Optimal Compensation Rules: Pay-for-Performance vs Fee-for-Service*

Yaping Wu† Sanxi Li‡

May 20, 2015

Abstract

This paper examines the optimal non-linear compensation rule for physicians under pay-for-performance, fee-for-service and capitation payment in the presence of both adverse selection and moral hazard on the supply side. We identify the screening effect of fee-for-service. We provide an argument for the criticism on the shortcomings of fee-for-service. More importantly, we also provide a rationale for the continued use of fee-for-service payment even though serious problems with fee-for-service have been widely acknowledged. When moral hazard is the only problem, we find that fee-for-service can only lead to the substitution of treatment quantity to physician's effort, which is inefficient. Consequently, fee-for-service payments should not be used in this case. However, when moral hazard is combined with the adverse selection issue, an efficient screening requires a continued use of fee-for-service and less pay-for-performance for the lower productivity physicians. The design of the use of fee-for-service effectively improves screening.

JEL Code: I18, D82

Keywords: Physician compensation, fee-for-service, pay-for-performance, capitation

*I am grateful to David Bardey, Helmuth Cremer, Guido Friebel, Catarina Goulao, Christian Hellwig, Bruno Jullien, Jean-Marie Lozachmeur, Patrick Rey for their support, advice and comments.

†Southwestern University of Finance and Economics, ChengDu. Email: wuyp@swufe.edu.cn

‡Renmin University of China. Email: sanxi@ruc.edu.cn

1 Introduction

It has been widely believed that one of the key reasons for the high level of health care spending is the predominance of the fee-for-service payment system, which rewards quantity over quality. Instances of overuse, wasteful use, and misuse of care have been widely documented.[1] On the other hand, quality issue and patient's safty have attracted a lot of attention in the debate on health care policies in many countries. The economic literature has identified a number of triggers for these problems, among which are asymmetric informational problems.[2] Since the physician possesses more information regarding his private characteristics and behavior than the health care payer, he may abuse or cheat the health care program in favor of his own self-interests. The policymakers believe that one of the solutions for this problem is to provide incentives to the physicians. Many health innovators and reformers are experimenting with new pay-for-performance models (initially in the United States and the United Kingdom), which assert that payers should hold providers of health care accountable for both costs and quality of care. Physicians may receive bonuses based on performance indicators. However, many payment reforms do not change the fee-for-service payment in any fundamental way, but merely add new forms of pay-for-performance bonuses or additional penalties.

This paper provides a theoretical framework to analyze the debate between fee-for-service and pay-for-performance. The main contribution of this paper is to identify the screening effect of fee-for-service. We provide a rationale for fee-for-service payment from the perspective of the adverse selection issue.

We build a model to analyze the optimal compensation rule of physicians in the presence of both moral hazard and adverse selection on the supply side. We consider a nonlinear compensation scheme where the possible policy instruments include a capitation payment, a fee-for-service and a pay-for-performance system. The physician, privately knowing his professional productivity, performs a single task yielding a recovery probability that depends on two alterable and substitutable dimensions: the treatment quantity and the effort. One dimension, the treatment quantity is contractable, while the other dimension, the effort, is not contractable. Pay-for-performance is the reward that is based on the treatment outcome, that is, the recovery probability.

We confirm with the literature that the fee-for-service induces the substitution behavior of the physician. The fee-for-service is criticised for this reason. When only moral hazard problem is

[1] See UnitedHealth Center for Health Reform and Modernization (2012).

[2] See McGuire (2000).

considered, the optimal compensation policy does not pay for the services. By making the physician residual claimant, the payer solves the moral hazard problem. Payments for services can only lead to substitution behavior of the physician, which is inefficient and inflates costs. Hence the fee-for-service should not be used in this case.

However, the adverse selection problem is often ignored during the analysis of physician compensation in the literature, especially the screening problem on the physician's side. When both adverse selection and moral hazard are considered, the previous result no longer holds. The design of the use of fee-for-service effectively improves screening. We show that efficient screening requires using some fee-for-service and less pay-for-performance for the lower productivity physicians . This is due to the fact that the effort is not contractable and can only be indirectly contracted by using the observed outcome. As is usual in adverse selection models, informational rents of the better types can be mitigated by reducing the performance pay of the less efficient types. In order to give partial incentives for exerting effort, the payer has to reduce the pay-for-performance. However the reduced pay-for-performance may unintentionally induce an excessive decrease in treatment quantity which reduces health care quality. Hence, to avoid an excessive degradation in health care quality, it is desirable to also use fee-for-service payments. Assuming a risk neutral physician, we prove that, because of its screening effect, fee-for-service continues to be used as a provider payment method.

Whether the adverse selection issue is important or not depends on the volatility of the realization of the type. Smaller (larger) the support of the type is, less (more) volatile is the type's realization, and less (more) important is the adverse selection issue. When the screening is an important issue, it is desirable to use the fee-for-service payment method to avoid excessive degradations in quality.

Rosenthal et al. (2004) states that although in practice clinical outcome measures were rarely the basis of payment overall, they are more common among hospital incentive programs where in-hospital mortality, complication, and readmission rates are widely used metrics. The current pay-for-performance incentive programs in developed countries are a mixture of input-oriented and outcome-based payments. They reward the use of pre-specified health inputs such as the percentage of patients with diabetes on the register who have a record of retinal screening in the preceding 12 months. They also reward the performance of quality and efficiency measures using the treatment outcome, such as the percentage of patients with diabetes on the register, in whom the last IFCC-HbA1c (blood sugar) is 64 mmol/mol or less in the preceding 12 months. These are examples found in the Quality and Outcomes Framework (QOF) in the National Health Service (NHS) in the United Kingdom. Miller and Singer Babiarz (2013), in their study of performance payment in low- and middle-income countries, argue that if the primary objective of a health care program is the patient

or population health outcome, it would seem natural for performance incentives to reward good health or health improvement directly rather than the use of health services or other health inputs. However, because of limitations such as a small variation in health outcomes under provider control, uncontrollable patient behavior and the measurement issue of contracted outcomes, policymakers have to rely on pre-specified inputs. In this paper, we focus on the benchmark case where the recovery percentage is used by the payer to provide incentives to the physicians. In previous health economics literature, the quality is often a black box modeled as a deterministic variable entering into the benefit function or the demand function; in this paper, we open the black box by allowing the physician to invest in the performance, that is, the recovery probability.

Literature review

This paper is closely related to the physician compensation literature dealing with quality issues. K. Eggleston (2005)'s result paralels Chalkley and Malcomson (1998) which shows that in some case a payer may find it optimal to set supply-side cost sharing to zero to promote quality effort, but K. Eggleston (2005) considers multitasking problem. The reason that the cost reimbursement should be used is that it reduces the multitasking problem when the effort of that service is noncontractable. In this sense, following the above two papers, the present paper also provides rationales for the continued use of cost reimbursement, but with an additional argument that comes from the perspective of the adverse selection issue.

Ma (1994) studies the quality and cost-reducing-effort tradeoff. The optimal scheme is a pure capitation if dumping is not possible; a piecewise linear scheme is optimal if dumping is not forbidden. Ellis and McGuire (1986) derive the classic result on the debate between capitation and fee-for-service. They show that a mixed payment system, where hospitals are paid partly prospectively and partly cost-based, is superior to the other two single payment systems. Allen and Gertler (1991) assume heterogeneity of patients, and endogenize quantity and quality at the same time. Chalkley and Malcomson (1998 a) accentuate the tradeoff between quality and quantity, and introduce multidimensional quality. They intend to resume the quality issue by making the payment not only depend on the number of patients actually treated but also the number of patients who demand services. They assume that demand responds to some extent to the quality offered by the hospital. The contract is in fact composed of three margins: a lump sum transfer, a fixed price per patient treated and a fixed amount per patient wanting treatment. Chalkley and Malcomson (1998 b) extend their previous paper to analyse contracts when patient demand does not reflect quality.

For the most part, the previous literature focuses on the interplay between the capitation and

fee-for-service. This present paper analyzes the interplay between the three payment methods: fee-for-service, capitation and pay-for-performance. As capitation is a payment which is based on the number of patients, it is still a quantity-based payment. This paper mainly focuses on the debate between fee-for-service and pay-for-performance.

The paper proceeds as follows. Section 2 sets up the model. Section 3 derives the first best optimum. Section 4 studies the pure moral hazard problem and its implementation. Section 5 analyzes the problem with both moral hazard and adverse selection, as well as the contract design. Section 6 provides discussions and robustness checks. Section 7 concludes.

2 The Model

Consider three agents in the model, which are the physician, the patient and the payer. The physician is a profit-maximizer whose profit depends on the payment S received from the payer, the number of patients treated $x \in \mathcal{R}^+$, the treatment quantity per patient $q \in \mathcal{R}^+$, and his effort per patient $e \in \mathcal{R}^+$:

$$\pi = S - C(xq) - \varphi(xe), \tag{1}$$

where $C(xq)$ denotes the total monetary cost of the aggregate treatment quantity, and $\varphi(xe)$ denotes the total disutility of effort. Both of these functions are increasing and convex in their arguments. The physician chooses the number of patients to be treated, the treatment quantity per patient and the effort per patient. In practice, the proxy for the treatment quantity can be the number of repeated office visits, the quantity of disposable medical appliances, the length of hospital stay, or the extent of surgical versus drug interventions. The proxy for the effort can be the time spent on the patient, or a costly mental and manual work.

