

Procurement with specialized firms

Jan Boone and Christoph Schottmüller*

April 12, 2012

Abstract

This paper analyzes optimal procurement mechanisms in a setting typical for post-liberalization industries. The procurement agency has incomplete information concerning the firms' cost functions and cares about quality as well as price. Low type firms are cheaper than high type firms in providing low quality but more expensive when providing high quality. Hence, each type is specialized in a certain quality level. We show that this specialization leads to a bunching of types on profits, i.e. a range of firms with different cost functions receives zero profits and therefore no informational rents. If first best welfare is monotone in the efficiency parameter, the optimal mechanism can be implemented by a simple auction. If first best welfare is U-shaped in type, the optimal mechanism is not efficient in the sense that types providing a lower second best welfare win against types providing a higher second best welfare.

JEL codes: D82, H75, L51

keywords: procurement, specialization, participation constraint

*Boone: Department of Economics, University of Tilburg, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands; Tilec, CentER and CEPR. Schottmüller: Department of Economics, Tilec and CentER, Tilburg University. We gratefully acknowledge financial support from the Dutch National Science Foundation (VICI 453.07.003).

1. Introduction

The literature on procurement and optimal incentive regulation, see for example Laffont and Tirole (1987), Laffont and Tirole (1993) or Che (1993), assumes that firms have private information with regard to its cost function. As usual in screening models, this private information is represented by a “type” which is assumed to be a scalar. It is then assumed that higher types are better in the sense that higher types have lower costs. If costs, for example, depend on the quality produced, this means that a higher type has lower costs for any quality level.

On the other hand, the private information of a firm is often interpreted as the production technology it uses. This technology was determined in the past and can therefore be treated as given in the context of one specific procurement contract. Following this interpretation, one should expect that firms chose production technologies that are not obviously inferior to alternative technologies, i.e. there should be some level of quality for which the technology of a firm is efficient. Put differently, firms are specialized in the production of a certain quality level. As we argue below, this specialization is not covered by standard procurement models which assume that higher types are better in order to obtain a single crossing property.

Put differently, the procurement literature so far focusses on private information concerning absolute efficiency of a firm. We will think of private information in terms of production technology and specialization.

To illustrate what we mean by that, we describe the market for home care in the Netherlands which was recently liberalized. Local governments now procure home care for their citizens and money saved on the procurement can be used now by local governments for other things, like sports facilities (that is, the money received from the central government to pay for home care is not earmarked). However, the local government does have a duty to provide care of some minimum standard. It used to be the case that regional care offices procured without much incentive to save money. Due to this liberalization, new players have entered the market. For example, cleaning companies considered moving into home care. As these new players have no experience with care (to illustrate, they did not use to hire nurses or other professionals with a medical background), they are seen as low quality players. At this low quality level, however, they are cheaper than traditional firms. That is, they can provide simple services like house cleaning and shopping more cheaply than incumbent home care companies. In this sense, incumbents are specialized in high quality production while entrants are specialized in

low quality/low costs production.¹

This pattern –where incumbents are specialized in high quality service while entrants specialize in a low quality (low price) service– is typical after liberalization. Many European countries have liberalized sectors like post, taxis, air transport, railway or local transport. This has led to entry by players who offer lower quality in, for instance, the following sense: only make deliveries twice a week (instead of 6 days a week), drive cars substantially cheaper than a Mercedes (see <http://www.tuktukcompany.nl/> for an example), operate planes with reduced seat pitch and limited on board service as well as offering less connections, use old trains and buses to transport people. A reaction often heard by customers and/or incumbents is that the liberalization is bad for welfare because of the lower quality.²

In each of these cases, one could argue either that quality did not decrease at all or that before liberalization quality was inefficiently high. In the former case, incumbents spread rumors to reduce the probability that entrants win contracts. In the latter case, after liberalization quality goes down but total surplus rises. Presumably, in some of the examples mentioned either of these two cases arise. However, we are interested in the case where indeed entrants offer both lower quality and lower total surplus than incumbents.³ The question we ask is: How should a planner (in the home care example: the municipality) who wants to maximize welfare optimally organize the procurement in the face of such low quality entrants?

We show the following results. Think of low (high) type firms as firms specialized in low (high) quality production. First, if low types (e.g. entrants in the examples above) are indeed worse high types (incumbents in the examples) with respect to first best welfare, the incumbents (under the optimal procurement rules) do not lose from entry. Second, only if first best welfare first decreases and then increases in type, types specialized in high quality can lose in the

¹To a certain extent this can be resolved through market separation in *care* and *support*. People who do not need medical attention but only someone to clean their home, can be served by cleaning companies. While patients who stay at home and need a nurse can be served by the incumbents. Hence at the extremes of the home care spectrum, market separation can alleviate the issue. However, many cases in home care are not so clear cut. To illustrate, a nurse helping an elderly woman putting on her clothes in the morning and cleaning the house may recognize the first signs of dementia that would be overlooked by an employee of a cleaning company. In such a separated market, what we write above applies to the *support* segment of the market.

²One could even take a broader point of view and also consider the case of foreign workers entering a domestic labor market. In the EU there was a heated debate about Polish workers coming to the west in case Poland would join the EU. Again some people argued that this is bad since Polish workers are supposedly less qualified than domestic workers.

³There are two reasons for this focus. First, as argued below, in the home care example mentioned above there is evidence suggesting that the entrants offer lower surplus. Second, if the entrants offer higher surplus than the incumbents, it is clear that they should be used by a value maximizing planner. Further, if the cheaper entrants offer higher surplus, no service should be procured from the incumbents.

following way: A low quality provider (entrant) can win the procurement even though the high quality provider (incumbent) would provide higher welfare under the optimal procurement rules. Third, in this latter case, quality is distorted above first best for some types and below first best for others. Fourth, in both cases an interval of types has zero profits (“profit bunching”). Although all types in this interval have zero profits, they produce different qualities when winning the contract. Put differently, a mass of types will have no economic rents under the optimal contract although types are perfectly separated in equilibrium. Fifth, if first best welfare is monotone in type, relatively simple auctions can implement the optimal mechanism.

Technically speaking, a contribution of the paper is to solve a two-dimensional mechanism design problem. A technical challenge is that local incentive compatibility is not straightforwardly sufficient for non-local incentive compatibility, i.e. non-local incentive constraints have to be checked explicitly. To illustrate the problem, view profits as a function of the probability of getting the contract. The assumption that firms are specialized implies then that “marginal profits” (where marginal refers to a slightly higher probability of getting the contract) are not monotone in type. This is equivalent to a violation of single crossing in one dimensional models. As is well known, non-local incentive compatibility does not follow from local incentive compatibility if single crossing is not satisfied.

Our paper is related to the literature on procurement, especially to those papers in which more than price matters, e.g. Laffont and Tirole (1987), Che (1993), Branco (1997), Asker and Cantillon (2008) or Asker and Cantillon (2010). This literature shows how quality (or quantity) is distorted away from first best for rent extraction purposes. It also analyzes how simple auctions can implement the optimal mechanism. These papers assume that firms are not specialized, i.e. higher types have lower costs for all quality levels. This assumption seems to be too strong in many settings, e.g. newly liberalized industries. We show that relaxing it leads to zero economic rents for a mass of types which is, to our knowledge, a new result in the literature on procurement auctions. We also show that implementation of the optimal mechanism by standard auctions, e.g. scoring rule actions, is no longer straightforward when firms are specialized.

Our paper connects the literature on competitive procurement with the literature on countervailing incentives, see Lewis and Sappington (1989) for the seminal contribution and Jullien (2000) for the most general treatment. By assuming that firms are specialized, our paper uses a cost function that resembles the utility functions of the countervailing incentives literature. Our result that the participation constraint is binding for a mass of types is also typical for

this literature. We contribute by allowing for several agents bidding for the contract while the countervailing incentive literature focuses on settings with one principal and one agent. As a consequence of this one agent setting, the probability of being contracted is one for the agent. Consequently, local incentive compatibility constraints are sufficient for non-local incentive compatibility and many of the technical challenges encountered in our paper do not occur. From an applied point of view, having more than one firm leads to the result that optimal procurement auctions can be second best inefficient.

An exception from the focus on one agent in the countervailing incentives literature are papers on auctions with valuation dependent externalities, see Carrillo (1998), Figueroa and Skreta (2009) or Brocas (2011). The outside option is type dependent since the agent suffers the externality even if he does not participate. The main difference between our paper and this literature is the existence of another variable, i.e. quality in our paper, while the auction literature focuses on the problem of allocating one exogenously given good. Hence, the only variable is the probability of getting the good. Since the used preferences satisfy single crossing, non-local incentive constraints play again no role.

As we solve a mechanism design problem with two variables, i.e. quality and the probability of winning, our paper is also related to the literature on multidimensional screening as surveyed in Rochet and Stole (2003). We contribute here by analyzing a two-dimensional screening model with countervailing incentives. Other screening models with one-dimensional type and multidimensional decisions include, for example, Matthews and Moore (1987) or Guesnerie and Laffont (1984). In these papers, single crossing is assumed in each dimension which rules out the specialization we have in mind.

The set up of the paper is as follows. In section 2, we present the model. Section 3 analyzes the case where first best welfare is monotonically increasing in type while section 4 deals with U-shaped first best welfare. In the latter case we find a discrimination result, i.e. some types with lower second best welfare are preferred to types with higher second best welfare. Section 5 shows how the model extends to situations in which the assumptions of section 2 are not met and section 6 concludes. Proofs are relegated to the appendix.

2. Model

We consider the case where a social planner procures a service of quality $q \in Q \subset \mathbb{R}_+$ where Q is a convex set. The gross value of this service is denoted by $S(q)$ where we normalize quality in such a way that $S(q) = Sq$ for some $S > 0$.⁴ The cost of production is denoted by the three times continuously differentiable cost function $c(q, \theta)$ where a firm's type θ is private information of the firm. We assume that each firm's type is drawn independently from a distribution F on $[\underline{\theta}, \bar{\theta}]$ which has a strictly positive density f . We also assume that c is (at least) three times continuously differentiable.

