (α) and the number of hospitals (n) are two indicators of competition in our model.

Since we assume that hospitals act as Cournot competitors, the quality elasticity (η) for the demand of hospitals *i* (q_i) with respect to the perceived quality of care (x_i) can be written as equation [3] under the symmetric equilibrium ($q_i = q$ and $x_i = x$). [1]

$$\eta = \alpha \left(1 - \frac{1-\beta}{n} \right) \qquad (3)$$

(2) Price

The government imposes an expenditure cap (E) on the hospital sector, from which hospitals derive their revenue on a fee-for-service basis. The reimbursement payment per admission (P) for each hospital is determined by the fixed budget divided by total health care quantity. Hence, we have

$$P = \frac{E}{\sum_{j=1}^{n} q_j} \qquad (4)$$

where, E is the expenditure cap imposed on the health care system.

[1] The formal proof is displayed in Appendix 1.

(3) Cost and Profit

Since medical staffs value a slower pace of work or having extra supplies on hand, the marginal cost includes a slack variable, modeling pure managerial inefficiency and consumption at workplace (such as non-business telephone usage and internet surfing for self-interest during the business hours). Thus, the hospital's marginal cost of an admission and the profit of hospital *i* can be written as:

$$MC_i = c(x_i) + s_i \tag{5}$$

The notation MC_i is the unit cost of treating a patient when the hospital *i* chooses quality (x_i) and slack (s_i).The cost function for the quality of care (x_i) is defined by $c(x_i) = kx_i$, $k > 0$. We assume that $c(x_i) > s_i$, meaning that cost of slack is less than the expense of quality. In addition, the fixed cost is normalized to be zero for the purpose of mathematical tractability. Thus, the profit of hospital *i* (π_i) under the expenditure cap policy is written as:

$$\pi_i^G = (P - MC_i) * q_i = (P - c(x_i) - s_i) * q_i \tag{6}$$

Superscript *G* here as well as in equations (7)-(8) below indicates the expenditure cap policy.

(4) Hospital's problem

Hospitals maximize profits plus utility from managerial slack, so the hospital's problem is written as:

$$Max\ U_i(\pi_i^G, S_i) = \pi_i^G + F(S_i)\ s.t.\ \pi_i^G \geq 0 \tag{7}$$

The notion $F(S_i)$ is the utility obtained from total slack (S_i) where $S_i = s_i * q_i$. We assume that $F(S_i) > 0$, $\forall S_i > 0$. The assumption of concavity is imposed on $F(S_i)$. Thus, $F^{'}(S_i) > 0$, and $F^{''}(S_i) < 0$, imply a diminishing marginal utility of total slack. In order to ensure per admission slack (s_i) to be positive ($s_i > 0$), the condition $F^{'}(0) > 1$ is also imposed on $F(S_i)$. Equation (7) is a standard Kuhn-Tucker problem, which can be presented as maximizing the Lagrangian function below.

Solving the first-order conditions for the utility maximization, which are $\partial L^G / \partial x_i = 0$ and $\partial L^G / \partial s_i = 0$, the symmetric equilibrium of quality of care and slack is given by: [2]

$$c(x_G) = \frac{\eta}{1+\eta}(1+\xi)P \tag{9}$$

where $\xi = \frac{\partial P}{\partial q_i} * \frac{q_i}{P} = -\frac{1}{n}$ with imposing symmetric equilibrium of quantity.

$$F^{'}(S_G) = 1 + \lambda_G \text{, where } \lambda_G \geq 0 \tag{10}$$

The subscript *G* in equations (9)-(10) as well as in the discussion below indicates the symmetric equilibrium of equation (8). The Lagrange multiplier (λ_G) in equation (10) defines an equilibrium level of slack (s_G). If the constraint is not binding ($\pi_i^G > 0$, $\lambda_G = 0$), that implies total slack (S_G) to be independent of quality of care (x_G). If the constraint is binding ($\pi_i^G = 0$, $\lambda_G > 0$), the slack level can be solved by $\pi_i = 0$. The intuitions behind these results are reasonable because the hospital's administrator may not care about the slack level as long as the hospital's profit is positive. Contrarily, if the hospital's profit is so low that it may fail to meet zero-profit restriction regulated by owners, the hospital's administrator is more likely to take some aggressive management strategies to decrease slack but increases quality in order to attract more patients. In addition, the marginal cost under a positive-profit condition is the same as equation (5), while marginal cost under a zero-profit condition is reduced to

$$MC_G = c(x_G) + s_G = P \tag{11}$$

[2] The second order condition is assumed to be satisfied throughout this paper. The same assumption is also used in [4]. More detailed derivation for the solution is presented in Appendix 2.