A patient gets a benefit $b > 0$ if he recovers after the physician's intervention, and 0 if he does not recover. The benefit from recovery can be measured by the reacquisition of potential economic or non-economic losses due to the disease in case of no intervention from the physician.[3] But in the case of non-recovery, even with the intervention from the physician, the patient still suffers from these losses and does not reacquire these potential losses. Hence b measures the benefit that a patient expects to enjoy but unfortunately does not in the case of non-recovery.

[3]Economic losses include financial losses such as lost wages (sometimes called lost earning capacity). These losses may be assessed for future losses due to the disease in the case of no intervention. Non-economic losses are assessed for the patients themselves: physical and psychological harm, such as loss of vision, loss of a limb or organ, the reduced enjoyment of life due to a disability, severe pain and emotional distress in case of no intervention.

Central to our analysis is the assumption that the provision of treatment to the patients generates recovery only with some probability P:[4]

$$P = P(e, q, \theta; \alpha), \tag{2}$$

where $\alpha \in (0, 1)$ denotes the nature of the disease, with higher α measuring a more serious disease. Let $\theta \in \mathcal{R}^+$ denote the professional productivity of the physician, being distributed according to the cumulative distribution $F(.)$ with density function $f(.)$ on the interval $[\theta_0, \theta_1]$, with higher θ designating higher professional productivity. For notation, let the lower index refer to the first order derivation and the double lower index refer to the second order derivation. We make the following usual assumption on the recovery probability:

Assumption 1 *$P_\theta > 0$, $P_q > 0$, $P_e > 0$ and $P_{qq} < 0$, $P_{ee} < 0$, $P_{q\theta} > 0$, $P_{e\theta} > 0$.*

Assumption 1 states that, given the nature of the disease, high professional productivity increases the recovery probability, and larger quantity of treatment or higher level of physician effort increases the recovery probability at a decreasing rate. The last two conditions state that higher professional productivity also increases both the marginal productivity of treatment quantity and the marginal productivity of the effort. These are the usual Spence-Mirrlees conditions which simply state that a more efficient type is also more efficient at the margin.

To simplify, we assume that the recovery probability takes the following function:[5]

$$P = P(e, q, \theta; \alpha) = 1 - \alpha exp\big(-\theta f(e, q)\big), \tag{3}$$

where $f(e, q)$ is the constant elasticity of substitution production function $f(e, q) = (e^\rho + q^\rho)^{\frac{1}{\rho}}$ with the coefficient of substitution $\rho \in (-\infty, 1]$. Therefore, more serious the disease is, smaller is the recovery probability. Moreover, more serious the disease is, more important are the physician's ability, his effort and the treatment quantity for the recovery probability. A higher α also implies higher marginal productivity of ability, effort and treatment quantity.

[4]In the case of chronic disease, the patient may never be able to recover. In this case, another performance measure is needed, and our P function in the model can still be used for the analysis, albeit with another interpretation. For example, a possible performance measure is how long on average the patient's life has been extended under the physician's intervention. For example, the disability-adjusted life year (DALY) is a measure of overall disease burden, expressed as the number of years lost due to ill-health, disability or early death.

[5]Under the general form of probability function, our main results on the first best and second best policy mix remain unaffected. The only thing that is affected is that the direction of the distortion on treatment quantity is ambiguous in the second best.

Since x is continuous, according to the law of large numbers, the proportion of patients that successfully recover from the intervention almost surely converges to the recovery probability P. Therefore, the recovery probability can be inferred from the proportion of patients that successfully recover (that is, the success rate).

Given the nature of the disease, the expected benefit to a patient treated by a type-θ physician is

$$Pb = P(e(\theta), q(\theta), \theta; \alpha)b. \tag{4}$$

As in Ma and Alger (2003) and Ma and McGuire (1997), we consider a risk-neutral payer who is assumed to operate in a competitive market and to set its policy to maximize the patients' expected utility, or a public regulator who cares only about the benefit of the patient because there is a shadow cost of public fund, where all rents paid to providers are costly. Given the nature of the disease, the payer's preference is given by the difference between the patients' expected benefit and the payment S paid to the physician. We denote its objective function as SW defined thus as follows:

$$\int_{\theta_0}^{\theta_1} [x(\theta)P(e(\theta), q(\theta), \theta; \alpha)b - S(\theta)]f(\theta)d\theta. \tag{5}$$

We consider a non-linear expected payment scheme $S = S(x, q, P)$, which possibly depends on the number of patients treated, the treatment quantity and the recovery probability. Therefore, the possible instruments are defined respectively as follows. Let $S_q(x, q, P)$ denote the fee-for-service which is the payment based on the treatment quantity. Let $S_x(x, q, P)$ denote the capitation payment which is based on the number of patients being treated. And let $S_P(x, q, P)$ denote the pay-for-performance which is based on the success rate of the physician's intervention.

We consider alternative information structures. Regardless of which case we consider, we assume that the number of patients and treatment quantity can always be observed and verified by the payer. In the first best benchmark, we assume that both the type and the effort are observable. Next, we consider the pure moral hazard where the type is still observable but the effort is not observable. Then, in the second best, we consider the case where neither the type nor the effort is observable. As stated above, according to the law of large numbers, the recovery probability can be inferred from the the proportion of patients that successfully recover (that is, the success rate), which is observable. However, the observation of this success rate does not allow the payer to perfectly disentangle the type of the physician and his level of effort. Although the payment scheme cannot be based on the effort which is generally not contractable, the success rate is observable. The payer can take advantage of this additional piece of information by using it as an additional screening variable.

The timing of the game is as follows. At the first stage, the payer sets the payment policies. Then, the type of the physician realizes. At the second step, the physician selects the number of patients, the treatment quantity as well as the effort level. Finally, the payment policies are implemented.

3 The first best optimum

In the first best benchmark, both the type of physician and the effort are observable. In this section, we characterize for each type of physician, the first best number of patients, the treatment quantity per patient, and the effort per patient. Rewriting the profit function (1) we have $S\big(x, q, P(e, q, \theta; \alpha)\big) = \pi + C(xq) + \varphi(xe)$. Replacing it into the payer's objective function (5), the problem of the payer is as follows:

$$\max_{x(\theta), q(\theta), e(\theta), \pi(\theta)} \int_{\theta_0}^{\theta_1} [x(\theta)P(e(\theta), q(\theta), \theta; \alpha)b - \pi(\theta) - C(x(\theta)q(\theta)) - \varphi(x(\theta)e(\theta))]f(\theta)d\theta \tag{6}$$

$$s.t. \quad \pi(\theta) \geq 0. \tag{7}$$

Since all information is publicly observable, the payer will leave no informational rent to the physician by choosing $\pi(\theta) = 0, \forall\theta$. Assuming an interior solution, deleting the argument inside the parenthesis refering to the individual type, the first order conditions with respect to x, q and e are respectively as follows:

$$P(e, q, \theta; \alpha)b = qC'(xq) + e\varphi'(xe), \tag{8}$$

$$P_q(e, q, \theta; \alpha)b = C'(xq), \tag{9}$$

$$P_e(e, q, \theta; \alpha)b = \varphi'(xe), \tag{10}$$

for every $\theta \in [\theta_0, \theta_1]$.

At the first best optimum, the marginal cost of patient is equal to the marginal benefit. The marginal cost of treatment is equal to the marginal benefit of treatment, and the marginal cost of effort is equal to the marginal benefit of effort. If second order cross derivative of recovery probability with respect to e and q are sufficiently small in absolute value, equation (10) implies that under the Spence-Mirrlees condition $P_{e\theta}(e, q, \theta; \alpha) > 0$, a higher productivity physician exerts a higher level of effort on aggregate xe.[6] From equation (9), under the Spence-Mirrlees condition $P_{q\theta}(e, q, \theta; \alpha) > 0$, a

[6]If the effort is interpreted as the time spent on the patient, the higher productivity physician exerts longer time on aggregate.

higher productivity physician is allocated with a higher level of total treatment quantity xq, because he/she is more efficient at the margin. Furthermore, more serious disease requires higher level of effort in aggregate and higher level of total treatment quantity. We summarise our findings in the next proposition:

Proposition 1 (First best optimum) *The first best allocation $\{x(\theta), q(\theta), e(\theta)\}$ for each θ are described by equation (8)(9) and (10). Under the Spence-Mirrlees conditions, the higher productivity physician exerts a higher level of effort xe and chooses a higher level of total treatment quantity xq.*

4 The pure moral hazard

In this section, we study the pure moral hazard where we assume that the physician's effort is not observable but the type is observable. Since the link between effort, types and the recovery probability is completely deterministic, it entails no randomness. Given a target value of the recovery probability P, which is a contractual variable available to the payer, given the physician's type and the nature of the disease, effort is completely determined by the condition $e = e(P, q, \theta; \alpha)$, where $e(.)$ is implicitly defined by the identity $P = P(e(P, q, \theta; \alpha), q, \theta; \alpha)$ for all θ in Θ and all P. In fact, the physician has no freedom in choosing his effort level when he takes his decision. As a result, the first best allocations can be achieved. Before moving to the contract design in the pure moral hazard setting, let us first analyze the physician's behavior.