We make the following assumptions on the cost function c and distribution function F .

Assumption 1 *We assume that*

- *the function $c(q, \theta)$ satisfies $c_q, c_{qq} > 0, c_{q\theta} < 0, c_{\theta\theta} \geq 0$,*
- *for $q \in Q$ it is the case that S is high enough compared to $c(q, \theta)$ that the planner always wishes to procure (regardless of the type realization) and*
- *the function F satisfies $\frac{d((1-F(\theta))/f(\theta))}{d\theta} < 0$ and $\frac{d(F(\theta)/f(\theta))}{d\theta} > 0$.*

These assumptions are standard in the literature. The first part says that c is increasing and convex in q . Higher θ implies lower marginal costs c_q (the Spence-Mirrlees condition) and c is convex in θ . It will become clear that this convexity is part of the idea of specialized firms. The second assumption formalizes the idea in our home care application that the government cannot decide not to provide the service. That is, it is always socially desirable for the service to be supplied. The third part is the monotone hazard rate (MHR) assumption. Usually this assumption is only made “in one direction”. However, in the literature on countervailing incentives it is standard to have MHR “in both directions”, see for example Lewis and Sappington (1989) or Jullien (2000). Well known distributions that satisfy MHR include normal, uniform and exponential distributions.⁵ In section 5, we discuss what happens if MHR is not satisfied.

The following assumption states that firms are specialized which is the case we want to analyze in this paper.

⁴This is, given our assumptions on the cost function, without loss of generality for weakly concave gross values $S(q)$.

⁵See Bagnoli and Bergstrom (2005) for a more complete overview.

Assumption 2 For each $\theta \in [\underline{\theta}, \bar{\theta}]$, there exists $k(\theta) > 0$ such that

$$c_\theta(q, \theta) \begin{cases} > 0 & \text{if } q < k(\theta) \\ < 0 & \text{if } q > k(\theta) \end{cases}$$

Further,

$$c_{q\theta\theta}(q, \theta) \begin{cases} \leq 0 & \text{if } q < k(\theta) \\ \geq 0 & \text{if } q > k(\theta) \end{cases}$$

$$c_{qq\theta}(q, \theta) \begin{cases} \geq 0 & \text{if } q < k(\theta) \\ \leq 0 & \text{if } q > k(\theta) \end{cases}$$

Hence, for high values of q , a higher type θ produces q more cheaply. This is the usual assumption. We allow for the possibility where low values of q are actually more cheaply produced by lower θ types. To illustrate, high θ incumbents may have invested in (human) capital that makes it actually relatively expensive to produce low quality. If the quality of the product is mainly determined by the qualification of the staff, incumbents might have more expensive but also more qualified workers. Replacing these workers is, especially in Europe, costly because of labor market rigidities and search costs. Consequently, it is more expensive for incumbents to produce low q than for entrants (and the other way around for high q). The function $k(\theta)$ is implicitly defined by $c_\theta(k, \theta) = 0$. By assumption 1, $k(\theta)$ is differentiable and monotonically increasing. Put differently, as θ increases the quality level $k(\theta)$ where $c_\theta = 0$ (weakly) increases.

In some sense, our assumption that c_θ switches sign in q follows naturally from the sorting condition $c_{q\theta} < 0$. However, it is the main departure from the existing literature on procurement which assumes $c_\theta < 0$ or equivalently that $k(\theta) < 0$ which implies that $c_\theta < 0$ in the relevant domain. Put differently, the existing literature assumes that types can be ranked in terms of efficiency irrespective of q . We allow efficiency advantages to depend on q and therefore firms can be specialized in producing a certain quality.⁶

To make sure that (i) the planner's objective function is concave in q and (ii) quality q increases in type, it is standard in the literature to make assumptions on third derivatives $c_{q\theta\theta}, c_{qq\theta}$. If c_θ does not switch sign, the usual assumption is that these derivatives should not switch sign either. This is different in our case. To ease the exposition we make the assumptions on the third derivatives above and discuss in section 5 what changes if these assumptions are

⁶If q is interpreted as quantity, we allow firms to be specialized in a certain scale of production.

not satisfied. Note that we allow for the simple case where these third derivatives are equal to zero.

As c_θ can be positive, it is not clear how first best welfare varies with θ . Below we define the two cases that we consider here. In order to do this, we introduce the following notation. First best output is defined as

$$q^{fb}(\theta) = \arg \max_q Sq - c(q, \theta) \quad (1)$$

which is a singleton as $c_{qq} > 0$ by assumption 1. First best welfare is denoted by

$$W^{fb}(\theta) = Sq^{fb}(\theta) - c(q^{fb}(\theta), \theta). \quad (2)$$

Our final assumption makes sure that we can focus on two relevant cases only.

Assumption 3 *Assume that $c_{\theta\theta}^2 > c_{\theta\theta}c_{qq}$.*

This assumption implies that first best welfare is convex in θ . Hence, we only need to consider two cases. Either first best welfare is increasing in θ or it is first decreasing and then increasing in θ . Further, we can show that first best quality increases faster with θ than $k(\theta)$ and therefore k can intersect q^{fb} at (at most) one type; a result that we use below.

Lemma 1 *First best welfare $W^{fb}(\theta)$ is convex in θ and $q_\theta^{fb}(\theta) > k_\theta(\theta)$.*

Now we define the two cases that we focus on in this paper.

Definition 1 *We consider the two cases*

(WM) *where first best welfare is monotone in θ : $\frac{dW^{fb}(\theta)}{d\theta} > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ or*

(WNM) *where a θ_w exists such that $\frac{dW^{fb}(\theta)}{d\theta} < 0$ for $\theta \in [\underline{\theta}, \theta_w)$ and $\frac{dW^{fb}(\theta)}{d\theta} > 0$ for $\theta \in (\theta_w, \bar{\theta}]$; further $W^{fb}(\bar{\theta}) > W^{fb}(\underline{\theta})$.*

Hence, we exclude the case where $W^{fb}(\bar{\theta}) < W^{fb}(\underline{\theta})$ (and by lemma 1 this is the only case we exclude). In words, we keep on thinking of high (enough) θ as better.⁷

The following two examples give cost and surplus functions that correspond to cases (WM) and (WNM) resp.

⁷It will become clear that the opposite case with $W^{fb}(\bar{\theta}) < W^{fb}(\underline{\theta})$ is symmetric and does not need to be considered separately.

Example 1 Assume $S(q) = q$ and $c(q, \theta) = (q - \theta)^2 + q(1 - \theta/2)$ where θ is distributed uniformly on $[0, 1]$. With these functions $k(\theta) = 4\theta/5$ and $q^{fb}(\theta) = 5\theta/4$. First best welfare is $W^{fb}(\theta) = \frac{9}{16}\theta^2$ which is increasing in $\theta \in [0, 1]$.

The interpretation of this example could be that by the qualification of its staff a firm has the “natural quality level” θ . Producing at different qualities involves adjustment costs that increase with the distance $|q - \theta|$. Additionally, there is a linear cost of quality, e.g. from additional (non-staff) input factors. A high type firm, e.g. a firm that traditionally has had highly qualified staff and therefore is experienced in high quality production, has lower additional costs of quality.

Example 2 Assume $S(q) = Sq$ and $c(q, \theta) = \frac{1}{2}q^2 - \theta q + \theta k$ with $k \in (S + \underline{\theta}, S + \bar{\theta})$. Thus $k(\theta) = k$ in assumption 2. Then we find that $q^{fb}(\theta) = S + \theta$ and $dW^{fb}(\theta)/d\theta = S + \theta - k$. Hence, with $(k - S) \in (\underline{\theta}, \bar{\theta})$ first best welfare increases for $\theta > k - S$ and decreases for $\theta < k - S$.

The second example reflects the standard idea that a firm with high fixed costs ($k\theta$) has lower marginal costs ($c_q = q - \theta$) of producing quality. That is, a firm that produces with a more capital intensive technology might have lower marginal costs for quality but higher fixed costs.

Now we are able to set up the mechanism design problem. The planner only needs one firm to supply the desired service or product. Since $n \geq 2$ firms are able to supply, the planner needs to determine: which firm wins the procurement, what quality level should this firm supply and how much money should be transferred to firms in return for this.

Let $t(\theta)$ denote the (expected) transfer paid by the planner to a firm of type θ and $x(\theta)$ the probability that type θ wins the procurement. That is, the planner offers a menu of choices for firms and each firm chooses the option that maximizes its profits. The planner’s objective is to maximize the expected value of $Sq - t$. The payoff for a type θ player that chooses option (q, x, t) is written as $t - xc(q, \theta)$.⁸

Following Myerson (1981), we use a direct revelation mechanism. That is, we design a menu of choices where $(q(\theta), x(\theta), t(\theta))$ is the choice “meant for” type θ . Then we make it incentive compatible (IC) for type θ to choose this option. That is, it is IC for θ to truthfully reveal his type.

⁸Note that since firms’ and planner’s utility is quasilinear in money, it is without loss of generality to assume that transfer payments t are paid without conditioning on winning: A price p which is paid only when winning the auction is equivalent to an unconditional transfer $t = px$.

Type θ can misrepresent as $\hat{\theta}$ and its profits equal

$$\pi(\hat{\theta}, \theta) = t(\hat{\theta}) - x(\hat{\theta})c(q(\hat{\theta}), \theta) \quad (3)$$

A menu $q(\cdot), x(\cdot), t(\cdot)$ is IC if and only if

$$\Phi(\hat{\theta}, \theta) \equiv \pi(\theta, \theta) - \pi(\hat{\theta}, \theta) \geq 0 \quad (4)$$

for all $\theta, \hat{\theta} \in [\underline{\theta}, \bar{\theta}]$.