The implication from equation (9) is that quality of care provided by hospitals is independent of the slack level and profit constraint. This condition does not mean that slack level and profit constraint do not have any impact on the choice of quality level, but implies that the slack level and profit constraint influence the choice of quality through other mechanism, which is the change of marginal cost per admission in accordance with the change of slack level and the binding of profit constraint. For example, when market competition drives the hospital's profit to zero, marginal cost (slack plus quality expense per admission) equals price so that the quality depends upon the slack and profit constraint (see equation (11)). Given equations (9)-(10) as well as equations (5) and (11), we are ready to prove the following proposition:

Proposition 1: Treadmill Effect

Under the expenditure cap policy, the following conditions hold.

(1-A) Competition in the health care market is measured by the mobility of patients (α) and the number of hospitals in the market (n). If competition is increasing (either $\alpha \uparrow$ or $n \uparrow$), hospitals will increase their quality of care ($x_G \uparrow$) but decrease their managerial slack ($s_G \downarrow$). Namely, $\frac{\partial x_G}{\partial \alpha} > 0$, $\frac{\partial x_G}{\partial n} > 0$, $\frac{\partial s_G}{\partial \alpha} < 0$, and $\frac{\partial s_G}{\partial n} < 0$.

(1-B) When the hospital is operated at a positive-profit status, the marginal cost is upward sloping. However, the marginal cost is downward sloping if the hospital is operated at a zero-profit status. Namely, If $\pi_i > 0$, then $x_G \uparrow \Rightarrow MC \uparrow (\frac{\partial MC}{\partial x_G} > 0)$. However, if $\pi_i = 0$, then $x_G \uparrow \Rightarrow MC \downarrow (\frac{\partial MC}{\partial x_G} < 0)$.

Proof: See Appendix 3.

The implications of Proposition 1 can be illustrated clearly by Figure 1. The hyperbolic curve is a fixed budget curve representing the underlying inverse relationship between price and quantity. Note that quantity is positively associated with quality in our model, so it is relevant for us to present the fixed budget curve in a geometric space generated by quality and price. The marginal cost curve is illustrated as the bold lines $\overline{DB}$ and $\overline{BC}$. Suppose that a hospital initially has positive profits so that it is operated at point A and its marginal cost is positive sloping at segment $\overline{DB}$ in Figure 1 (see Proposition 1-B).