4.1 The physician's behavior

In this section, we study the decision choice of the physician under a given contract. Given the nature of the disease, the physician, privately knowing his type θ, selects the number of patients, the treatment quantity per patient and the effort per patient. Deleting the argument inside the parenthesis refering to the individual type, the problem of the physician is:

$$\max_{x,q,e} S(x, q, P(e, q, \theta; \alpha)) - C(xq) - \varphi(xe). \tag{11}$$

Assuming an interior solution, the first order conditions with respect to x, q and e are respectively given by:

$$S_x(x, q, P) = qC'(xq) + e\varphi'(xe), \tag{12}$$

$$S_q(x, q, P) + S_P(x, q, P)P_q(e, q, \theta; \alpha) = xC'(xq), \tag{13}$$

$$S_P(x, q, P)P_e(e, q, \theta; \alpha) = x\varphi'(xe). \tag{14}$$

Therefore, from equation (12), (13) and (14), we observe that if we increase fee-for-service S_q, it induces substitution of either treatment quantity and/or number of patients to physician effort. Keeping the capitation payment and pay-for-performance constant, if we increase the fee-for-service, equation (13) implies that the physician will choose either higher treatment, or more patients, or both. At least one allocation, q or x, will be higher. If the physician chooses both higher treatment quantity and more patients, equation (12) implies that effort will be lower. If the physician chooses only higher treatment quantity, then following equation (12) effort will be lower too. If the physician chooses only more patients, then equation (14) implies that effort will be lower too. However, the physician may choose a much larger amount of treatment with fewer patients or a higher number of patients with less treatment. In these two cases, which allocations are induced by the fee-for-service is ambiguous. But according to equation (13) the marginal cost of treatment is higher, at least one allocation, x or q, is selected to be higher. Hence, the fee-for-service induces substitution of either treatment quantity and/or number of patients to physician effort.

Since the payment method cannot be based on the effort, the only method that can induce a profit-maximizer physician's effort is pay-for-performance. If the percentage of recovered patients is too costly to be observed or be verified so that the payer cannot use the pay-for-performance, equation (14) implies that $e = 0$: there is no incentive to exert any effort.[7]

4.2 Implementation in a pure moral hazard setting

In this section, we consider the implementation of the allocations in a pure moral hazard environment.

Lemma 1 *When there is no adverse selection problem, by making the physician residual claimant:* $S^*(x, P) = xPb - M(\theta)$ *where* $M(\theta)$ *is a constant depending on the type, the payer solves the moral hazard problem.*

The treatment q is contractable while the effort e is not contractable. By making the physician residual claimant, the physician's profit is perfectly in line with the payer's objective. Consequently, there is no need to contract on any contractable dimension q. Combining equation (8)(9)(10) with

[7] Equations (12) and (13) imply that the number of patients will not be zero because the fee-for-service and the capitation is still positive.

equation (12)(13)(14), we derive the optimal policy in a pure moral hazard environment:

$$S_P^*(x, q, P) = xb, \tag{15}$$

$$S_q^*(x, q, P) = 0, \tag{16}$$

$$S_x^*(x, q, P) = Pb. \tag{17}$$

Thus, we summarise our findings in the next proposition:

Proposition 2 (Pure moral hazard policy) *When there is pure moral hazard and no adverse selection, the optimal policy mix can be described by equations (15),(16) and (17). The allocation is the same as in the first best. The optimal compensation policy does not pay for the services; the pay-for-performance incentivizes the physician to fully internalize the patients' benefit.*

There is no need to reward services. Rewards for services only induce substitution behaviors, which are inefficient and inflate costs. The physician performs a single task: the recovery probability, which has two alterable dimensions, that is, the treatment quantity and the effort. Both dimensions contribute to the recovery probability and they are substitutable to some degree measured by ρ. One dimension, the treatment quantity is contractable, while the other dimension, the effort, is not contractable. The effort can and can only be indirectly contracted by using the payment based on the outcome, the recovery probability. Given that two dimensions equally contribute to the outcome, if one dimension is paid, while the other is not and the outcome is paid, the physician will naturally tend to select the higher level on the dimension which is paid, since it also increases the outcome which is also paid. We thus get a result which is similar to the findings of Holmstrom and Milgrom (1991). Incentive payment on the contractable dimension will induce substitution of this dimension to the other uncontractable dimension. In order not to induce such substitution behavior, the payer should not pay upon any dimensions, especially the contractable dimension in the pure moral hazard case. If only the outcome is paid, the physician will select the two dimensions by comparing their contributions to the benefit and their cost just as the payer does. Hence in the presence of pure moral hazard, the optimal policy mix includes no fee-for-service at all. The reason why the capitation based on the contractable number of patients is still used is that the number of patients does not affect the patients' recovery probability.

We confirm with the literature that the fee-for-service induces the substitution behavior of the physician. The fee-for-service is criticised for this reason. When only moral hazard is considered, the optimal compensation policy does not pay for the services. However, the adverse selection problem is often ignored during the analysis of physician compensation in the literature, especially

the screening problem on the physician's side. In the next section, we introduce the adverse selection issue with respect to the physician's professional ability and analyze the second best payment policy in this case.

5 Moral hazard and adverse selection

5.1 The characterization

The first best and pure moral hazard solutions, and their decentralization, have been derived under the assumption that the payer observes the type of the physician. When there is asymmetric information on the type, the optimal scheme with pure moral hazard is generally not feasible. Consider two types of physician $\theta_1 < \theta_2$. In the first best or the pure moral hazard environment, all types of physician get zero profit. In particular,

$$\pi(\theta_1) = S(x_1, q_1, P_1) - C(x_1 q_1) - \varphi(x_1 e(P_1, q_1, \theta_1; \alpha)) = 0,$$
$$\pi(\theta_2) = S(x_2, q_2, P_2) - C(x_2 q_2) - \varphi(x_2 e(P_2, q_2, \theta_2; \alpha)) = 0.$$

When the type of the physician is unknown and the effort is unobservable, if the payer still offers the first best/MH allocations, the high-productivity physician gets a positive profit by mimicking the low-productivity type:

$$\widetilde{\pi}(\theta_2) = S(x_1, q_1, P_1) - C(x_1 q_1) - \varphi(x_1 e(P_1, q_1, \theta_2; \alpha)) > 0,$$

because $\frac{\partial e(P,q,\theta;\alpha)}{\partial \theta} < 0$, for a given level of P and q.

In this section, we consider an information structure where neither the type nor the effort is observable. We characterize the second best optimum.

Given the nature of the disease, the physician's productivity, the treatment quantity, the effort and the recovery probability is deterministically linked by the probability function. As a result, after having chosen the treatment quantity, choosing the effort is equivalent to choosing a recovery probability. We apply the change of variable: $e = e(P, q, \theta; \alpha)$, where $e(.)$ is implicitly defined by the identity $P = P(e(P, q, \theta; \alpha), q, \theta; \alpha)$ for all θ in Θ and all P. With observables being x, q and P, we consider the direct revelation mechanisms $\{S(\hat{\theta}), x(\hat{\theta}), q(\hat{\theta}), P(\hat{\theta})\}_{\hat{\theta} \in [\theta_0, \theta_1]}$ which are truth telling. The incentive compatibility constraint is written as follows: $\forall \theta$ and $\hat{\theta} \neq \theta$,

$$S(\theta) - C(x(\theta)q(\theta)) - \varphi(x(\theta)e(P(\theta), q(\theta), \theta; \alpha)) \geq S(\hat{\theta}) - C(x(\hat{\theta})q(\hat{\theta})) - \varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)). \tag{18}$$

We use the first order approach.[8] The payer's problem in the second best is as follows:

$$\max_{x(\theta),q(\theta),P(\theta;\alpha),S(\theta)} \int_{\theta_0}^{\theta_1} [x(\theta)P(\theta;\alpha)b - S(\theta)]f(\theta)d\theta \tag{19}$$

$$s.t. \quad S(\theta) - C\big(x(\theta)q(\theta)\big) - \varphi\big(x(\theta)e(P(\theta;\alpha), q(\theta), \theta; \alpha)\big)$$

$$= \max_{\hat{\theta}\in[\theta_0,\theta_1]} \big\{S(\hat{\theta}) - C\big(x(\hat{\theta})q(\hat{\theta})\big) - \varphi\big(x(\hat{\theta})e(P(\hat{\theta};\alpha), q(\hat{\theta}), \theta; \alpha)\big)\big\}, \tag{20}$$

$$S(\theta) - C\big(x(\theta)q(\theta)\big) - \varphi\big(x(\theta)e(P(\theta;\alpha), q(\theta), \theta; \alpha)\big) \geq 0, \tag{21}$$

$$x(\theta) \geq 0, \quad q(\theta) \geq 0, \quad e(\theta) \geq 0, \quad 0 \leq P(\theta;\alpha) \leq 1, \quad \text{for all} \quad \theta \in [\theta_0, \theta_1]. \tag{22}$$