With a slight abuse of notation we define the function $\pi(\theta)$ as:

$$\pi(\theta) = \max_{\hat{\theta}} \pi(\hat{\theta}, \theta)$$

Hence, using an envelope argument, incentive compatibility implies

$$\pi_{\theta}(\theta) = -x(\theta)c_{\theta}(q(\theta), \theta) \quad (5)$$

This equation makes sure that the first order condition for truthful revelation of θ is satisfied. The next result derives a tractable form for the local second order condition.

Lemma 2 *For the second order conditions to be locally satisfied, we also need that*

$$x_{\theta}(\theta)c_{\theta}(q(\theta), \theta) + x(\theta)c_{q\theta}(q(\theta), \theta)q_{\theta}(\theta) \leq 0. \quad (\text{SOC})$$

As shown in textbooks like Fudenberg and Tirole (1991), first and second order conditions above imply global IC (as in equation (4)) if $c_{\theta} < 0$ for all $q \in Q$. Because we assume that firms are specialized (assumption 2), local IC does not automatically imply global IC. Hence, we need to verify explicitly below that global IC is satisfied.

Intuitively, assumption 2 is similar to a violation of single crossing. Viewing firm's payoff, $t - xc(q, \theta)$ as a function of x , the standard single crossing assumption would require that the derivative of $t - xc(q, \theta)$ with respect to x is monotone in type, i.e. single crossing would require that c_{θ} does not change sign. But assumption 2 states exactly the opposite. It is well known that in models without single crossing non-local IC can become relevant, see for example Araujo and Moreira (2010) or Schottmüller (2011). We will first neglect these non-local incentive constraints and verify ex post that they do not bind. Although there are some issues with defining single crossing in multidimensional models (see for example McAfee and McMillan (1988)), we refer to c_{θ} switching sign as a “violation of single crossing”.

Finally, because c_θ can switch sign, it is not clear that $\pi(\underline{\theta}) = 0$ under the optimal mechanism. That is, we cannot rule out that $\pi(\underline{\theta}) > 0$ while $\pi(\theta) = 0$ for some $\theta > \underline{\theta}$. Hence, we need to explicitly track the individual rationality constraint

$$\pi(\theta) \geq 0 \quad (6)$$

where we normalize firms' outside option to zero.

We assume that the planner maximizes utility Sq minus the transfer paid to firms. If the planner assigns the project to player i with probability x^i where i produces quality q^i and receives transfer t^i , the planner's utility from i can be written as $x^i Sq^i - t^i = x^i(Sq^i - c^i) - \pi^i$. Above, we did not index q and x by $i = 1 \dots n$ although we have n firms. It will be shown now that this is indeed unnecessary because of the symmetry of the problem. To do so, we write the planner's optimization problem⁹ including the firm identifier i

$$\begin{aligned} & \max_{q^i, x^i, \pi^i} \int_{\underline{\theta}}^{\bar{\theta}} \dots \int_{\underline{\theta}}^{\bar{\theta}} \sum_{i=1}^n f(\theta^1) \dots f(\theta^n) / f(\theta^i) \{ f(\theta^i) [x^i(\Theta)(Sq^i(\theta^i) - c(q^i(\theta^i), \theta^i)) - \pi^i(\Theta)] \\ & + \lambda^i(\theta^i)(\pi_{\theta^i}^i(\Theta) + x^i(\Theta)c_{\theta^i}(q^i(\theta^i), \theta^i)) \\ & - \mu^i(\theta^i)(x_{\theta^i}^i(\Theta)c_{\theta^i}(q^i(\theta^i), \theta^i) + x^i(\Theta)c_{q\theta^i}(q^i(\theta^i), \theta^i)q_{\theta^i}^i(\theta^i)) \\ & + \eta^i(\theta)\pi^i(\Theta) \} - \sum_i \{ \tau^i(\Theta)x^i(\Theta) \} + \sigma(\Theta) \left(1 - \sum_i x^i(\Theta) \right) d\theta_1 \dots d\theta_n \end{aligned}$$

where $\lambda^i(\cdot)$ and $\mu^i(\cdot)$, $\eta^i(\cdot) \geq 0$ are the Lagrange multipliers (co-state variables) of the constraints (5), (SOC) and (6). Here, $x^i(\Theta)$ denotes the probability of firm i being contracted when types are $\Theta = (\theta^1 \dots \theta^n)$. The last constraint ensures that probabilities sum to no more than 1. Because of assumption 1, this constraint will bind and $\sigma(\Theta)$ will therefore be positive. The second but last term secures nonnegativity of the contracting probabilities where the Lagrange multiplier $\tau^i(\Theta) \geq 0$.

The Euler equation for $x^i(\Theta)$ can be rewritten as

$$\begin{aligned} f(\theta^1) \dots f(\theta^n) / f(\theta^i) \{ f(\theta^i) [Sq^i(\theta^i) - c(q^i(\theta^i), \theta^i)] + \lambda^i(\theta^i)c_{\theta^i}(q^i(\theta^i), \theta^i) \\ + \mu_{\theta^i}^i(\theta^i)c_{\theta^i}(q^i(\theta^i), \theta^i) + \mu^i(\theta^i)c_{q\theta^i}(q^i(\theta^i), \theta^i) \} = \sigma(\Theta) + \tau^i(\Theta). \quad (7) \end{aligned}$$

As the objective function is linear in $x^i(\cdot)$, we get what is called a ‘‘bang-bang’’ solution in optimal control theory: For any Θ , the firm i with the highest left hand side in (7) is contracted, i.e. $x^i(\Theta) = 1$, while the other firms are not, i.e. $x^j(\Theta) = 0$ for all $j \neq i$.

⁹We immediately focus on the case with non-random qualities, i.e. each type's quality is a deterministic function of his type only. Appendix E in Jullien (2000) can be easily adapted to our setting to show that optimal mechanisms are indeed deterministic under our assumptions.

With this simple structure for the decision $x(\Theta)$, the maximization problem is totally symmetric across all i . In particular, the first order conditions for $q^i(\cdot)$ and $\pi^i(\cdot)$ are the same for all i . Consequently, we can use a notationally much simpler formulation of the maximization problem

$$\begin{aligned} & \max \int_{\underline{\theta}}^{\bar{\theta}} f(\theta)[x(\theta)(Sq(\theta) - c(q(\theta), \theta)) - \pi(\theta)] \\ & + \lambda(\theta)(\pi_\theta(\theta) + x(\theta)c_\theta(q(\theta), \theta)) \\ & - \mu(\theta)(x_\theta(\theta)c_\theta(q(\theta), \theta) + x(\theta)c_{q\theta}(q(\theta), \theta)q_\theta(\theta)) \\ & + \eta(\theta)\pi(\theta)d\theta \end{aligned} \quad (8)$$

where $\lambda(\cdot)$ and $\mu(\cdot), \eta(\cdot) \geq 0$ are the Lagrange multipliers (co-state variables) of the constraints (5), (SOC) and (6) respectively.

The Euler equation for $\pi(\cdot)$ implies

$$\lambda_\theta(\theta) = -f(\theta) + \eta(\theta) \quad (9)$$

The first order condition for $q(\cdot)$ can be written as

$$f(\theta)(S - c_q(q(\theta), \theta)) + \lambda(\theta)c_{q\theta}(q(\theta), \theta) + \mu(\theta)c_{q\theta\theta}(q(\theta), \theta) = -\mu_\theta(\theta)c_{q\theta}(q(\theta), \theta). \quad (10)$$

Define the virtual valuation of type θ as

$$VV(\theta) = Sq(\theta) - c(q(\theta), \theta) + \frac{\lambda(\theta)}{f(\theta)}c_\theta(q(\theta), \theta). \quad (11)$$

If constraint (SOC) is not binding, the planner's objective function is linear in $x(\theta)$, where $x(\theta)$ is multiplied by $VV(\theta)$. Hence, using standard arguments, the firm with the highest VV wins the procurement contract. The virtual valuation includes next to the first best welfare a rent extraction term. Roughly speaking, contracting a type more often, i.e. increasing $x(\theta)$, changes the slope of the rent function $\pi(\theta)$; see equation (5). If, for example, the incentive constraints is downward binding and the rent function is increasing more steeply, types above θ will get a higher rent. $\lambda(\theta)$ is basically the weight of the types that benefit from such a change.

The following two lemmas are useful in the analysis below. The first lemma establishes that we have a monotone hazard rate property for our case with specialized firms.

Lemma 3 *If either*

(i) $\lambda(\theta) \geq 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $\lambda(\bar{\theta}) = 0$ or

(ii) $\lambda(\underline{\theta}) = \lambda(\bar{\theta}) = 0$,

then

$$\frac{d(\lambda(\theta)/f(\theta))}{d\theta} < 0 \quad (12)$$

for values of θ with $\eta(\theta) = 0$.

As we will see below, the property in equation (12) is useful to have. It is part of the set of conditions to make quality q monotone in θ . Case (i) is relevant for the (WM) case and case (ii) for (WNM). If $\eta(\theta) > 0$, it turns out that the monotonicity of quality is easy to prove (see the discussion of profit bunching below).

Lemma 4 Assume $\mu(\theta) = 0$ and $\frac{d(\lambda(\theta)/f(\theta))}{d\theta} < 0$ for all θ with $\eta(\theta) = 0$. Then

1. if there is $\hat{\theta}$ such that $q(\hat{\theta}) = k(\hat{\theta})$ then $q_\theta(\hat{\theta}) \geq k_\theta(\hat{\theta})$,
2. if there is θ' such that $c_\theta(q(\theta'), \theta') \leq 0$ then $c_\theta(q(\theta), \theta) \leq 0$ for all $\theta > \theta'$ and
3. if there exist $\theta_1, \theta_2 > \theta_1$ with $\pi(\theta_1) = \pi(\theta_2) = 0$ then $\pi(\theta) = 0$ for all $\theta \in [\theta_1, \theta_2]$.