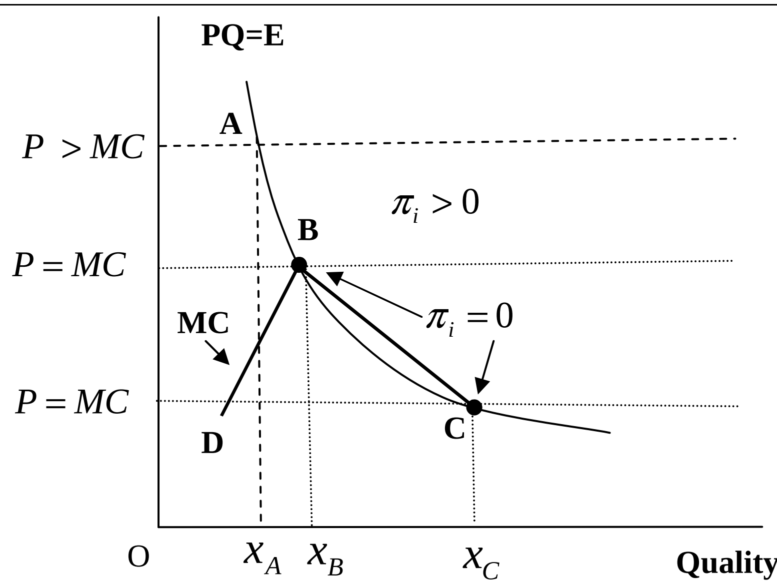

Figure 1: Illustration of Treadmill Effect

It is possible that the hospital prefers to stay at point A since the profit level is higher and so is the slack level. However, market competition (the increase of α or n ,or both) forces the hospital into choosing a higher level of quality (see Proposition 1-A), say the equilibrium at point B, the start point that could drive hospitals to a zero-profit status if market competition is increasing continuously. With zero profits, the increase of α or n not only increases quality also decreases the slack (see Proposition 1-A). In order to maintain zero profits, the cost-saving effect from the reduction of slack should be able to overcome the cost-increasing effect from quality-enhancing and price-decreasing effect resulting from the nature of the fixed budget regulated by the government. Therefore, the negative slope of marginal cost curve at segment $\overline{BC}$ is induced by competition (see Proposition 1-B), and the new equilibrium appears at point C. The phenomenon mentioned above, describing a possible downward sloping supply curve under the expenditure cap policy, is the treadmill effect illustrated by [1]. The treadmill effect presented here actually leads hospitals to put more effort into quality enhancement (higher quality level) and managerial efficiency improvement (lower slack level).

Expenditure Target Policy

The expenditure target policy means that a fixed quota of health care services is assigned to be performed by each hospital at a fixed reimbursement payment ($\overline{P}$), but the reimbursement payment ($\overline{P}$) will be reduced up to a specific proportion ($\gamma\overline{P}$, $\gamma \in (0,1)$) if quantities of health care services are provided above the quota. Under the design of this policy, the profit of hospital *i* is

$$\pi_i^T = (\overline{P} - MC_i) * q_i \text{ , if } q_i \le \overline{q}$$
$$= (\overline{P} - MC_i) * \overline{q} + (\gamma\overline{P} - MC_i) * (q_i - \overline{q}) \text{ , if } q_i > \overline{q}$$

Superscript *T* here as well as in equations (12)-(14) below indicates the expenditure target policy. Since the policy interest is the case that health care provision is higher than quota, the profit of hospital *i* is defined by

$$\pi_i^T = (1-\gamma)\overline{P} * \overline{q} + (\gamma\overline{P} - MC_i) * q_i = M + (\gamma\overline{P} - MC_i) * q_i \tag{12}$$

where, $M = (1-\gamma)\overline{P} * \overline{q}$ and $\gamma \in (0,1)$

The hospital's problem is to maximize its utility subject to the positive profit constraint, which can be defined as equation (13) or the Lagrangian function in the equation (14).

$$Max\ U_i(\pi_i^T, S_i) = \pi_i^T + F(S_i) \ s.t.\ \pi_i^T \ge 0 \tag{13}$$

$$L^T = (1+\lambda_T)\pi_i^T + F(S_i) \tag{14}$$

Again, solving the first-order conditions for utility maximization, $\partial L^T / \partial x_i = 0$ and $\partial L^T / \partial s_i = 0$, the symmetric equilibrium of quality of care and managerial slack is given by:[3]

$$c(x_T) = \frac{\eta}{1+\eta}\gamma\overline{P} \tag{15}$$

$F'(S_T) = 1 + \lambda_T$, where $\lambda_T \ge 0$ is the same as equation (10). (16)

The marginal cost under a positive-profit condition is the same as equation (5), while marginal cost under a zero-profit condition is reduced to

$$MC_T = \gamma\overline{P} + \frac{M}{q_T} \tag{17}$$

The subscript *T* in equations (15)-(17) indicates the symmetric equilibrium of equation (13) (or equation (14)). Given equations (5),(9)-(11), and (15)-(17), we are ready to prove the following proposition:

[3] See Appendix 4 for detailed derivation.

Proposition 2: Expenditure target policy

Under the expenditure target policy with a sufficiently large number of hospitals in the health care market, if the hospital's optimal level of quantity is above the quota set by the government ($q_i > \bar{q}$ *), given the same medical expenditure, the quality level obtained under the expenditure target policy is lower than the quality level obtained under the expenditure cap policy, while the slack level obtained under the expenditure target policy is higher than the slack level obtained under the expenditure cap policy. Namely,* $x_G > x_T$ *and* $s_G < s_T$ *as* $n \to \infty$.

Proof: See Appendix 5.

There are two implications from Proposition 2 in the case that market competition is strong enough (the number of hospitals is large). First, the slack level under the expenditure cap policy is lower than that under the expenditure target policy. This condition indicates that the expenditure target policy will exacerbate the technical inefficiency among hospitals. Second, given the same medical expenditure, the optimal quality level under the expenditure cap policy is higher than that under the expenditure target policy. This condition implies that if the government switches their reimbursement mechanism from the expenditure cap policy to the expenditure target policy, the beneficiary welfare will be lower due to a lower quality of care generated by the expenditure target policy.