We leave the procedure of solving the above program to the appendix. After changing of variables, for an interior solution, the associated first order conditions with respect to x, q, e are as follows:

$$P(e,q,\theta;\alpha)b = qC'(xq) + e\varphi'(xe) - \frac{1-F(\theta)}{f(\theta)}\Delta_x, \tag{23}$$

$$xP_q(e,q,\theta;\alpha)b = xC'(xq) - \frac{1-F(\theta)}{f(\theta)}\Delta_q, \tag{24}$$

$$xP_e(e,q,\theta;\alpha)b = x\varphi'(xe) - \frac{1-F(\theta)}{f(\theta)}\Delta_e, \tag{25}$$

for every $\theta \in [\theta_0, \theta_1]$, where

$$\Delta_x = \big(\varphi'(xe) + xe\varphi''(xe)\big)\frac{\partial e(P(e,q,\theta;\alpha), q, \theta; \alpha)}{\partial \theta}, \tag{26}$$

$$\Delta_q = x\varphi'(xe)\frac{\partial^2 e(P(e,q,\theta;\alpha), q, \theta; \alpha)}{\partial\theta\partial q}, \tag{27}$$

$$\Delta_e = \frac{\partial\left[x\varphi'(xe)\frac{\partial e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta}\right]}{\partial e}. \tag{28}$$

As usual in adverse selection models, for $\theta = \theta_1$, $F(\theta_1) = 1$, there is no distortion for the highest type θ_1. As it is proved in the appendix that $\Delta_x < 0$, $\Delta_q < 0$, $\Delta_e < 0$, at the second best optimum, compared to the first best tradeoff, for all types such that $\theta < \theta_1$, the number of patients is distorted downwards. The marginal benefit of effort is greater than the marginal cost. These

[8] The first order approach is valid as long as the second order condition is satisfied. A sufficient condition for the second order condition to be satisfied is that $\dot{x}(\theta) > 0$, $\dot{P}(\theta) > 0$ and $\dot{q}(\theta) < 0$, where a dot means that the variable is the derivative with respect to θ. The proof is given in the Appendix.

types of physician under-provide effort. The treatment quantity is also distorted downwards.[9] In fact, rewriting the incentive compatibility constraint (18), we have $\forall\theta$ and $\hat{\theta} \neq \theta$,

$$\pi(\theta) \geq \pi(\hat{\theta}) + [\varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \hat{\theta}; \alpha)) - \varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha))]. \tag{29}$$

The second term on the right hand side of equation (29) is the informational rent given to the type θ. It is proved in the appendix that this informational rent is increasing in $x(\hat{\theta})$, increasing in $e(\hat{\theta})$ and also increasing in the treatment quantity $q(\hat{\theta})$ of the mimicked type $\hat{\theta}$. In fact, starting from the first best tradeoff, $\forall\theta < \theta_1$, variations $dx(\theta) < 0$, $de(\theta) < 0$ along with a variation $dq(\theta) < 0$ have no first order effect on the efficiency, but they decrease the profit of the mimicking type, and hence decrease the informational rent given to the mimicking type. Consequently, the downward distortion in x, e and q, $\forall\theta < \theta_1$ is a way to relax the otherwise binding incentive compatibility constraints.

5.2 Exclusion

For efficiency reasons, the payer may not want to offer contracts to all the types of physicians. In this section, we study the conditions under which a subset of types are excluded from the second best optimal contracts. We first summarise the conditions in the following lemma:

Lemma 2 (Exclusion) *Given the nature of the disease, assume that the hazard rate* $\frac{1-F(\theta)}{f(\theta)}$ *is strictly decreasing in* θ. *If* $\rho \nrightarrow 0$ *and* $C'(0) > 0$, $\varphi'(0) > 0$, *then*
i) If there exists $\tilde{\theta}$ *such that* $C'(0) \geq P_q(0, 0, \tilde{\theta}; \alpha)b$ *and* $\varphi'(0) \geq P_e(0, 0, \tilde{\theta}; \alpha)b$, *then all types* $\theta_0 \leq \theta \leq \tilde{\theta}$ *are excluded.*
ii) If $\forall\theta$, $C'(0) < P_q(0, 0, \theta; \alpha)b$ *and* $\varphi'(0) < P_e(0, 0, \theta; \alpha)b$, *then, if there exists* $\tilde{\theta}$ *such that,*

$$\frac{f(\tilde{\theta})}{1 - F(\tilde{\theta})} < Min\left\{ \frac{\varphi'(0)}{\tilde{\theta}(2^{\frac{1-\rho}{\rho}}\alpha\tilde{\theta}b - C'(0))}, \frac{\varphi'(0)}{\tilde{\theta}(2^{\frac{1-\rho}{\rho}}\alpha\tilde{\theta}b - \varphi'(0))} \right\}, \tag{30}$$

then all types $\theta_0 \leq \theta \leq \tilde{\theta}$ *are excluded.*

Note that if $\rho \to 0$, the Inada conditions are satisfied: $\forall\theta$

$$P_q(0, 0, \theta; \alpha) = +\infty,$$
$$P_e(0, 0, \theta; \alpha) = +\infty.$$

[9] Although the downward distortion of treatment is derived under CES function, whether the treatment is downward or upward distorted does not affect our main conclusion in the second best policy mix.

If in addition, $C'(0) = 0$ and $\varphi'(0) = 0$, exclusion is never desirable. It means that if, whatever the type, the marginal benefits are infinity and the marginal costs are zero at zero allocations, no type is excluded, even the least efficient type. If $\rho \nrightarrow 0$, and $C'(0) > 0$ and $\varphi'(0) > 0$, then, the result $i)$ of Lemma 2 states that if there exists a type such that the marginal costs at zero are higher than the marginal benefits at zero, then the Spence-Mirrlees conditions guarantee that all types that are lower than this type are excluded from the optimal contract. If for all types, the marginal benefits at zero are higher than the marginal costs at zero, then result $ii)$ states that the types whose productivity is close to zero, and whose presence among the physicians are insignificant, are excluded from the optimal contract. These conditions depend only on the primitives.

5.3 The second best compensation policy

In this section we derive the second best compensation scheme for $\theta \in (\tilde{\theta}, \theta_1]$ who are not excluded from the contract. By combining equations (12), (13) and (14) with equations (23), (24) and (25), we obtain the following prices:

$$S_q(x, q, \mathcal{P}) = \frac{1 - F(\theta)}{f(\theta)}(-|\Delta_q| + |\Delta_e|\frac{P_q}{P_e}), \tag{31}$$

$$S_x(x, q, \mathcal{P}) = \mathcal{P}b - \frac{1 - F(\theta)}{f(\theta)}|\Delta_x|, \tag{32}$$

$$S_{\mathcal{P}}(x, q, \mathcal{P}) = xb - \frac{1 - F(\theta)}{f(\theta)}\frac{|\Delta_e|}{P_e}. \tag{33}$$

We observe that there is no distortion of prices at the top. The highest productivity physician θ_1 continues to receive the first best prices and is fully residual claimant for his effort. For $\theta \in (\tilde{\theta}, \theta_1)$, for the same level of x, q and e, we obtain the prices as follows:

$$S_q^{SB} > 0, \tag{34}$$

$$S_x^{SB} < S_x^*, \tag{35}$$

$$S_{\mathcal{P}}^{SB} < xb = S_{\mathcal{P}}^*. \tag{36}$$

Compared to the first best prices (15), (16) and (17), for the same level of x, q and e, all types lower than the highest type get a positive fee-for-service, less capitation payment and less pay-for-performance. As they are rewarded less on the pay-for-performance, they are just given partial incentive to exert effort, hence are partially residual claimants for their effort. Surprisingly, we find that the second best fee-for-service is positive. As is usual in adverse selection models informational

rents of the better types can be mitigated by reducing the performance pay of the less efficient types. However the reduced pay-for-performance may unintentionally induce an excessive decrease in treatment quantity which reduces quality. To avoid an excessive degradation in quality, it is then desirable to also use the fee-for-service payment. We prove that because of its screening effect, the fee-for-service continues to be used as a provider payment method, even if with a risk neutral physician. We summarize our findings in the following proposition:

Proposition 3 (Second best policy) *When moral hazard is combined with the adverse selection issue, efficient screening requires using some fee-for-service for the lower productivity physicians and less pay-for-performance. The fee-for-service effectively improves screening.*