The first result says that if q and k coincide for some value $\hat{\theta}$, then it cannot be the case that k exceeds q for higher values of θ . Further, it is the case that once $c_\theta \leq 0$ for the optimal $q(\theta)$ then c_θ stays non-positive for all higher θ . Finally, the third result implies that if two types have zero profits then all types in between have zero profits as well. That is, there cannot be a type $\theta \in [\theta_1, \theta_2]$ with positive profits (and negative profits are excluded by equation (6)).

Finally, we use the following notation. Let $q^h(\theta)$ denote the solution to¹⁰

$$S - c_q(q(\theta), \theta) + \frac{1 - F(\theta)}{f(\theta)} c_{q\theta}(q(\theta), \theta) = 0 \quad (13)$$

and $q^l(\theta)$ the solution to¹¹

$$S - c_q(q(\theta), \theta) - \frac{F(\theta)}{f(\theta)} c_{q\theta}(q(\theta), \theta) = 0 \quad (14)$$

¹⁰If several q solve this equation, we denote the highest by q^h . By assumption 1 and 2, there can be at most one $q > k(\theta)$ satisfying equation (13).

¹¹If the solution to this equation is not unique, let the lowest solution be q^l . By assumption 1 and 2, there is at most one $q < k(\theta)$ satisfying equation (14).

In the following two sections we solve the problem for the WM and then the WNM case. The strategy will be to solve the first order condition and then to verify ex post that (SOC) and non-local incentive constraints do not bind under our assumptions. Section 5 returns to the case where (SOC) is not satisfied.

3. First best welfare monotone

We will now characterize the optimal mechanism for the WM-case. The following lemma is useful to characterize the optimal menu. The lowest type $\underline{\theta}$ receives lowest profits (zero) and the IC constraint (5) is binding downwards. That is, high types would like to mimic low types (not the other way around).

Lemma 5 *In the WM-case we have: $\pi(\underline{\theta}) = 0$ and $\lambda(\theta) \geq 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.*

Now we are able to characterize the solution for the WM case. There are two cases to consider. In the first case, the solution (given by equation (13)) is such that the specialization of the firms plays no role. This is basically the solution to a standard problem. In the second case, low types up to a type θ_b are bunched on zero profits (but with different quality levels) and from $\theta_b \geq \underline{\theta}$ onwards, $q(\theta)$ follows the solution in equation (13).

Proposition 1 *There are two cases:*

1. *If $c_\theta(q^h(\underline{\theta}), \underline{\theta}) < 0$, then $q^h(\theta)$ in equation (13) gives the optimal quality for all $\theta \in [\underline{\theta}, \bar{\theta}]$. We have $\pi_\theta(\theta), q_\theta(\theta), x_\theta(\theta) > 0$ for each $\theta \in [\underline{\theta}, \bar{\theta}]$.*
2. *If $c_\theta(q^h(\underline{\theta}), \underline{\theta}) \geq 0$ then there exists a largest $\theta_b \geq \underline{\theta}$ such that*

$$q(\theta) = k(\theta) \text{ for all } \theta \in [\underline{\theta}, \theta_b]$$

and θ_b is determined by the unique solution to

$$S - c_q(k(\theta_b), \theta_b) + \frac{1 - F(\theta_b)}{f(\theta_b)} c_{q\theta}(k(\theta_b), \theta_b) = 0 \quad (15)$$

For all $\theta > \theta_b$ quality $q(\theta) = q^h(\theta)$. We have

$$\begin{aligned} \pi(\theta) &= 0 \text{ for all } \theta \in [\underline{\theta}, \theta_b], \\ \pi_\theta(\theta) &> 0 \text{ for all } \theta \in (\theta_b, \bar{\theta}], \text{ and} \\ x_\theta(\theta), q_\theta(\theta) &\geq 0 \text{ for all } \theta \in [\underline{\theta}, \bar{\theta}]. \end{aligned}$$

The relaxed solution is globally incentive compatible.

In the first case of proposition 1, the possibility that c_θ can change sign does not play a role in the relevant range of q . Therefore, the standard menu as in Che (1993) results. In the second case, c_θ would be positive for some types in the standard quality menu which is given by (13). A direct corollary of lemma 1 is that $c_\theta \leq 0$ at the first best quality level. Hence, the standard downward distortion of q caused by the rent extraction motive is responsible for having $c_\theta > 0$ for some types under q^h . By (5), profits are decreasing at types where $c_\theta > 0$. If q^h was implemented, type θ^b would therefore have zero profits while lower types would have positive profits. But now the principal can do better than q^h : By assigning $k(\theta)$ to types below θ^b , the principal (i) saves rents as those types remain at zero profits and (ii) reduces distortion compared to q^h . Because each type is most cost efficient at his $k(\theta)$, no other type can profitably misrepresent as θ if θ expects zero profits and produces quality $k(\theta)$. Put differently, the incentive constraint is lax in this situation. Therefore, it is not necessary to distort quality further down than $k(\theta)$ for rent extraction purposes. In some sense, specialization leads to “less distortion at the bottom” and more rent extraction.

In conclusion, the menu in case 2 of proposition 1 consists of a standard part for high types and one part where types produce at $k(\theta)$ and consequently the incentive constraint is lax.

4. First best welfare non-monotone

In this section, first best welfare is U-shaped. The lowest type $\underline{\theta}$ is no longer worst (in a first best sense) and therefore he might have positive profits under the optimal mechanism. The following lemma confirms this intuition.

Lemma 6 *Under WNM, $\pi(\underline{\theta}) > 0$, $\pi(\bar{\theta}) > 0$ and $\lambda(\underline{\theta}) = \lambda(\bar{\theta}) = 0$.*

One can think of the WNM case as having two standard menus. One for lower θ in which lower types are better, profits are decreasing in type and quality is distorted upwards. The other for higher θ with higher types being better, profits increasing in type and quality distorted downwards. These two menus have to be reconciled.

In the principal agent literature, irregularities are often dealt with bunching types on one decision¹². Hence, a first idea could be that bunching on quality might be used to connect the

¹²See, for instance, Guesnerie and Laffont (1984) or Fudenberg and Tirole (1991, ch. 7).

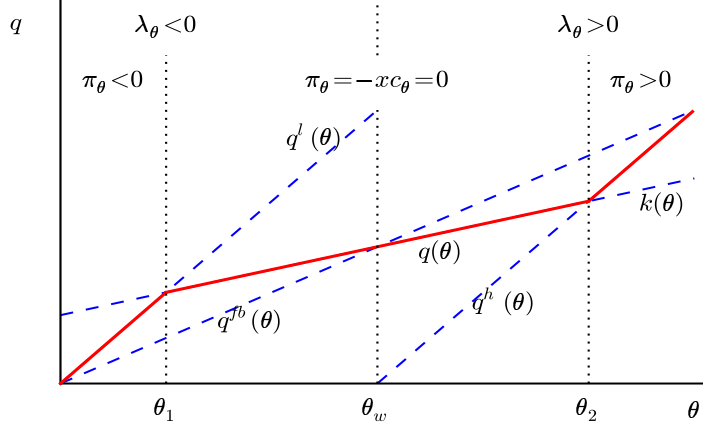


Figure 1: Optimal $q(\theta)$ (solid, red) in the WNM case, together with (dashed) $q^l(\theta)$, $q^{fb}(\theta)$, $k(\theta)$, $q^h(\theta)$.

two menus. It is quickly shown that this does not work. To see this, suppose –by contradiction– that $q(\theta) = q^b$ for types θ in the bunching interval. As profits are decreasing in θ for low θ and increasing in θ for high θ , the type θ' with the lowest profits ($\pi(\theta') = 0$) would have to be in the bunching interval. From (5), the profit minimizing type has to satisfy $c_\theta(q^b, \theta') = 0$. Hence, he produces at $q^b = k(\theta')$ and is for this quality level the most efficient type. But then he has the highest profits of all types in the quality-bunching interval. This contradiction implies that a menu with quality bunching cannot be the solution.¹³

The right way to reconcile the two standard menus is an interval of types with zero profits (but differing quality levels). Incentive compatibility within the bunched interval is no problem here. Each bunched type θ will produce at quality level $k(\theta)$ at which he has lower costs than any other type. The following proposition describes the optimal menu in the WNM case.

Proposition 2 *There exist unique θ_1 and θ_2 , with $\theta_1 < \theta_2$, such that $q^l(\theta_1) = k(\theta_1)$ and $q^h(\theta_2) = k(\theta_2)$. Quality is determined by*

$$q(\theta) = \begin{cases} q^h(\theta) & \text{for all } \theta > \theta_2 \\ k(\theta) & \text{for all } \theta \in [\theta_1, \theta_2] \\ q^l(\theta) & \text{for all } \theta < \theta_1. \end{cases} \quad (16)$$

¹³Unless it happens at $q^b = k$ in the case where $k(\theta) = k$ is constant (on a subset of $[\underline{\theta}, \bar{\theta}]$). In this case, those types that are bunched on zero profits would also have the same quality $q(\theta) = k$.

We have

$$\begin{aligned}\pi(\theta) &= 0 \text{ for all } \theta \in [\theta_1, \theta_2] \\ \pi_\theta(\theta) &< 0 \text{ for all } \theta < \theta_1 \\ \pi_\theta(\theta) &> 0 \text{ for all } \theta > \theta_2 \\ q_\theta(\theta) &\geq 0.\end{aligned}$$

Type θ_w , who has the lowest first best welfare of all types, is in the zero profit interval and produces his first best quality. It holds that

$$\begin{aligned}x_\theta(\theta) &< 0 \text{ for all } \theta < \theta_w \\ x_\theta(\theta) &> 0 \text{ for all } \theta > \theta_w.\end{aligned}$$

The relaxed decision is globally incentive compatible.