Intuition behind these two implications is obvious. Under the expenditure target policy, hospitals can independently choose quality and slack to maximize their utility due to no limitation of total medical expenditure. The expenditure cap policy, however, introduces competition into the health care market through fixing the total medical expenditure. Therefore, the optimal quality level under the expenditure target policy is more likely to be lower than that under the expenditure cap policy, while the optimal slack level under the expenditure target policy is more likely to be higher than that under the expenditure cap policy. Since the quantity depends on the quality of care in the demand function (see equation (1)), the higher quality level generated by the expenditure cap policy compared to the expenditure target policy implies that the expenditure cap policy will induce hospitals to provide a higher quantity of care than the expenditure target policy. This finding is consistent with [2] in which they prove that the optimal quantity under the expenditure cap policy is higher than the expenditure target policy as long as number of providers is sufficiently high.

Summary

The main objective in this paper is to investigate two important concerns of when a global budgeting system is applied to the expenditure cap policy and the expenditure target policy (1) whether the treadmill effect resulting from the expenditure cap policy is an adverse outcome that decreases both quality of care and managerial efficiency; (2) whether the expenditure cap policy is a better reimbursement policy than the expenditure target policy in terms of quantity of care and managerial efficiency.

In respond to the first concern, our theoretical analyses in Proposition 1 suggest that the treadmill effect would be a quality-enhancing and efficiency-improving outcome if hospitals are not operated at cost minimization. In respond to the second concern, our theoretical analyses in Proposition 2 suggest that a higher quality of care but a lower slack level will be derived from the expenditure cap policy rather than the expenditure target. Since quality of care is positively related to quantity in our model, our results, consistent with [2], conclude that the expenditure cap policy a better reimbursement policy than the expenditure target policy from both higher quality and higher quantity of health care perspectives.

REFERENCES

[1] Benstetter F., and Wambach A(2006).The treadmill effect in a fixed budget system. *Journal of Health Economics,* Vol. 25, Issue 1, p146-169.

[2] Fan, CP., Chen, KP. and Kan, K.(1998). The design of payment systems for physicians under global budget-an experimental study. *Journal of Economic Behavior & Organization*, Vol. 34, p295-311.

[3] Harris J.E.(1977). The internal organization of hospitals: some economic implications. *Bell Journal of Economics*, 8, 467-482.

[4] Pope, G.(1989). Hospitals non-price competition and medicare reimbursement policy. *Journal of Health Economics*, Vol. 8, 147-172.

[5] Sloan, F.A. (2000). Not-for-profit ownership and hospitals behavior. in the *Handbook of Health Economics*, Vol. 1B, 1141-1174. Elsevier, Amsterdam.

APPENDICES :

Appendix 1: Proof for the quality elasticity of demand in equation [3]

From equations (1) and (2) in the text, we have

[A1] $\ln q_i = \ln(A/n) + \alpha \ln x_i + (\beta - 1)\ln\left(\sum_{j=1}^{n} x_j^{\alpha}\right)$

[A2] $\frac{\partial \ln q_i}{\partial x_i} * x_i = \left(\frac{\alpha}{x_i} + (\beta-1)\frac{\alpha (x_i)^{\alpha-1}}{\sum_{j=1}^{n} x_j^{\alpha}}\right) * x_i = \alpha\left(1 + (\beta-1)\frac{(x_i)^{\alpha}}{\sum_{j=1}^{n} x_j^{\alpha}}\right)$

where, we assume Cournot competition among hospitals ($\frac{\partial x_i}{\partial x_j} = 0$). At a symmetric equilibrium ($x_i = x_j = x$ for all *i* and *j*), we have [A3] $\eta = \alpha\left(1 - \frac{1-\beta}{n}\right)$ which is the equation (3) in the text. *QED.*

Appendix 2: Kuhn-Tucker conditions for the Expenditure Cap Policy

The Largrangean for this problem is equation (8) in the text.

$$L = (1+\lambda)[(P - c(x_i) - s_i) * q_i] + F(S_i)$$

$$\frac{\partial L}{\partial x_i} = (1+\lambda)[(P(1+\xi_i) - c(x_i) - s_i)\eta_i - c(x_i)] + F^{'}(S_i)s_i\eta_i \le 0$$

where, $\xi_i = \frac{\partial P}{\partial q_i}\frac{q_i}{P}$, and $\eta_i = \frac{\partial q_i}{\partial x_i}\frac{x_i}{q_i}$