Proposition 4 provides a rationale for the continued use of fee-for-service scheme in spite of the widely acknowledged problems associated with this payment scheme. When moral hazard is the only problem, fee-for-service can only lead to the substitution of treatment quantity to physician effort, which is inefficient and inflates costs. Consequently, fee-for-service payments should not be used in this case. When moral hazard is combined with the adverse selection issue, the design of the use of fee-for-service can effectively improve screening. The contractable dimension of the task, the treatment quantity, is paid for the lower ability types. This is the main, and perhaps surprising, difference with the pure moral hazard policy, similar to the findings of Holmstrom and Milgrom (1991) when adverse selection is not considered. This is because the effort is not contractable and can only be indirectly contracted by using the observed outcomes, that is, the success rate. In the presence of adverse selection, the informational rents of the better types can be mitigated by reducing the performance pay of the less efficient types. In order to give partial incentives for exerting effort, the payer has to reduce the pay-for-performance. However the reduced pay-for-performance may unintentionally induce an excessive decrease in treatment quantity which reduces quality. To avoid an excessive degradation in quality, it is then desirable to also use the fee-for-service payment.[10]

Whether the adverse selection issue is important or not depends on the volatility of the realization of the type. Smaller (larger) the support of the type $[\theta_0, \theta_1]$ is, less (more) volatile is the type's realization, and less (more) important is the adverse selection issue. When screening is an important issue, it is desirable to use the fee-for-service payment method to avoid excessive degradations in quality.

[10]Indeed, if under other forms of probability function the treatment quantities for the lower types are distorted upwards in the second best, it even reinforces our result that the fee-for-service is positive for these types in order to encourage treatment quantity. The proof is provided in the appendix.

Moreover, we observe from (31) that the larger the distortion on treatment, the lower the fee-for-service; the smaller the distortion on treatment, the higher the fee-for-service. It follows directly from equation (31) which states that the level of fee-for-service depends exactly on the magnitude of the downward distortion on treatment quantity. If the distortion is larger in the second best, it is better to give a lower fee-for-service; if the distortion is small, a higher fee-for-service is needed to guarantee a sufficient level of quality.

To summarize, although moral hazard calls for no incentive pay on the contractable dimension, adverse selection requires a continued use of this incentive pay. Due to the fact that effort can only be indirectly contracted by the pay-for-performance, the incentive payment on the contractable treatment quantity is used as an instrument to correct excessive distortions.

6 Discussion

6.1 Time lag and noisy observation of performance

Our work can also be extended to the cases where the performance is a noisy observation. This could occur when heterogeneity of patients within one certain DRG is considered. Patients may differ in their severities even within one DRG, and consequently they may recover at different moments. At the moment of survey, a patient may recover after the survey, therefore this data is missing. Or, it seems that he has recovered but actually he has not. Hence, the noisy observation of P depends on the patients' severities.

Within our framework, we propose a variation of our model. Let γ denote the severity within one DRG, being distributed according to a cumulative function $G(.)$. We only observe P': $P' = P + \varepsilon(\gamma)$ where $\varepsilon(.)$ measures the noise from the heterogeneity of severities. Then, when only moral hazard is considered, the payer can offer a payment scheme such that $S'(x, P') = xP'b - x\mathbb{E}(\varepsilon(.))b - M$, where $\mathbb{E}(\varepsilon(.))$ is the expectation of the noise. Hence, noisy observation is not a problem for the pure moral hazard if $\mathbb{E}(\varepsilon(.))$ is known. There is still no need for fee-for-service. When both moral hazard and adverse selection are considered, noisy observation would not be a problem if other higher moments of $\varepsilon(.)$ can be estimated from medical statistics. The second best non-linear compensation scheme can be implemented through a menu of quadratic schemes which are corrected by the moments of the degree of the quadratic schemes. This would be an application of the polynomial approximation method proposed by Caillaud, Guesnerie and Rey (1992), in regard to physician compensation. In practice, if we can estimate the distribution of patients' severities in the population within one certain DRG, and then try to estimate how the noise of performance observation depends on the

patients' severities (linearly, polynomially, exponentially, etc) from medical statistics, then we can first find the menu of quadratic schemes to approximate our optimal non-linear scheme, and correct it by its moments. And finally, pay-for-performance based on the noisy observation would implement our second best optimal allocations.

7 Conclusion

This paper examines the optimal compensation rule for physicians under three payment methods: pay-for-performance, fee-for-service and capitation in the presence of both adverse selection and moral hazard on the supply side. We identify the screening effect of fee-for-service. We provide an argument for the criticism on the shortcomings of fee-for-service. More importantly, we also provide a rationale for the continued use of fee-for-service payment from the perspective of the adverse selection issue. We show that fee-for-service induces substitution of either treatment quantity and/or number of patients to physician effort. When moral hazard is the only problem, the optimal compensation policy includes a capitation payment and pay-for-performance without fee-for-service. When moral hazard is combined with the adverse selection issue, in order to avoid an excessive degradation in quality, efficient screening requires a continued use of fee-for-service for the lower productivity physicians and less pay-for-performance.

This paper has proposed a framework by which to study the pay-for-performance incentive programs. A possible extension would be to study an alternative pay-for-performance system also used in practice which rewards physicians according to how well they perform relative to their peers on various quality or cost measures. To summarize, pay-for-performance remains an area of possible promising advances, which are worthy of significant new research.

References

[1] Alger I. and C.-t.A. Ma, 2003, Moral hazard, insurance, and some collusion, Journal of Economic Behavior and Organization, 50, 225-247.

[2] Allen R. and P. Gertler, 1991, Regulation and the provision of quality to heterogenous consumers: the case of prospective pricing of medical services, National Bureau of Economic Research, Working Paper No. 2269.

[3] Caillaud B., R. Guesnerie and P. Rey, 1992, Noisy observation in adverse selection models, The Review of Economic Studies, 59, 3, 595-61

[4] Chalkley M. and J.M. Malcomson, 1998 a, Contracting for health services with unmonitored quality, The Economic Journal, 108, 1093-111

[5] Chalkley M. and J.M. Malcomson, 1998 b, Contracting for health services when patient demand does not reflect quality, Journal of Health Economics, 17, 1-1

[6] Eggleston, K., 2005, Multitasking and mixed systems for provider payment, Journal of Health Economics, 24(1), 211-223.

[7] Ellis R.P. and T.G. McGuire, 1986, Provider behavior under prospective reimbursement cost sharing and supply, Journal of Health Economics, 129-151.

[8] Holmstrom B. and P. Milgrom, 1991, Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, Journal of Law, Economics and Organization, Oxford University Press, 7(0), 24-52, Special I.

[9] Ma C.-t.A., 1994, Health care payment systems: Cost and quality incentives, Journal of Economics and Management Strategy, 3 (1), 93-112.

[10] Ma C.-t.A. and T.G. McGuire, 1997, Optimal health insurance and provider payment, The American Economic Review, 685-704.

[11] McGuire T.G., 2000, Physician agency, Handbook of Health Economics.

[12] Miller G. and K.S. Babiarz, 2013, Pay-for-performance incentives in low- and middle-income country health programs, National bureau of economic research, Working Paper 18932.

[13] Rosenthal MB., RG. Frank, Z. Li and AM. Epstein, 2005, Early evidence with pay-for-performance: from concept to practice, Journal of American Medical Association, 294, 1788-1793.

[14] Rosenthal, MB., BE. Landon, SLT. Normand, RG. Frank and AM. Epstein, 2006, Pay-for-performance in commercial HMOs, New England Journal of Medicine, 355, 1895-1902.

[15] Rosenthal, M.B., Fernandopulle, R., Song, H.R., Landon, B., 2004, Paying for quality: providers' incentives for quality improvement. Health Affairs 23 (2), 127-141.

[16] The Commonwealth Fund, 2010, International Profiles of Health Care Systems.

[17] UnitedHealth Center for Health Reform and Modernization, 2012, Farewell to fee-for-service? A real world strategy for health care payment reform, Working Paper 8.

[18] World Health Organization, 2010, World Health Statistics.