Figure 1 illustrates proposition 2. Quality is above first best, i.e. upwards distorted, for low θ and downwards distorted for high θ . This is a consequence of the U-shaped first best welfare which implies that low types are better around $\underline{\theta}$ and high types are better around $\bar{\theta}$. Quality is not distorted at the (locally) best types $\underline{\theta}$ and $\bar{\theta}$ which resembles the well known “no distortion at the top” result. Quality is also undistorted for the worst type θ_w which allows a continuous transition from upwards to downwards distortion.

The boundaries of the zero profit interval $[\theta_1, \theta_2]$ are at those types where the low standard menu and the high standard menu feature $q(\theta) = k(\theta)$. In the zero profit interval, each type produces the quality for which he is the cost minimizing type, i.e. $k(\theta)$. Any other quality could not be incentive compatible within a zero profit interval as either types slightly higher or slightly lower would be more efficient. But then they could achieve positive profits by misrepresenting. From $q(\theta) = k(\theta)$, it is evident that misrepresenting as any other type $\theta \in [\theta_1, \theta_2]$ cannot be profitable and this is exactly the reason why the zero profit types do not receive any informational rent.

At θ_1 and θ_2 , $q(\theta)$ is kinked. At θ_1 , for example, the quality according to the standard low menu (q^l) would include additional informational distortion pushing quality upwards. Therefore $q^l(\theta) > k(\theta)$ for types slightly above θ_1 while $q(\theta) = k(\theta)$ is necessary to stay in the zero profit interval.

Note that for types above θ_w the optimal contract is similar to the one derived in proposition 1, i.e. quality and virtual valuation are the same. This is quite intuitive as first best welfare is increasing for those types. In this sense, proposition 2 “extends” proposition 1.

The following proposition formalizes the “grudge” of high θ incumbents against low θ entrants: although in second best the incumbent generates higher quality and higher welfare than the entrant, it can happen that the entrant wins the procurement contract. Incidentally, the opposite can happen as well: an incumbent wins from an entrant who generates higher (second best) welfare.

Proposition 3 *The optimal allocation is not (second best) efficient in the sense that there exist types θ', θ'' such that θ' wins against θ'' while $W^{sb}(\theta'') > W^{sb}(\theta')$.*¹⁴

A similar result is well known in auctions with asymmetric bidders. Myerson (1981) shows that it is optimal to discriminate between bidders drawing their valuations from different distributions. For example, if bidder A draws his valuation from a distribution putting more weight on high values and bidder B draws from a distribution with low values, the auction will favor B. This decreases the rents A will get by stimulating him to bid more aggressively. In our case, there is only one distribution from which types are drawn. Nevertheless, the intuition is similar. The reason for discrimination are informational distortions. For the lower standard menu, the relevant term inducing distortion in the virtual valuation is $-F(\cdot)c_\theta(\cdot)$. For high θ , the respective term is $(1 - F(\cdot))c_\theta(\cdot)$. While discrimination in Myerson (1981) results from the fact that different distributions govern the distortion, discrimination in our model is due to different parts of the same distribution governing distortion: For low θ , the left tail is relevant and for high types the right tail of the distribution matters for distortion. The reason is that the local incentive constraint is upward binding in the lower standard menu and downward binding in the upper standard menu. On a more intuitive level, by ex ante committing to let a worse low type $\theta' < \theta^w$ win against a better high type $\theta'' > \theta^w$, one can save informational rents for θ'' and all types above him. The reason is that the probability that θ'' wins the auction, i.e. $x(\theta'')$, decreases and therefore the slope of the rent function $\pi_\theta(\theta'') = x(\theta'')c_\theta(q(\theta''), \theta'')$ decreases. Loosely speaking, one stimulates θ'' and higher types to bid more aggressively.

We conclude this section with a brief discussion of how to implement the optimal menus in propositions 1 and 2. We argue that this is more straightforward for the WM than for the WNM case. In each case, we have in mind that the government announces at the start its willingness p to pay (conditional on winning) for each quality level q . Hence, p corresponds to t/x in the mechanism design notation used so far. In case 1 in proposition 1, the government

¹⁴We use the term *second best efficient* to describe a situation where the selection rule picks the firm providing the highest W^{sb} . W^{sb} is welfare under the optimal quality schedule derived in propositions 1 and 2.

can then organize a second price auction to determine the firm that wins the contract. The firm with the highest bid, wins and pays the second highest bid. This firm is then allowed to choose its combination (q, p) from the menu announced by the government. Since the planner wants the highest type to win and profits are strictly increasing in θ , such an auction selects the right type as winner. Since we assume that the service is valuable enough that it has to be supplied, there is no reserve price in this auction.

However, in the second case in proposition 1 there are a number of types with equal (zero) profits while the planner prefers higher θ for the case where quality increases in θ . The auction described above is not optimal here since it cannot discriminate between types with the same profits. In that case, the selection mechanism must be based on quality directly. To be more precise, let firms bid qualities. The firm bidding the highest quality wins, produces this quality and receives payment p (according to the menu announced by the government).

In the WNM case, the auctions described above do not implement the optimal mechanism. The planner's preference over winning types is given by the virtual valuation VV in equation (11). Again, a price auction cannot work because of the zero profit interval: These types have the same valuation for winning the auction but should have different probabilities of winning ($VV(\theta)$ and thus $x(\theta)$ is not constant over $\theta \in [\theta_1, \theta_2]$). An auction based on quality does not work either because q_θ and x_θ do not have the same sign for all types. Furthermore, a scoring rule auction, as analyzed in Che (1993), cannot implement the optimal mechanism either. This can be seen as follows.

In a scoring rule auction, each bidder bids a score and the bidder with the highest score is contracted. Consider a scoring rule of the form $score(p, q) = s(q) - p$, where the price p is (only) paid to the winner of the auction. In a second score auction, the winner has to provide (q, p) which corresponds to the second highest score. Hence, the second highest score, $score^{(2)}$, determines the rents going to the winner. Consequently, it is a dominant strategy to bid the maximum score that one can deliver at non-negative profits. Thus,

$$bid(\theta) = \max_q \{s(q) - c(q, \theta)\}.$$

To implement the optimal mechanism, it must be the case that the first order condition

$$s_q(q) - c_q(q, \theta) = 0$$

yields $q(\theta)$ as given by equation (16). As shown by Che (1993), it then follows that

$$s(q) = Sq + \int_{q(\underline{\theta})}^q \frac{\lambda(q^{-1}(s))}{f(q^{-1}(s))} c_{q\theta}(s, q^{-1}(s)) ds \quad \text{for} \quad q \in [q(\underline{\theta}), q(\bar{\theta})]$$

and $-\infty$ for all other q ; where $q^{-1}(s)$ is the inverse of $q(\theta)$ and $\lambda(\theta)$ is the Lagrange multiplier (co-state) of the optimal menu derived in proposition 2.

This implies that the winner is determined by the firm bidding the highest value of

$$bid(\theta) = Sq - c(q, \theta) + \int_{q(\underline{\theta})}^q \frac{\lambda(q^{-1}(s))}{f(q^{-1}(s))} c_{q\theta}(s, q^{-1}(s)) ds$$

while in the optimal mechanism, the winner has the highest value of VV as given by equation (11).¹⁵ Put differently, if the scoring rule implements the optimal mechanism it has to hold that $bid(\theta') = bid(\theta'')$ whenever $VV(\theta') = VV(\theta'')$ under the optimal mechanism. The following proposition says that generically this is untrue under WNM.

Proposition 4 *Generically, a simple scoring rule auction cannot implement the optimal mechanism in the WNM case.*

Consequently, more general mechanisms are needed for implementation in the WNM case. As shown in proposition 3, the (optimal) government's decision may be criticized ex post in case a firm loses from a winner generating lower (second best) welfare. If the government cannot implement the optimal mechanism because of its complexity, more inefficiencies will be introduced in the WNM case.

5. Robustness

Above we made some assumptions on third derivatives of the cost function and the distribution of θ for ease of exposition. Here we discuss how the solution changes when these assumptions are no longer satisfied. In principle, there are two possible problems that can arise: First, the second order condition (SOC) could be violated in the derived solution. Second, the program is no longer globally concave.

5.1. Violation second order condition

For concreteness, we focus here on the WM case and assume that the problems arise because of a violation of the MHR assumption. The cases where third derivatives cause problems with

¹⁵Note that there is also an issue in choosing the right tie-breaking rule. From the envelope theorem, $\frac{d bid(\theta)}{d\theta} = -c_{\theta}(q(\theta), \theta)$. Therefore, all types with zero profits have the same bid because $q(\theta) = k(\theta)$ for them which implies $\frac{d bid(\theta)}{d\theta} = 0$. The tie breaking rule should follow $VV(\theta)$ here, making the mechanism not easy to implement.

(SOC) are dealt with analogously. In the WM case, the change in q for $\theta > \theta_b$ is given by

$$q_\theta(\theta) = \frac{c_{q\theta}(q(\theta), \theta) - c_{q\theta\theta}(q(\theta), \theta) \frac{1-F(\theta)}{f(\theta)} - c_{q\theta}(q(\theta), \theta) \frac{d(\frac{1-F(\theta)}{f(\theta)})}{d\theta}}{-c_{qq}(q(\theta), \theta) + c_{qq\theta}(q(\theta), \theta) \frac{1-F(\theta)}{f(\theta)}}. \quad (17)$$

The assumptions made above are sufficient conditions for $q_\theta(\theta) \geq 0$. Hence, if F does not satisfy the MHR assumption, it can still be the case that $q_\theta(\theta) \geq 0$ and $x_\theta(\theta) \geq 0$.¹⁶ If q and x are non-decreasing in θ , we know that the second order condition (SOC) is satisfied. Even if, say, $q_\theta(\theta) < 0$ while $x_\theta(\theta) > 0$, equation (SOC) can still be satisfied.