At a symmetric equilibrium, $\forall x_i = x_G$, $\forall q_i = q_G$, and $\forall s_i = s_G \Rightarrow$

[A4] $[P(1+\xi) - c(x_G)]\eta - c(x_G) \le s_G\eta\left(1 - \frac{F^{'}(S_G)}{1+\lambda}\right)$

The complementary slackness condition for x_i is

[A5] $x_G\left\{[P\ (1+\xi) - c(x_G)]\eta - c(x_G) - s_G\eta\left(1 - \frac{F^{'}(S_G)}{1+\lambda}\right)\right\} = 0$

$$\frac{\partial L}{\partial s_i} = -(1+\lambda)q_i + F^{'}(S_i)q_i \le 0$$

[A6] $F^{'}(S_G) \le 1+\lambda$, where, "=" is imposed due to the assumption $s_G > 0$.

$\Rightarrow F^{'}(S_G) = 1 + \lambda_G$ which is equation [10] in the text. *QED.*

The complementary slackness condition for s_G is [A7] $s_G\left(F^{'}(S_G) - (1+\lambda_G)\right) = 0$

This condition implies that either $s_G = 0$ or $F^{'}(S_G) = 1 + \lambda_G$, or both, at an equilibrium. In any of these cases, that will eliminate s_G from equation [A4] for x_G. Thus, [A4] can be written as

[A8] $c(x_G) \ge \frac{\eta}{1+\eta}(1+\xi)P$, $x_G = 0$ is infeasible since "$\ge$" will not hold.

We have $c(x_G) = \frac{\eta}{1+\eta}(1+\xi)P$, which is the equation (9) in the text. *QED.*

Appendix 3: Proof for Proposition 1

To prove (1-A) $\frac{\partial x_G}{\partial \alpha} > 0$, $\frac{\partial x_G}{\partial n} > 0$, $\frac{\partial s_G}{\partial \alpha} < 0$, and $\frac{\partial s_G}{\partial n} < 0$.

(1-A-I) $\frac{\partial x_G}{\partial \alpha} > 0$;

Total differentiate equation [9] with respect to α , we have $\frac{\partial x_G}{\partial \alpha} > 0$.

$$\frac{\partial x_G}{\partial \alpha} = \left(\frac{1+\xi}{k}\right)\left[\frac{\partial}{\partial \alpha}\left(\frac{\eta}{1+\eta}\right)P + \left(\frac{\eta}{1+\eta}\right)\left(\frac{\partial P}{\partial q_G}\frac{\partial q_G}{\partial x_G}\frac{\partial x_G}{\partial \alpha}\right)\right]$$

After doing some algebra, we reach

$$\frac{\partial x_G}{\partial \alpha} = \left[\left(\frac{1+\xi}{k}\right)\left(\frac{\partial}{\partial \alpha}\left(\frac{\eta}{1+\eta}\right)\right)P\right] \Big/ \left[1-\left(\frac{1+\xi}{k}\right)\left(\frac{\eta}{1+\eta}\right)\left(\frac{\partial P}{\partial q_G}\frac{\partial q_G}{\partial x_G}\right)\right] = \frac{G_1}{G_2}$$

$$\frac{\partial}{\partial \alpha}\left(\frac{\eta}{1+\eta}\right) = \frac{\partial}{\partial \eta}\left(\frac{\eta}{1+\eta}\right)\frac{\partial \eta}{\partial \alpha} = \frac{1}{(1+\eta)^2}\left[1-\frac{1-\beta}{n}\right] > 0 \Rightarrow G_1 > 0$$

$$\frac{\partial P}{\partial q_G} < 0 \;;\; \frac{\partial q_G}{\partial x_G} > 0 \Rightarrow G_2 > 0 \Rightarrow \frac{\partial x_G}{\partial \alpha} > 0 \quad \textit{QED.}$$

(1-A-II) $\frac{\partial x_G}{\partial n} > 0$

Total differentiate equation [9] with respect to n , we have $\frac{\partial x_G}{\partial n} > 0$

$$\frac{\partial x_G}{\partial n} = \frac{1}{\alpha\beta+1}\left(\frac{E}{Ak}\right)\left(\frac{E}{Ak}\frac{(1+\xi)\eta}{1+\eta}\right)^{\frac{-\alpha\beta}{\alpha\beta+1}}\left[\left(\frac{\xi^2\eta}{1+\eta}\right)+\frac{\alpha(1-\beta)(1+\xi)\xi^2}{(1+\eta)^2}\right] > 0 \;\; \textit{QED.}$$

(1-A-III) $\frac{\partial s_G^{\pi>0}}{\partial \alpha} < 0$; $\frac{\partial s_G^{\pi>0}}{\partial n} < 0$

When $\pi > 0$, $F'(S_G) = 1$ $\Rightarrow$ Total differentiate $F'(S_G) = 1$ with respect to α and n .