A Appendix

A.1 Second order condition for the first order approach

The local incentive constraint (20) implies that the following first order condition for the optimal report $\hat{\theta}$ chosen by type θ is satisfied: $\forall\theta \in \Theta$

$$\begin{aligned}
&\dot{S}(\hat{\theta}) - C'(x(\hat{\theta})q(\hat{\theta}))[\dot{x}(\hat{\theta})q(\hat{\theta}) + x(\hat{\theta})\dot{q}(\hat{\theta})] \\
&- \varphi'(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha))\{\dot{x}(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha) \\
&+ x(\hat{\theta})[e_{P(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta), \alpha]\dot{P}(\hat{\theta}) + e_{q(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\dot{q}(\hat{\theta})]\} \mid_{\hat{\theta}=\theta} = 0. \qquad (37)
\end{aligned}$$

It is also necessary to satisfy the local second order condition,

$$\begin{aligned}
&\ddot{S}(\hat{\theta}) - \{C''(x(\hat{\theta})q(\hat{\theta}))(\dot{x}(\hat{\theta})q(\hat{\theta}) + x(\hat{\theta})\dot{q}(\hat{\theta}))[\dot{x}(\hat{\theta})q(\hat{\theta}) + x(\hat{\theta})\dot{q}(\hat{\theta})] \\
&+ C'(x(\hat{\theta})q(\hat{\theta}))[\ddot{x}(\hat{\theta})q(\hat{\theta}) + \dot{x}(\hat{\theta})\dot{q}(\hat{\theta}) + \dot{x}(\hat{\theta})\dot{q}(\hat{\theta}) + x(\hat{\theta})\ddot{q}(\hat{\theta})]\} \\
&- \varphi''(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha))[\dot{x}(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha) \\
&+ x(\hat{\theta})[e_{P(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\dot{P}(\hat{\theta}) + e_{q(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\dot{q}(\hat{\theta})]] \\
&\{\dot{x}(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha) \\
&+ x(\hat{\theta})[e_{P(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta), \alpha]\dot{P}(\hat{\theta}) + e_{q(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\dot{q}(\hat{\theta})]\} \\
&- \varphi'(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha))\{\ddot{x}(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\} \\
&+ \dot{x}(\hat{\theta})[e_{P(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\dot{P(\hat{\theta})} + e_{q(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\dot{q}(\hat{\theta})] \\
&+ \dot{x}(\hat{\theta})[e_{P(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta), \alpha]\dot{P}(\hat{\theta}) + e_{q(\hat{\theta})}(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha)\dot{q}(\hat{\theta})] \\
&+ x(\hat{\theta})[(e_{P(\hat{\theta})P(\hat{\theta})}\dot{P}(\hat{\theta}) + e_{P(\hat{\theta})q(\hat{\theta})}\dot{q}(\hat{\theta}))\dot{P}(\hat{\theta})] \\
&+ e_{P(\hat{\theta})}\ddot{P}(\hat{\theta}) + (e_{q(\hat{\theta})P(\hat{\theta})}\dot{P}(\hat{\theta}) + e_{q(\hat{\theta})q(\hat{\theta})}\dot{q}(\hat{\theta}))\dot{q}(\hat{\theta}) + e_{q(\hat{\theta})}\ddot{q}(\hat{\theta})\} \mid_{\hat{\theta}=\theta} \leq 0. \qquad (38)
\end{aligned}$$

But totally differentiating (37) with respect to θ, (38) can be written more simply as

$$\begin{aligned}&-\varphi''(xe)xe_\theta\{\dot{x}(\theta)e+x(\theta)[e_P\dot{P}(\theta)+e_q\dot{q}(\theta)]\}\\&-\varphi'(xe)\{\dot{x}(\theta)e_\theta+x(\theta)[e_{P\theta}\dot{P}(\theta)+e_{q\theta}\dot{q}(\theta)]\}\geq 0,\end{aligned}$$

By simplifying it, we have the equation (38) being written as

$$-\{\Delta_x\dot{x}(\theta)+x(\theta)[\frac{\Delta_e}{P_e}+\varphi'(xe)(-\frac{P_{e\theta}}{p_e^2})]\dot{P}(\theta)+[\Delta_q-\frac{\Delta_e}{P_e}P_q]\dot{q}(\theta)\}\geq 0, \tag{39}$$

where Δ_x, Δ_q, Δ_e was defined by equation (26)(27) and (28). As it will be proved in Appendix $A.3$ that $\Delta_x<0$, $\Delta_q<0$ and $\Delta_e<0$; Hence, $[\frac{\Delta_e}{P_e}+\varphi'(xe)(-\frac{P_{e\theta}}{p_e^2})]<0$. In appendix $A.6$, we show that $[\Delta_q-\frac{\Delta_e}{P_e}P_q]>0$. Thus, a sufficient condition for that the local second order condition (39) to be satisfied is that $\dot{x}(\theta)>0$, $\dot{P}(\theta)>0$ and $\dot{q}(\theta)<0$.

A.2 The Hamiltonian

In this section we solve the second best program using the Hamiltonian approach. Applying Envelop Theorem to the incentive contraint (20) implies that

$$\dot{\pi}(\theta)=-x\varphi'\big(xe(P,q,\theta;\alpha)\big)\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta}.$$

It follows that $\dot{\pi}(\theta)\geq 0$.[11]

Deleting the argument inside the parenthesis refering to the individual type, the Hamiltonian function writes as:

$$\mathcal{H}(\theta,\pi,x,q,P,\lambda;\alpha)=\Big[xPb-\pi-C(xq)-\varphi(xe(P,q,\theta;\alpha))\Big]f(\theta)+\lambda(\theta)\Big[-\varphi'(xe(P,q,\theta;\alpha))\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta}\Big], \tag{40}$$

where the state variable is π, the control variables are x,q,P, the co-state variable is λ. The Hamilton-Jacobi system yields that

$$\dot{\lambda}^{SB}(\theta)=-\frac{\partial\mathcal{H}}{\partial\pi}=f(\theta). \tag{41}$$

Integrating equation (41) and using the transversality condition $\lambda^{SB}(\theta_1)=0$, we have

$$\lambda^{SB}(\theta)=F(\theta)-1. \tag{42}$$

[11]This is because $\varphi'(xe)>0$, and $\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta}$ is the partial derivative of effort with respect to the ability given the recovery probability and the treatment quantity. Define $P(e,q,\theta;\alpha)=\bar{P}$. Applying the implicit function theorem to this equation, we have $\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta}=-\frac{P_\theta(e,q,\theta;\alpha)}{P_e(e,q,\theta;\alpha)}$ which is negative.

For each type of physician, the effort, type, treatment quantity and the recovery probability is deterministically linked by the relationship $P = P(e, q, \theta; \alpha)$. Optimizing with respect to the P, x, q amounts to optimizing with respect to e, x, q. Expressing the payer's objective function in terms of effort instead of the recovery probability and substituting equation (42) into the Hamiltonian (40), we have

$$\mathcal{H}(\theta, \alpha, \pi, x, q, e, \lambda^{SB}) = \Big[xP(e, q, \theta; \alpha)b - \pi - C(xq) - x\varphi(e)\Big] f(\theta) \\ - \big(1 - F(\theta)\big)\Big[- x\varphi'\big(xe(P(e, q, \theta; \alpha), q, \theta; \alpha)\big)\frac{\partial e(P(e, q, \theta; \alpha), q, \theta; \alpha)}{\partial \theta}\Big]. \quad (43)$$

The associated first order conditions can be derived from equation (43).

A.3 Downward distortion of the second best optimal allocations for lower types

In this section, we prove that the second best allocations are distorted downwards for the lower types. For this, we have to prove that

$$\Delta_x = \big(\varphi'(xe) + xe\varphi''(xe)\big)\frac{\partial e(P(e, q, \theta; \alpha), q, \theta; \alpha)}{\partial \theta} < 0,$$
$$\Delta_q = x\varphi'(xe)\frac{\partial^2 e(P(e, q, \theta; \alpha), q, \theta; \alpha)}{\partial \theta \partial q} < 0,$$
$$\Delta_e = \frac{\partial\Big[x\varphi'(xe)\frac{\partial e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial \theta}\Big]}{\partial e} < 0.$$

We first prove that $\Delta_x < 0$. $\frac{\partial e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial \theta}$ is the partial derivative of effort with respect to its own ability θ. By implicit function theorem:

$$\frac{\partial e(P, q, \theta; \alpha)}{\partial \theta} = -\frac{P_\theta(e, q, \theta; \alpha)}{P_e(e, q, \theta; \alpha)} < 0.$$

Moreover, $\big(\varphi'(xe) + xe\varphi''(xe)\big) > 0$. It follows that $\Delta_x < 0$.

We next prove that $\Delta_e < 0$.

$$\Delta_e = x^2\varphi''(xe)\frac{\partial e(P(e, q, \theta; \alpha), q, \theta; \alpha)}{\partial \theta} + x\varphi'(xe)\frac{\partial^2 e(P(e, q, \theta; \alpha), q, \theta; \alpha)}{\partial \theta \partial e}.$$

We have shown that

$$\frac{\partial e(P, q, \theta; \alpha)}{\partial \theta} = -\frac{P_\theta(e, q, \theta; \alpha)}{P_e(e, q, \theta; \alpha)} < 0.$$

Then,

$$\begin{aligned}
&\frac{\partial^2 e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta\partial e} = \frac{\partial^2 e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta\partial P}\frac{\partial\big(P(e,q,\theta;\alpha)\big)}{\partial e} \\
&= \frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial P}}{\partial\theta}\frac{\partial\big(P(e,q,\theta;\alpha)\big)}{\partial e} \\
&= \frac{\partial\left(\frac{1}{\partial P(e,q,\theta;\alpha)/\partial e}\right)}{\partial\theta}\frac{\partial P(e,q,\theta;\alpha)}{\partial e} \\
&= \frac{\partial\left(\frac{1}{\partial P(e,q,\theta;\alpha)/\partial e}\right)}{\partial\theta}\frac{\partial P(e,q,\theta;\alpha)}{\partial e} < 0,
\end{aligned}$$

by the single-crossing property.