Now we consider the case where $d((1 - F(\theta))/f(\theta))/d\theta > 0$ for $\theta > \theta_b$ in such a way that $q_\theta < 0$ causes a violation of (SOC). We first sketch how this is dealt with in general. Then we work out an example. As shown by Guesnerie and Laffont (1984) and Fudenberg and Tirole (1991) for the case of a single dimensional decision (say, only quality), a violation of the second order condition leads to bunching: several θ -types produce the same quality. However, in our case the decision is two dimensional: quality q and the probability of winning x . In fact, below we do not work with x but with the virtual valuation VV as there is a one-to-one relation between the two (i.e. higher VV implies higher x and the other way around). We show that in this two-dimensional case, it is not necessarily true that a violation of (SOC) leads to bunching of types θ to the same quality q and probability of winning x .¹⁷

We use figure 2 to illustrate the procedure. This figure shows equation (SOC) (where it holds with equality) in (q, VV) space and the solution $(q(\theta), VV(\theta))$ that follows from the planner's optimization problem while ignoring the second order condition; i.e. assuming $\mu_\theta(\theta) = 0$ for all θ . The former curve is downward sloping in the WM case since

$$\frac{dx}{dq} = \frac{x_\theta(\theta)}{q_\theta(\theta)} = -x(\theta) \frac{c_{q\theta}(q(\theta), \theta)}{c_\theta(q(\theta), \theta)} < 0$$

In the simple case (that we also use in the example below) where $c_{\theta\theta} = 0$, this curve boils down to

$$x(\theta)c_\theta(q(\theta), \theta) = -K < 0 \quad (18)$$

for some constant $K > 0$, as differentiating equation (18) with respect to θ indeed gives $x_\theta c_\theta + x c_{q\theta} q_\theta = 0$.

¹⁶Whether $x_\theta \geq 0$ can be derived from the expression for $dVV(\theta)/d\theta$ in equation (37).

¹⁷A related point is already made by García (2005). He shows in a multidimensional screening model where single crossing holds in all dimensions that non-monotone decisions can be optimal (even if second order conditions do not bind).

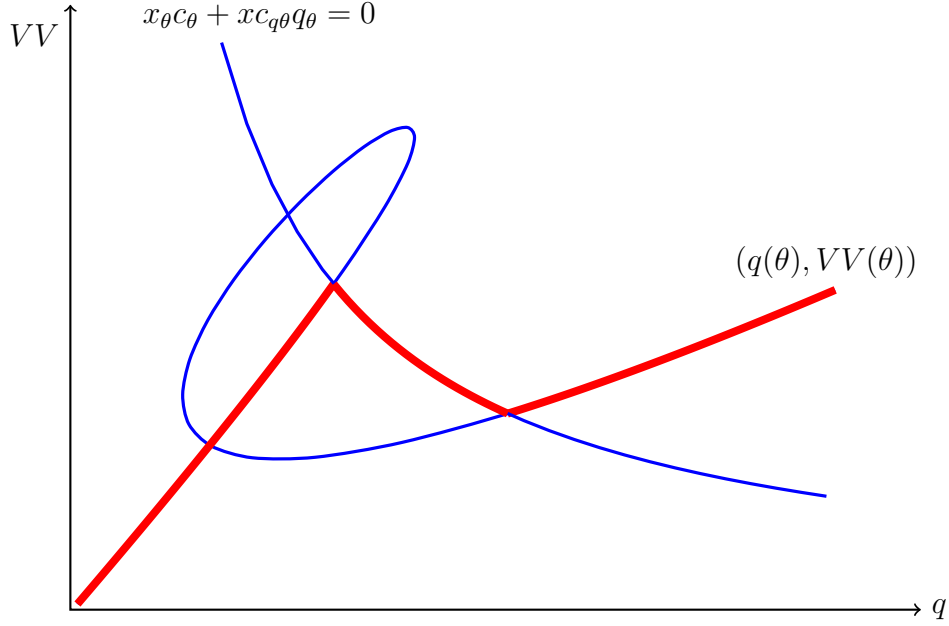


Figure 2: Solution for quality $q(\theta)$ and virtual valuation $VV(\theta)$ for the case where (second order) condition (SOC) is violated.

The solution $(q(\theta), VV(\theta))$ ignoring the second order constraint, starts at $\underline{\theta}$ in the bottom left corner and moves first over the thick (red) part of this curve, then follows the thin (blue) part, curving back (i.e. both q and x fall with θ) then both q and x increase again with θ and we end on the thick (red) part of the curve. The part of the curve where $q_\theta, x_\theta < 0$ violates equation (SOC).

Hence, we need to find θ_a, θ_b where (SOC) starts to bind and $\mu(\theta) > 0$. Then from θ_a onwards, we follow the binding constraint till we arrive at θ_b , from which point onwards we follow the solution $(q(\theta), VV(\theta))$ again. As shown in figure, the choice of θ_a determines both the trajectory $(\tilde{q}(\theta), \tilde{V}V(\theta))$ satisfying equation (SOC) and the end point of this trajectory θ_b . Since $\mu(\theta) = 0$ both for $\theta < \theta_a$ and for $\theta > \theta_b$, it must be the case that $\int_{\theta_a}^{\theta_b} \mu_\theta(\theta) d\theta = 0$. To illustrate, for the case where $c_{q\theta\theta} = 0$,¹⁸ this can be written as (using equation (10)):

$$\int_{\theta_a}^{\theta_b} \frac{f(\theta)(S - c_q(\tilde{q}(\theta), \theta)) + (1 - F(\theta))c_{q\theta}(\tilde{q}(\theta), \theta)}{c_{q\theta}(\tilde{q}(\theta), \theta)} d\theta = 0 \quad (19)$$

We now illustrate this approach with an example.

¹⁸If $c_{q\theta\theta} \neq 0$, the differential equation (10) has to be solved for $\mu(\theta)$. Although a bit tedious, this is do-able since the differential equation is linear and first order in $\mu(\theta)$.

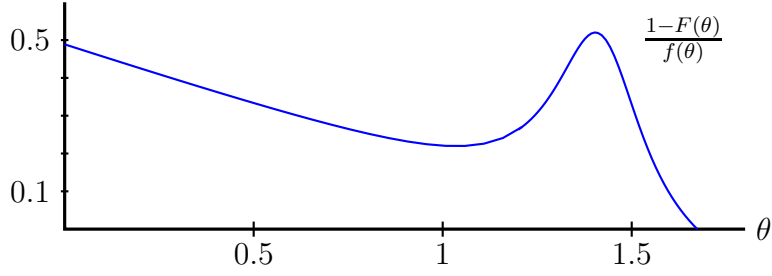


Figure 3: Inverse hazard rate with $f(\theta) = (\theta - a)^2 + 1/50$

Example 3 To violate the monotone hazard rate assumption we use the density $f(\theta) = (\theta - a)^2 + 1/50$ with support $[0, a + 1/4]$ where a has to be approximately 1.42 to satisfy the requirements of a probability distribution. The hazard rate of this distribution is depicted in figure 3.

Assume that there are two firms and that $c(q, \theta) = \frac{1}{2}q^2 - q\theta + \theta$. Then $c_\theta(q, \theta) = 1 - q$ which changes sign at $q = 1$. As $c_{\theta\theta} = 0$, the binding second order condition takes the form of (18):

$$x = \frac{K}{q - 1}$$

for some $K > 0$. Note that this equation does not depend on θ . Hence, in this case, “following the constraint” takes the form of bunching $\theta \in [\theta_a, \theta_b]$ on some point

$$(\tilde{q}, \tilde{V}\tilde{V}) \tag{20}$$

where $\tilde{V}\tilde{V}$ corresponds to the probability $\tilde{x} = \frac{K}{\tilde{q} - 1}$. Choosing θ_a , fixes $\tilde{q} = q(\theta_a)$ and θ_b since $q(\theta_b) = \tilde{q}$. Writing the dependency of \tilde{q}, θ_b on θ_a explicitly, θ_a solves equation (19):

$$\int_{\theta_a}^{\theta_b(\theta_a)} f(\theta)(S - (\tilde{q}(\theta_a) - \theta)) - (1 - F(\theta))d\theta = 0 \tag{21}$$

Since equation (SOC) will already start to bind for θ_a where $q_\theta(\theta_a) > 0$, it is routine to verify that this equation is downward sloping in θ_a . The unique solution in this example is $\theta_a \approx 1.1685$ which gives a corresponding $\theta_b = 1.428$ and $\tilde{q} = 1.923$.

While the ironing procedure described above takes care of the local second order condition (SOC), this does not necessarily imply global incentive compatibility. Global constraints are mathematically intractable in general frameworks; see Araujo and Moreira (2010) and Schottmüller (2011) for special examples of how to handle global constraints. However, the

following proposition establishes that global constraints do not bind for a family of cost functions. This family includes the functions we used in the example and the most commonly used linear-quadratic cost functions.

Proposition 5 *If $c_{\theta\theta} = 0$ and the local second order condition (SOC) is satisfied, the solution is globally incentive compatible.*

5.2. Concavity in q

The second possible problem with $c_{qq\theta}$ not satisfying assumption 2 is that the planner’s objective function (8) is not necessarily globally concave in $q(\cdot)$. However, in principle, the solution will still satisfy the first order conditions derived before. In particular, it is never optimal to choose $q \rightarrow \infty$: Since costs are convex and the principal’s utility is linear in q , costs are higher than benefits for q high enough and therefore optimal qualities cannot be arbitrarily high. Hence, if Q is not bounded, the solution will be interior and satisfy the conditions derived above.

If the set of available qualities is a compact subset of \mathbb{R}_+ , corner solutions could play a role; e.g. if quality cannot be higher than some level \bar{q} , some types might have $q(\theta) = \bar{q}$ and the first order conditions do not apply for them. However, such a situation can be easily approximated by a continuous cost function which is very steep around \bar{q} (instead of jumping discontinuously to infinity) and to which our analysis would apply.