After doing some algebra, we have $\frac{\partial s_G^{\pi>0}}{\partial \alpha} = -\frac{s_G^{\pi>0}}{q_G}\frac{\partial q_G}{\partial x_G}\frac{\partial x_G}{\partial \alpha} < 0$; $\frac{\partial s_G^{\pi>0}}{\partial n} = -\frac{s_G^{\pi>0}}{q_G}\frac{\partial q_G}{\partial x_G}\frac{\partial x_G}{\partial n} < 0$

(1-A-IV) $\frac{\partial s_G^{\pi=0}}{\partial \alpha} < 0$

If $\pi_i = 0$, we solve $\pi_i = 0$ for s_G . Namely, we solve $\left[(P - c(x_G) - s_G) * q_G\right] = 0$ for s_G . We have

[A9] $s_G^{\pi=0} = \left[1-\left(\frac{\eta}{1+\eta}\right)(1+\xi)\right]P$

Differentiate equation [A9] with respect to α , we have

$$\frac{\partial s_G^{\pi=0}}{\partial \alpha} = \left\{\frac{\partial}{\partial \alpha}\left[1-\left(\frac{\eta}{1+\eta}\right)\left(\frac{n-1}{n}\right)\right]P\right\} + \left\{\left[1-\left(\frac{\eta}{1+\eta}\right)\left(\frac{n-1}{n}\right)\right]\left(\frac{\partial P}{\partial q_G}\frac{\partial q_G}{\partial x_G}\frac{\partial x_G}{\partial \alpha}\right)\right\} = G_3 + G_4$$

Since $\frac{\partial}{\partial \alpha}\left[1-\left(\frac{\eta}{1+\eta}\right)\left(\frac{n-1}{n}\right)\right] = -\frac{1}{(1+\eta)^2}\left(1-\frac{1-\beta}{n}\right)\left(\frac{n-1}{n}\right) \Rightarrow G_3 < 0$

$$\frac{\partial P}{\partial q_G} < 0 \;;\; \frac{\partial q_G}{\partial x_G} > 0 ;\; \frac{\partial x_G}{\partial \alpha} > 0 \;\Rightarrow G_4 < 0 \Rightarrow \frac{\partial s_G^{\pi=0}}{\partial \alpha} < 0 \;\; \textit{QED.}$$

(1-A-V) $\frac{\partial s_G^{\pi=0}}{\partial n} < 0$

Differentiate equation [A9] with respect to n, we have

$$\frac{\partial s_G^{\pi=0}}{\partial n} = \left\{\frac{\partial}{\partial n}\left[1-\left(\frac{\eta}{1+\eta}\right)\left(\frac{n-1}{n}\right)\right]P\right\} + \left\{\left[1-\left(\frac{\eta}{1+\eta}\right)\left(\frac{n-1}{n}\right)\right]\left(\frac{\partial P}{\partial q_G}\frac{\partial q_G}{\partial x_G}\frac{\partial x_G}{\partial n}\right)\right\} = G_5 + G_6$$

Since $\frac{\partial}{\partial n}\left[1-\left(\frac{\eta}{1+\eta}\right)\left(1-\frac{1}{n}\right)\right] = -\frac{1}{n^2}\left[\frac{\alpha(1-\beta)}{(1+\eta)^2}\left(1-\frac{1}{n}\right)+\frac{1}{1+\eta}\right] \Rightarrow G_5 < 0$

$$\frac{\partial P}{\partial q} < 0\ ;\ \frac{\partial q_G}{\partial x_G} > 0;\ \frac{\partial x_G}{\partial n} > 0 \ \Rightarrow G_6 < 0 \ \Rightarrow \frac{\partial s_G^{\pi=0}}{\partial n} < 0 \quad QED.$$

(1-B) To prove $\frac{\partial MC}{\partial x} > 0$ if $\pi_i > 0$; $\frac{\partial MC}{\partial x} < 0$ if $\pi_i = 0$.