Moreover, $\varphi''(xe) > 0$ and $\varphi'(xe) > 0$ by the convexity and the increasing property of the disutility function. Thus, it follows that $\Delta_e < 0$.

Finally, we prove that $\Delta_q < 0$, using the following calculation:

$$\begin{aligned}
&\frac{\partial^2 e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta\partial q} \\
&= \frac{\partial^2 e(P,q,\theta;\alpha)}{\partial\theta\partial P}\frac{\partial\big(P(e,q,\theta;\alpha)\big)}{\partial q} + \frac{\partial^2 e(P,q,\theta;\alpha)}{\partial\theta\partial q} \\
&= \frac{\partial\frac{1}{\partial P(e,q,\theta;\alpha)/\partial e}}{\partial\theta}\frac{\partial P(e,q,\theta;\alpha)}{\partial q} + \frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial q}}{\partial\theta}.
\end{aligned}$$

We observe that the first term is negative following the single-crossing property. Applying the implicit function theorem to the second term, it becomes

$$\frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial q}}{\partial\theta} = -\frac{1}{P_e}(P_{q\theta} - \frac{P_q}{P_e}P_{e\theta}).$$

For CES production function and the exponential function for the recovery probability,

$$\frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial q}}{\partial\theta} = -\frac{1}{P_e}(P_{q\theta} - \frac{P_q}{P_e}P_{e\theta}) = 0.$$

Moreover, $\varphi'(xe) > 0$. It follows that $\Delta_q < 0$.

A.4 Informational rent

It is straightforward to see that the informational rent

$$[\varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \hat{\theta}; \alpha)) - \varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha))]$$

is increasing in $e(\hat{\theta})$. To see that it is also increasing in $x(\hat{\theta})$ and $q(\hat{\theta})$, since $\Delta_q < 0$ and $\Delta_e < 0$, it is obvious to prove that

$$\frac{\partial[\varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \hat{\theta}; \alpha)) - \varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha))]}{\partial x(\hat{\theta})} > 0,$$
$$\frac{\partial[\varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \hat{\theta}; \alpha)) - \varphi(x(\hat{\theta})e(P(\hat{\theta}), q(\hat{\theta}), \theta; \alpha))]}{\partial q(\hat{\theta})} > 0.$$

A.5 Proof of Lemma 2

From the recovery probability function (3), we obtain

$$P_q = \alpha\theta exp(-\theta f(e,q)) f_q(e,q),$$

$$P_e = \alpha\theta exp(-\theta f(e,q)) f_e(e,q),$$

$$lim_{q\to 0^+, e\to 0^+} f_q(e,q) = lim_{q\to 0^+, e\to 0^+} (1 + q^\rho e^{-\rho})^{\frac{1-\rho}{\rho}} = 2^{\frac{1-\rho}{\rho}},$$
$$lim_{q\to 0^+, e\to 0^+} f_e(e,q) = lim_{q\to 0^+, e\to 0^+} (1 + e^\rho q^{-\rho})^{\frac{1-\rho}{\rho}} = 2^{\frac{1-\rho}{\rho}}.$$

It follows that

$$lim_{q\to 0^+, e\to 0^+} P_q(e,q,\theta;\alpha) = \theta lim_{q\to 0^+, e\to 0^+} f_q(q,e) = 2^{\frac{1-\rho}{\rho}}\alpha\theta,$$
$$lim_{q\to 0^+, e\to 0^+} P_e(e,q,\theta;\alpha) = \theta lim_{q\to 0^+, e\to 0^+} f_e(q,e) = 2^{\frac{1-\rho}{\rho}}\alpha\theta.$$

Therefore, when $\rho \to 0$ the Inada conditions are satisfied: $\forall\theta$

$$P_q(0,0,\theta;\alpha) = +\infty,$$
$$P_e(0,0,\theta;\alpha) = +\infty.$$

Then, if $C'(0) = 0$ and $\varphi'(0) = 0$, exclusion is never desirable.

If $\rho \nrightarrow 0$, Inada conditions are not satisfied:

$$P_q(0,0,\theta;\alpha) < +\infty,$$
$$P_e(0,0,\theta;\alpha) < +\infty.$$

if in addition $C'(0) > 0$ and $\varphi'(0) > 0$, result i) of Lemma 2 follows from equations (24) and (25) together with Spence-Mirrlees conditions, that is, $P_q(0,0,\theta;\alpha)$ and $P_e(0,0,\theta;\alpha)$ are increasing with type. Thus, if there exists $\tilde{\theta}$ such that $\forall x > 0$, $C'(0) \geq P_q(0,0,\tilde{\theta};\alpha)b$ and $\varphi'(0) \geq P_e(0,0,\tilde{\theta};\alpha)b$, then all types that are lower than $\tilde{\theta}$ will satisfy these inequalities. Then equations (24) and (25) imply that for these types, $q = 0$ and $e = 0$. Then, equation (23) implies that $x = 0$.

If $\forall\theta$, $\forall x(\theta) > 0$,

$$C'(0) < P_q(0,0,\theta;\alpha)b,$$
$$\varphi'(0) < P_e(0,0,\theta;\alpha)b,$$

then, suppose that $x(\theta)$ is interior, since the social welfare function is concave in its arguments, we can obtain $q(\theta) = 0$ and $e(\theta) = 0$ if and only if equations (24) and (25) are satisfied with inequalities[12] at $q(\theta) = 0$ and $e(\theta) = 0$. Rewriting these two inequalities evaluated at $q \to 0^+$ and $e \to 0^+$, we have

$$\frac{f(\theta)}{1-F(\theta)} < \frac{-\Delta_q|_{q\to 0^+,e\to 0^+}}{xP_q(e,q,\theta;\alpha)|_{q\to 0^+,e\to 0^+}b - xC'(0)}, \tag{44}$$
$$\frac{f(\theta)}{1-F(\theta)} < \frac{-\Delta_e|_{q\to 0^+,e\to 0^+}}{xP_e(e,q,\theta;\alpha)|_{q\to 0^+,e\to 0^+}b - x\varphi'(0)}. \tag{45}$$

Remember that

$$\Delta_e = x^2\varphi''(xe)\frac{\partial e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta} + x\varphi'(xe)\frac{\partial^2 e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta\partial e},$$

where

$$\frac{\partial^2 e(P(e,q,\theta;\alpha),q,\theta)}{\partial\theta\partial e} = \frac{\partial\left(\frac{1}{\partial P(e,q,\theta;\alpha)/\partial e}\right)}{\partial\theta}\frac{\partial P(e,q,\theta;\alpha)}{\partial e},$$

[12]The reason why we have to consider both inequalities together is as follows: If among (24) and (25) at the optimum, one is satisfied with equality and the other with inequality, this means that one of the allocations (q or e) is zero, while the other is interior. If the optimal x is interior, then this type is not excluded which is not the case that we are interested in. If the optimal x is zero, then this contradicts the assumed equality equation (either (24) or (25)), and this is impossible. Hence, considering one inequality with the other being equality either is not interesting or is impossible. Exclusion must mean that $q = 0$ and $e = 0$ at the same time. Hence (24) and (25) must be satisfied with inequality together to derive the condition of exclusion.

and

$$\begin{aligned}\Delta_q &= x\varphi'(xe)\frac{\partial^2 e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta\partial q}\\ &= x\varphi'(xe)\big[\frac{\partial\frac{1}{\partial P(e,q,\theta;\alpha)/\partial e}}{\partial\theta}\frac{\partial P(e,q,\theta;\alpha)}{\partial q}+\frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial q}}{\partial\theta}\big].\end{aligned}$$

Taking $P(e,q,\theta;\alpha)=1-\alpha exp\big(-\theta f(e,q)\big)$, where $f(e,q)=(e^\rho+q^\rho)^{\frac{1}{\rho}}$ with $\rho\in(-\infty,1]$,

$$\begin{aligned}P_q &= \alpha\theta exp(-\theta f(e,q))f_q(e,q),\\ P_e &= \alpha\theta exp(-\theta f(e,q))f_e(e,q),\\ P_{q\theta} &= \alpha M f_q(e,q),\\ P_{e\theta} &= \alpha M f_e(e,q),\end{aligned}$$

where $M=[exp(-\theta f(e,q))+\theta exp(-\theta f(e,q))(-f(e,q))]f_e(e,q)$. Then, we obtain

$$\frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial q}}{\partial\theta}=-\frac{1}{P_e}(P_{q\theta}-\frac{P_q}{P_e}P_{e\theta})=\frac{1}{P_e}[Mf_q(e,q)-Mf_q(e,q)]=0.$$