6. Conclusion

We analyzed a procurement setting in which the procurement agency cares not only about the price but also about the quality of the product. In many post liberalization situations incumbents seem to be good at producing high quality while entrants can produce low quality at very low costs. A similar pattern emerges if there are gains from specialization and firms can specialize in either high quality or low costs.

Standard procurement models do not account for this possibility because single crossing is assumed in all dimensions. More precisely, it is assumed that “type” denotes efficiency and not specialization. This implies that a more efficient type is simply better for all quality levels. We relax this assumption and allow each type to be specialized, i.e. to be the most efficient type for some quality level. This leads to a bunching of types on zero profits. The intuition is that distorting quality further than the quality level a type is specialized in (for rent extraction

reasons) is not necessary: A type producing “his quality level” with expected profits of zero cannot be mimicked by any other type. Hence, the incentive constraint is lax and an interval of zero profit types is feasible. In short, specialization limits distortion and helps the principal to extract rents.

If we assume that first best welfare is U-shaped, e.g. there are gains from specializing in low costs even from a welfare point of view, we get an interesting discrimination result. Types with lower second best welfare can be preferred to types with higher second best welfare. This is similar to auctions with asymmetric bidders where discriminatory mechanisms are well known. Contrary to this literature, bidders are drawn from the same distribution in our model. The intuition is that the incentive constraint is first upward and then, for higher types, downward binding. Therefore, different parts of the distribution govern the distortion for low and high types. By committing to let some worse types win against some better types, the principal can reduce the rents of the best types. Loosely speaking, the better types are incentivized to bid more aggressively. Put differently, competitive pressure can be exerted even by firms that are clearly worse. Further, in this case “gold plating” can be optimal in the sense that some types produce quality levels above their first best levels.

References

- ARAÚJO, A., AND H. MOREIRA (2010): “Adverse Selection Problems without the Spence-Mirrlees Condition,” *Journal of Economic Theory*, 145(5), 1113–1141.
- ASKER, J., AND E. CANTILLON (2008): “Properties of Scoring Auctions,” *RAND Journal of Economics*, 39(1), 69–85.
- (2010): “Procurement when Price and Quality Matter,” *RAND Journal of Economics*, 41(1), 1–34.
- BAGNOLI, M., AND T. BERGSTROM (2005): “Log-Concave Probability and its Applications,” *Economic Theory*, 26(2), 445–469.
- BRANCO, F. (1997): “The Design of Multidimensional Auctions,” *RAND Journal of Economics*, 28(1), 63–81.
- BROCAS, I. (2011): “Countervailing incentives in allocation mechanisms with type-dependent externalities,” Discussion Paper.
- CARRILLO, J. (1998): “Coordination and externalities,” *Journal of Economic Theory*, 78(1), 103–129.
- CHE, Y.-K. (1993): “Design Competition through Multidimensional Auctions,” *RAND Journal of Economics*, 24(4), 668–680.
- FIGUEROA, N., AND V. SKRETA (2009): “The role of optimal threats in auction design,” *Journal of Economic Theory*, 144(2), 884–897.
- FUDENBERG, D., AND J. TIROLE (1991): *Game theory*. MIT Press.
- GARCÍA, D. (2005): “Monotonicity in direct revelation mechanisms,” *Economics Letters*, 88(1), 21–26.
- GUESNERIE, R., AND J.-J. LAFFONT (1984): “A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm,” *Journal of Public Economics*, 25(3), 329 – 369.
- JULLIEN, B. (2000): “Participation Constraints in Adverse Selection Models,” *Journal of Economic Theory*, 93(1), 1–47.
- LAFFONT, J.-J., AND J. TIROLE (1987): “Auctioning Incentive Contracts,” *Journal of Political Economy*, 95(5), 921–937.

- (1993): *A theory of incentives in procurement and regulation*. MIT Press.
- LEWIS, T., AND D. SAPPINGTON (1989): “Countervailing incentives in agency problems,” *Journal of Economic Theory*, 49(2), 294–313.
- MATTHEWS, S., AND J. MOORE (1987): “Monopoly provision of quality and warranties: An exploration in the theory of multidimensional screening,” *Econometrica*, 55(2), 441–467.
- MCAFEE, R., AND J. MCMILLAN (1988): “Multidimensional incentive compatibility and mechanism design,” *Journal of Economic Theory*, 46(2), 335–354.
- MYERSON, R. (1981): “Optimal Auction Design,” *Mathematics of Operations Research*, 6, 58–73.
- ROCHET, J., AND L. STOLE (2003): “The economics of multidimensional screening,” in *Advances in economics and econometrics: Theory and applications, Eighth World Congress*, vol. 1, pp. 150–197. Cambridge University Press.
- SCHOTTMÜLLER, C. (2011): “Adverse selection without single crossing: The monotone solution,” *Center Discussion Paper, no. 2011–123*.

8. Appendix: Proofs

Proof of lemma 1 From the first order condition for q^{fb} we derive that

$$q_{\theta}^{fb} = \frac{-c_{q\theta}}{c_{qq}} > 0$$

Then we find that

$$W_{\theta\theta}^{fb}(\theta) = \frac{c_{q\theta}^2}{c_{qq}} - c_{\theta\theta} > 0$$

from the assumptions made on the function $c(q, \theta)$. Further, it follows from $c_{\theta}(k(\theta), \theta) \equiv 0$ that $c_{q\theta}k_{\theta}(\theta) + c_{\theta\theta} = 0$. Hence, $q_{\theta}^{fb} > k_{\theta}$ if and only if

$$\frac{-c_{q\theta}}{c_{qq}} > \frac{c_{\theta\theta}}{-c_{q\theta}}$$

which holds by assumption 3. *Q.E.D.*

Proof of lemma 2 Define the function

$$\Phi(\hat{\theta}, \theta) = \pi(\theta, \theta) - \pi(\hat{\theta}, \theta) \geq 0$$

By IC this function is always positive and equal to zero if $\hat{\theta} = \theta$. In other words, the function Φ reaches a minimum at $\hat{\theta} = \theta$. Thus truth-telling implies both

$$\left. \frac{\partial \Phi(\hat{\theta}, \theta)}{\partial \hat{\theta}} \right|_{\hat{\theta}=\theta} = 0 \tag{22}$$

and

$$\left. \frac{\partial^2 \Phi(\hat{\theta}, \theta)}{\partial \hat{\theta}^2} \right|_{\hat{\theta}=\theta} \geq 0 \tag{23}$$

Since equation (22) has to hold for all $\hat{\theta} = \theta$, differentiating with respect to θ gives

$$\left. \frac{\partial^2 \Phi(\hat{\theta}, \theta)}{\partial \hat{\theta}^2} \right|_{\hat{\theta}=\theta} + \left. \frac{\partial^2 \Phi(\hat{\theta}, \theta)}{\partial \hat{\theta} \partial \theta} \right|_{\hat{\theta}=\theta} = 0$$

Then equation (23) implies that

$$\left. \frac{\partial^2 \Phi(\hat{\theta}, \theta)}{\partial \hat{\theta} \partial \theta} \right|_{\hat{\theta}=\theta} \leq 0$$

It follows from the definition of Φ that

$$\left. \frac{\partial^2 \Phi(\hat{\theta}, \theta)}{\partial \hat{\theta} \partial \theta} \right|_{\hat{\theta}=\theta} = x_{\theta}(\theta)c_{\theta}(q(\theta), \theta) + x(\theta)c_{q\theta}(q(\theta), \theta)q_{\theta}(\theta) \leq 0$$

which is the inequality in the lemma.

Q.E.D.

Proof of lemma 3 We need to show that

$$\frac{d(\lambda(\theta)/f(\theta))}{d\theta} = \frac{\lambda_\theta(\theta)f(\theta) - \lambda(\theta)f_\theta(\theta)}{f(\theta)^2} < 0 \quad (24)$$

We consider the following four cases:

		$\lambda(\theta)$	
		≥ 0	< 0
$f_\theta(\theta)$	≥ 0	(α)	(β)
	< 0	(δ)	(γ)

Let's consider the two cases in the lemma in turn.

Case (i): We can solve

$$\lambda(\theta) = 1 - F(\theta) - \int_\theta^{\bar{\theta}} \eta(t)dt \quad (25)$$

Hence, we need to show

$$-f(\theta)^2 - \lambda(\theta)f_\theta(\theta) < 0 \quad (26)$$

where we use $\eta(\theta) = 0$. This is obviously satisfied in case (α) . In case (δ) we have

$$-f(\theta)^2 - (1 - F(\theta) - \int_\theta^{\bar{\theta}} \eta(t)dt)f_\theta(\theta) < 0$$

Then this inequality is implied by the MHR assumption 1 where we write $d((1-F(\theta))/f(\theta))/d\theta < 0$ as

$$-f(\theta)^2 - (1 - F(\theta))f_\theta(\theta) < 0 \quad (27)$$

because $\eta(t) \geq 0$. As we assume $\lambda(\theta) \geq 0$, we do not need to consider cases (β, γ) .

Case (ii): Here we have a second way in which we can write $\lambda(\theta)$:

$$\lambda(\theta) = -F(\theta) + \int_{\underline{\theta}}^\theta \eta(t)dt \quad (28)$$

Equation (26) is clearly satisfied in cases $(\alpha), (\gamma)$. Case (δ) is satisfied for the same reason as above. Hence, we only need to consider case (β) . Using equation (28), we write inequality (26) as

$$-f(\theta)^2 - (-F(\theta) + \int_{\underline{\theta}}^\theta \eta(t)dt)f_\theta(\theta) < 0$$

where we use $\eta(\theta) = 0$. Then this inequality is implied by the MHR assumption 1 where we write $d(F(\theta)/f(\theta))/d\theta > 0$ as

$$-f(\theta)^2 + F(\theta)f_\theta(\theta) < 0 \quad (29)$$

and $\eta(t) \geq 0$.

Q.E.D.