(1-B-I) $\frac{\partial MC}{\partial x} > 0$ if $\pi_i > 0$

$$\frac{\partial MC}{\partial x_G} = \frac{\partial c(x_G)}{\partial x_G} + \frac{\partial s_G}{\partial x_G} = k\left(1+\frac{\partial s_G}{\partial c(x_G)}\right) = \left(\frac{ks_G}{c(x_G)}\right)\left(\frac{c(x_G)}{s_G}+\sigma\right)$$

$\frac{\partial MC}{\partial x_G} > 0$ requires $\left(\frac{c(x_G)}{s_G}+\sigma\right)>0 \Rightarrow \sigma = \frac{\partial s_G}{\partial c(x_G)} * \frac{c(x_G)}{s_G} > -\frac{c(x_G)}{s_G}$

Since $F'(S_G) = 1$ when $\pi > 0$. $\Rightarrow S_G$ is independent of $x_G \Rightarrow \overline{S} = s_G * q_G = s_G * \frac{A}{n}(x_G)^{\alpha\beta}$

Take nature log on both sides, we have $\ln(\overline{S}) = \ln(s_G) + \ln\frac{A}{n} + \alpha\beta\ln(x_G)$

$$\Rightarrow \sigma = \frac{\partial s_G}{\partial x_G} * \frac{x_G}{s_G} = \frac{\partial s_G}{\partial c(x_G)} * \frac{c(x_G)}{s_G} = -\alpha\beta > -1,\ \because \alpha\beta \in (0,1).$$

$$\Rightarrow \left(\frac{c(x_G)}{s_G}+\sigma\right) = \left(\frac{c(x_G)}{s_G}-\alpha\beta\right) > 0,\ \because \frac{c(x_G)}{s_G} > 1 \text{ is assumed.} \Rightarrow \frac{\partial MC}{\partial x_G} > 0 \quad QED.$$

(1-C-II) $\frac{\partial MC}{\partial x_G} < 0$ if $\pi_i = 0$

From equations [11], we have $MC = c(x_G) + s_G = P$.

Hence, $\frac{\partial MC}{\partial x_G} = \frac{\partial c(x_G)}{\partial x_G} + \frac{\partial s_G}{\partial x_G} = \frac{\partial P}{\partial q_G}\frac{\partial q_G}{\partial x_G} < 0$ since $\frac{\partial P}{\partial q_G} < 0$ and $\frac{\partial q_G}{\partial x_G} > 0$. *QED.*

Appendix 4. Kuhn-Tucker conditions for the Expenditure Target Policy

The profit of hospital *i* is

$$\pi_i^T = (\overline{P} - MC_i) * \overline{q} + (\gamma\overline{P} - MC_i) * q_i \ , \text{ if } q_i > \overline{q}$$

The hospital's problem is $Max\ U_i(\pi_i^T, S_i) = \pi_i^T + F(S_i)\quad s.t.\ \pi_i^T \geq 0$

The Largrangean for this problem is equation [14] in the text.

$$L = (1+\lambda)\left[M + \left(\gamma\overline{P} - c(x_i) - s_i\right) * q_i\right] + F(S_i)\ , \text{ where } M = (1-\gamma)\overline{P}\overline{q}$$

$$\frac{\partial L}{\partial x_i} = (1+\lambda)\left[\left(\gamma\overline{P} - c(x_i) - s_i\right)\eta_i - c(x_i)\right] + F'(S_i)s_i\eta_i \leq 0$$

At a symmetric equilibrium, $\forall x_i = x_T$, $\forall q_i = q_T$, and $\forall s_i = s_T \Rightarrow$

[A10] $\left[\gamma \bar{P} - c(x_T)\right]\eta - c(x_T) \le \left(1 - \frac{F'(S_T)}{1+\lambda}\right) s_T \eta$

The complementary slackness condition for x_i is

[A11] $x_T\left\{\left[\bar{P} - c(x_T)\right]\eta - c(x_T) - \left(1 - \frac{F'(S_T)}{1+\lambda}\right) s_T \eta\right\} = 0$

$$\frac{\partial L}{\partial s_i} = -(1+\lambda)q_i + F'(S_i)q_i \le 0$$

[A12] $F'(S_T) \le 1 + \lambda_T$, where, "=" is imposed due to the assumption $s_T > 0$.