Thus,

$$\begin{aligned}\Delta_q|_{q\to0^+,e\to0^+} =& x\varphi'(0)\frac{\partial\frac{1}{\partial P(e,q,\theta;\alpha)/\partial e}}{\partial\theta}\frac{\partial P(e,q,\theta;\alpha)}{\partial q}\\ &= x\varphi'(0)\frac{\partial\frac{1}{\alpha\theta exp(-\theta f(e,q))f_e(e,q)}}{\partial\theta}\alpha\theta exp(-\theta f(e,q))f_q(e,q)\\ &= -x\varphi'(0)\frac{1}{\alpha^2\theta^2exp^2(-\theta f(e,q))f_e^2(e,q)}[\alpha Mf_e(e,q)]\alpha\theta exp(-\theta f(e,q))f_q(e,q)\\ &= -x\varphi'(0)\frac{1}{\theta exp(-\theta f(e,q))f_e(e,q)}Mf_q(e,q)\\ &= -x\varphi'(0)\frac{[1-\theta f(e,q)]f_q(e,q)}{\theta f_e(e,q)}|_{q\to0^+,e\to0^+}.\end{aligned}$$

With CES function,

$$\begin{aligned}&lim_{q\to0^+,e\to0^+}f(e,q)=0,\\ &lim_{q\to0^+,e\to0^+}f_q(e,q)=lim_{q\to0^+,e\to0^+}(1+q^\rho e^{-\rho})^{\frac{1-\rho}{\rho}}=2^{\frac{1-\rho}{\rho}},\\ &lim_{q\to0^+,e\to0^+}f_e(e,q)=lim_{q\to0^+,e\to0^+}(1+e^\rho q^{-\rho})^{\frac{1-\rho}{\rho}}=2^{\frac{1-\rho}{\rho}}.\end{aligned}$$

Hence,

$$lim_{q\to 0^+,e\to 0^+}\Delta_q = -\frac{1}{\theta}x\varphi'(0).$$

Then, the inverse function of $P(e,q,\theta;\alpha)$ given α, q and θ is

$$e(P,q,\theta;\alpha) = \Big[\big[-\frac{1}{\theta}ln\big(\frac{1}{\alpha}(1-P)\big)\big]^{\rho} - q^{\rho}\Big]^{\frac{1}{\rho}}.$$

Then,

$$\begin{aligned}
&\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta}|_{q\to 0^+,e\to 0^+}\\
&= \frac{1}{\rho}\Big[\big[-\frac{1}{\theta}ln\big(\frac{1}{\alpha}(1-P)\big)\big]^{\rho} - q^{\rho}\Big]^{\frac{1}{\rho}-1}\rho\big[-\frac{1}{\theta}ln\big(\frac{1}{\alpha}(1-P)\big)\big]^{\rho-1}\frac{1}{\theta^2}ln\big(\frac{1}{\alpha}(1-P)\big)|_{q\to 0^+,e\to 0^+}\\
&= 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
lim_{q\to 0^+,e\to 0^+}\Delta_e &= x\varphi'(0)\frac{\partial\frac{1}{\partial P(e,q,\theta;\alpha)/\partial e}}{\partial\theta}\frac{\partial P(e,q,\theta;\alpha)}{\partial e}\\
&= -x\varphi'(0)\frac{1}{\alpha^2\theta^2 exp^2(-\theta f(e,q))f_e^2(e,q)}[\alpha M f_e(e,q)]\alpha\theta exp(-\theta f(e,q))f_e(e,q)\\
&= -x\varphi'(0)\frac{[1-\theta f(e,q)]f_e(e,q)}{\theta f_e(e,q)}\\
&= -x\varphi'(0)\frac{1-\theta f(e,q)}{\theta}|_{q\to 0^+,e\to 0^+}\\
&= -\frac{1}{\theta}x\varphi'(0).
\end{aligned}$$

Thus

$$lim_{q\to 0^+,e\to 0^+}\Delta_q = lim_{q\to 0^+,e\to 0^+}\Delta_e = -\frac{1}{\theta}x\varphi'(0).$$

Moreover,

$$\begin{aligned}
lim_{q\to 0^+,e\to 0^+}P_q(e,q,\theta;\alpha) &= \alpha\theta lim_{q\to 0^+,e\to 0^+}f_q(q,e) = 2^{\frac{1-\rho}{\rho}}\alpha\theta,\\
lim_{q\to 0^+,e\to 0^+}P_e(e,q,\theta;\alpha) &= \alpha\theta lim_{q\to 0^+,e\to 0^+}f_e(q,e) = 2^{\frac{1-\rho}{\rho}}\alpha\theta.
\end{aligned}$$

Hence inequalities (44) and (45) evaluated at zero become

$$\frac{f(\theta)}{1-F(\theta)} < \frac{x\varphi'(0)}{\theta(2^{\frac{1-\rho}{\rho}}\alpha\theta xb - xC'(0))}, \tag{46}$$

$$\frac{f(\theta)}{1-F(\theta)} < \frac{x\varphi'(0)}{\theta(2^{\frac{1-\rho}{\rho}}\alpha\theta xb - x\varphi'(0))}. \tag{47}$$

It follows that if

$$\frac{f(\theta)}{1-F(\theta)} < Min\Big\{\frac{\varphi'(0)}{\theta(2^{\frac{1-\rho}{\rho}}\alpha\theta b - C'(0))}, \frac{\varphi'(0)}{\theta(2^{\frac{1-\rho}{\rho}}\alpha\theta b - \varphi'(0))}\Big\} \tag{48}$$

is satisfied, inequalities (46) and (47) are satisfied. Thus, the original equations (24) and (25) are both satisfied with inequalities at $q = 0$ and $e = 0$. Thus, the optimal solutions are indeed $q = 0$ and $e = 0$ for the type θ. Hence, from equation (48), for θ which is close to zero and $f(\theta)$ is close to zero, this condition is surely satisfied because the left hand side is close to zero and the right hand side goes to infinity.

Then, with monotonicity assumption of the hazard rate $\frac{1-F(\theta)}{f(\theta)}$, the left hand side of equation (48) is increasing in θ and the right hand side is decreasing in θ, thus for all types that are lower than this θ, this condition is also satisfied, which implies $q = 0$ and $e = 0$ for these types too. When the optimal treatment and effort are zero, equation (23) with inequality implies that the optimal number of patients is zero, which contradicts what we supposed at the beginning that x is interior. Consequently, with condition (48) being satisfied, it is impossible to have an interior solution. It thus proved result *ii*) of Lemma 2.

A.6 Second best fee-for-service is positive for lower types

Equation (31) is:

$$S_q = \frac{1-F(\theta)}{f(\theta)}(-|\Delta_q| + |\Delta_e|\frac{P_q}{P_e}).$$

We prove that $-|\Delta_q| + |\Delta_e|\frac{P_q}{P_e} = \Delta_q - \Delta_e\frac{P_q}{P_e} > 0$.

$$\Delta_q = x\varphi'(xe)\big[\frac{\partial^2 e(P,q,\theta;\alpha)}{\partial\theta\partial P}\frac{\partial P(e,q,\theta;\alpha)}{\partial q} + \frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial q}}{\partial\theta}\big],$$

where $\frac{\partial\frac{\partial e(P,q,\theta;\alpha)}{\partial q}}{\partial\theta} = 0$ under CES production function[13].

$$\Delta_e = x^2\varphi''(xe)\frac{\partial e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta} + x\varphi'(xe)\frac{\partial^2 e(P(e,q,\theta;\alpha),q,\theta;\alpha)}{\partial\theta\partial P}\frac{\partial\big(P(e,q,\theta;\alpha)\big)}{\partial e}.$$

[13]If under other functions, this term is not zero, then the Δ_q may or may not be negative. The treatment may be downward or upward distorted. If this term is negative, treatment is downward distorted and fee-for-service is positive as we prove in this section. If this term is positive and large enough, $\Delta_q > 0$, the treatment is upward distorted. The term $\Delta_q - \Delta_e\frac{P_q}{P_e}$ will be surely positive because Δ_e is negative under any form of probability function. Hence, fee-for-service is still positive.

Thus,

$$\begin{aligned}
&\Delta_q - \Delta_e \frac{P_q}{P_e} \\
&= x\varphi'(xe)\frac{\partial^2 e(P,q,\theta;\alpha)}{\partial\theta\partial P}\frac{\partial P(e,q,\theta;\alpha)}{\partial q} \\
&\quad - \left(x^2\varphi''(xe)\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta} + x\varphi'(xe)\frac{\partial^2 e(P,q,\theta;\alpha)}{\partial\theta\partial P}\frac{\partial\big(P(e,q,\theta;\alpha)\big)}{\partial e}\right)\frac{P_q}{P_e} \\
&= -x^2\varphi''(xe)\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta} > 0,
\end{aligned}$$

because $\frac{\partial e(P,q,\theta;\alpha)}{\partial\theta} < 0$, $x^2\varphi''(xe) > 0$, $P_q > 0$ and $P_e > 0$.

As a result, $S_q > 0, \forall\tilde{\theta} < \theta < \theta_1$.