The complementary slackness condition for s_T is

[A13] $s_T\left(F'(S_T) - (1+\lambda_T)\right) = 0$

This condition implies that either $s_T = 0$ or $F'(S_T) = 1 + \lambda_T$, or both, at an equilibrium. In any of these cases, that will eliminate s_T from equation [A10] for x_T. Thus, [A10] can be written as

[A14] $c(x_T) \ge \frac{\eta}{1+\eta}\gamma \bar{P}$

$x_T = 0$ is infeasible since "$\ge$" will never hold. Thus, we have equation [15] in the text.

Similar to discussion in the expenditure cap policy, equation [A13] implies that

$F'(S_T) = 1$, if $\pi_i^T > 0 \Rightarrow s_T > 0$ since we assume that $F'(0) \ge 1$ and $F''(S_T) \le 0$.

If $\pi_i^T = 0$, we solve $\pi_i^T = 0$ for s_T. Namely, $M + \left(\gamma \bar{P} - c(x_T) - s_T\right) * q_T = 0$

$$\Rightarrow s_T = \gamma \bar{P} - c(x_T) + \frac{M}{q_T} = \left(\frac{\eta}{1+\eta}\right)\gamma \bar{P} + \frac{M}{q_T} > 0. \Rightarrow F'(0) \ge 1 \text{ and } F''(S_i) \le 0 \text{ ensure } s_T > 0.$$

Appendix 5: Proof for Proposition 2

(2-I) $x_G > x_T$ if $n \to \infty$

By equations (9) and (15), we have

$c(x_G) = \frac{\eta}{1+\eta}(1+\xi)P_G$ where, $P_G = \frac{E}{nq_G} = \frac{\bar{P} * \bar{q} + \gamma \bar{P}(q_T - \bar{q})}{A(x_G)^{\alpha\beta}}$

$c(x_T) = \frac{\eta}{1+\eta}\gamma \bar{P}$

$$\frac{c(x_G)}{c(x_T)} = (1+\xi)\left(\frac{[\bar{P} * \bar{q} + \gamma \bar{P}(q_T - \bar{q})]}{A(x_G)^{\alpha\beta} * \gamma \bar{P}}\right) \Rightarrow$$

$$\frac{A(x_G)^{\alpha\beta} c(x_G)}{c(x_T)} = (1+\xi)\left(\frac{\bar{P} * \bar{q} + \gamma \bar{P}(q_T - \bar{q})}{\gamma \bar{P}}\right) = (1+\xi)\left(\left(\frac{1}{\gamma} - 1\right)\bar{q} + q_T\right) > q_T = A(x_T)^{\alpha\beta} \text{ if}$$

$n \to \infty$ ($\xi \to 0$)

$\Rightarrow (x_G)^{\alpha\beta} c(x_G) > (x_T)^{\alpha\beta} c(x_T) \Rightarrow x_T < x_G$ *QED.*

(2-II) $s_G < s_T$ if $n \to \infty$

If $\pi_i^T > 0$ then $\lambda = 0$, $F'(S_T) = 1 \Rightarrow \overline{S_T} = s_T * q_T = s_G * q_G = \overline{S_G}$.

$\because n \to \infty \Rightarrow x_G > x_T \Rightarrow q_G > q_T \Rightarrow s_T > s_G$.

If $\pi_i^T = 0$, s_T can be solved by $\pi_i^T = 0$. Hence, the equilibrium slack per admission is

$s_T = \gamma\overline{P} - c(x_T) + \frac{M}{q_T}$, where $M = (1-\gamma)\overline{P} * \overline{q}$ and $s_G = P_G - c(x_G)$ if $\pi_i^G = 0$

Thus, $s_T - s_G = \left(\gamma\overline{P} + \frac{M}{q_T} - \frac{\overline{P} * \overline{q} + \gamma\overline{P}(q_T - \overline{q})}{q_G}\right) + \left(c(x_G) - c(x_T)\right) = T_1 + T_2$

If $n \to \infty \Rightarrow x_G > x_T \Rightarrow c(x_G) > c(x_T) \Rightarrow T_1 > 0$

$T_2 = \gamma\overline{P} + \frac{M}{q_T} - \frac{\overline{P} * \overline{q} + \gamma\overline{P}(q_T - \overline{q})}{q_G} > \gamma\overline{P}\left(1 - \frac{q_T}{q_G}\right) > 0$

Therefore, $s_T > s_G$ *QED.*