Proof of lemma 4 We prove the parts in turn.

Part 1.: Suppose not, that is assume that $q(\hat{\theta}) = k(\hat{\theta})$ (i.e. $c_\theta(q(\hat{\theta}), \hat{\theta}) = 0$) and $q_\theta(\hat{\theta}) < k_\theta(\hat{\theta})$. Then for $\varepsilon > 0$ small enough, it is the case that

$$c_\theta(q(\hat{\theta} + \varepsilon), \hat{\theta} + \varepsilon) > 0$$

and thus (by (5))

$$\pi_\theta(\hat{\theta} + \varepsilon) < 0$$

This is only feasible if $\pi(\hat{\theta}) > 0$ and thus $\eta(\hat{\theta}) = 0$. With $\mu(\theta) = 0$ the first order condition (10) becomes

$$S - c_q(q(\theta), \theta) + \frac{\lambda(\theta)}{f(\theta)} c_{q\theta}(q(\theta), \theta) = 0 \quad (30)$$

Using the implicit function theorem we find

$$q_\theta = \frac{c_{q\theta}(-1 + (\lambda/f)') + c_{q\theta\theta}\lambda/f}{c_{qq} - c_{qq\theta}\lambda/f} \quad (31)$$

As derived in the proof of lemma 1, $k_\theta = \frac{c_{\theta\theta}}{-c_{q\theta}}$. Comparing q_θ and k_θ in the point $\hat{\theta}$ we can simplify the expression in (31) by noting that $c_{q\theta\theta} = c_{qq\theta} = 0$ for $\theta = \hat{\theta}$ by assumption 2. Using this we can write $q_\theta(\hat{\theta}) < k_\theta(\hat{\theta})$ as

$$c_{qq}c_{\theta\theta} - c_{q\theta}^2 > c_{q\theta\theta}^2(-\lambda/f)' \quad (32)$$

which leads to a contradiction because the left hand side is negative by assumption 3 and the right hand side is positive by assumption. Hence, it must be the case that $q_\theta(\hat{\theta}) \geq k_\theta(\hat{\theta})$ at such a point $\hat{\theta}$.

Part 2. Suppose not, that is there exists $\theta'' > \theta'$ such that $c_\theta(q(\theta''), \theta'') > 0$, i.e. such that $q(\theta'') < k(\theta'')$. This implies that there exists $\hat{\theta} \in [\theta', \theta'')$ such that $q(\hat{\theta}) = k(\hat{\theta})$ and $q_\theta(\hat{\theta}) < k_\theta(\hat{\theta})$. Part 1 of this lemma shows that this is not possible.

Part 3. The proof is again by contradiction. Suppose, profits were positive on some interval $(\hat{\theta}_1, \hat{\theta}_2)$ with $\theta_1 < \hat{\theta}_1 < \hat{\theta}_2 < \theta_2$.¹⁹ Quality $q(\theta)$ for $\theta \in (\hat{\theta}_1, \hat{\theta}_2)$ will be determined by (30) with $\lambda(\theta) = 1 - F(\theta) - \int_{\hat{\theta}_2}^{\hat{\theta}} \eta(\theta) d\theta$. Clearly, there has to be a type $\hat{\theta} \in (\hat{\theta}_1, \hat{\theta}_2)$ at which $\pi(\theta)$ attains a local maximum. Since profits are increasing for $\hat{\theta} - \varepsilon$ and decreasing for $\hat{\theta} + \varepsilon$, (5) implies that $q(\hat{\theta} - \varepsilon) > k(\hat{\theta} - \varepsilon)$ and $q(\hat{\theta} + \varepsilon) < k(\hat{\theta} + \varepsilon)$. Hence, $q(\hat{\theta}) = k(\hat{\theta})$ and

$$q_\theta(\hat{\theta}) < k_\theta(\hat{\theta})$$

¹⁹By continuity of $\pi(\theta)$, it cannot be the case that $\pi(\theta) > 0$ only in a point.

which is impossible by part 1 of this lemma. This is the required contradiction. *Q.E.D.*

Proof of lemma 5 In order to proof this, we need the following result.

Lemma 7 *At all qualities greater or equal to his first best quality $q^{fb}(\theta)$, the costs of type θ are lower than the costs of all $\tilde{\theta} < \theta$, i.e. $c(q, \theta) < c(q, \tilde{\theta}) \quad \forall q \geq q^{fb}(\theta)$ and $\tilde{\theta} < \theta$.*

Proof of lemma 7 At $q^{fb}(\theta)$ the claim follows from the strictly increasing first best net value assumption (WM): If a lower θ had the same or lower costs at $q^{fb}(\theta)$, he could produce at least the same net value by producing at $q^{fb}(\theta)$. Given that θ has lower costs for $q^{fb}(\theta)$, it is sufficient to show that the incremental costs of producing higher q , i.e. $c(q, \cdot) - c(q^{fb}(\theta), \cdot)$, is lower for θ than for any $\tilde{\theta} < \theta$. Since,

$$c(q, \tilde{\theta}) - c(q^{fb}(\theta), \tilde{\theta}) = \int_{q^{fb}(\theta)}^q c_q(x, \tilde{\theta}) dx \quad (33)$$

is strictly decreasing in $\tilde{\theta}$ because of $c_{q\theta} < 0$, the claim follows. Note that an implication of this claim is that

$$c_\theta(q, \theta) < 0 \text{ for all } q \geq q^{fb}(\theta) \quad (34)$$

as otherwise a marginally lower type would have lower costs. *Q.E.D.*

The proof of lemma 5 is by contradiction. Suppose there exists θ' such that $\lambda(\theta') < 0$.²⁰ Since the transversality condition implies $\lambda(\bar{\theta}) = 0$,²¹ it follows from the continuity of $\lambda(\theta)$ that there must be an interval of types in between θ' and $\bar{\theta}$ where $\lambda_\theta(\theta) = -f(\theta) + \eta(\theta) > 0$. This can only happen if $\eta(\theta) > 0$ or equivalently $\pi(\theta) = 0$ on this interval. On such a zero profit interval $c_\theta(q, \theta) = 0$ as $\pi_\theta(\theta)$ would not be zero otherwise.²² Hence, each type produces a quality such that he is the cost minimizing type for this quality. Furthermore, each of these types has $\pi(\theta) = 0$. These two facts imply that incentive compatibility cannot be a problem on the zero profit interval. In other words, we can ignore constraint (SOC) on this interval, i.e. $\mu(\theta) = 0$ on this interval.

Denote the lowest type with zero profits as $\theta_1 = \inf\{\theta | \pi(\theta) = 0, \theta \geq \theta_1\}$. Note that from what was said above $\lambda(\theta_1) < 0$ and $\mu(\theta_1) = 0$. Furthermore, $c_\theta(q(\theta_1 - \varepsilon), \theta_1 - \varepsilon) \geq 0$ for $\varepsilon > 0$ small enough.²³ Equation (34) then implies $q(\theta_1 - \varepsilon) < q^{fb}(\theta_1 - \varepsilon)$. However, this contradicts

²⁰As λ is continuous, it is without loss of generality to assume $\theta' > \underline{\theta}$.

²¹Because the highest type has strictly positive profits (see below).

²²The alternative would be $x(\theta) = 0$. But this is obviously not possible on an interval of types by assumption 1.

²³This follows from the definition of θ_1 : Since $\pi(\theta_1) = 0$ and $\pi(\theta_1 - \varepsilon) \geq 0$, profits have to be decreasing at $\theta_1 - \varepsilon$ for ε small enough.

the first order condition with respect to q :

$$f(\theta)(S_q(q(\theta)) - c_q(q(\theta), \theta)) + \lambda(\theta)c_{q\theta}(q(\theta), \theta) + \mu(\theta)c_{q\theta\theta}(q(\theta), \theta) = -\mu_\theta(\theta)c_{q\theta}(q(\theta), \theta) \quad (35)$$

Since $\mu(\theta_1) = 0$, we can ignore the $\mu(\theta_1 - \varepsilon)$ term for $\varepsilon > 0$ small enough. Further, $\mu_\theta(\theta_1 - \varepsilon) \leq 0$ since $\mu(\theta) \geq 0$. But then $c_{\theta q} < 0$ and $\lambda(\theta_1 - \varepsilon) < 0$ imply $S_q(q(\theta_1 - \varepsilon)) - c_q(q(\theta_1 - \varepsilon), \theta_1 - \varepsilon) < 0$ which contradicts $q(\theta_1 - \varepsilon) < q^{fb}(\theta_1 - \varepsilon)$. Hence, there is a contradiction and $\lambda(\theta) \geq 0$ has to hold.

To prove the other part of the lemma, suppose (again by contradiction) that $\pi(\underline{\theta}) > 0$. Consequently, the dynamic optimization problem will include the transversality condition $\lambda(\underline{\theta}) = 0$. Given that $\lambda_\theta(\theta) = -f(\theta) + \eta(\theta)$, this implies that $\lambda(\theta) < 0$ for some interval of θ starting at $\underline{\theta}$.²⁴ As we just proved, it is not possible to have $\lambda(\theta) < 0$. This is the required contradiction and we conclude that $\pi(\underline{\theta}) = 0$. *Q.E.D.*

Proof of proposition 1

We will analyze the problem without the incentive constraint (SOC) first, i.e. $\mu(\theta) = 0$, and show afterwards in lemma 9 that it is satisfied. The first order condition (10) becomes

$$S - c_q(q(\theta), \theta) + \frac{\lambda(\theta)}{f(\theta)}c_{q\theta}(q(\theta), \theta) = 0 \quad (36)$$

Since $\bar{\theta}$ is the best type (in a first best sense), we expect his profits to be positive and therefore $\eta(\bar{\theta}) = 0$ and also the transversality condition $\lambda(\bar{\theta}) = 0$ holds (indeed below we verify that $\pi(\bar{\theta}) > 0$). Therefore (9) implies $\lambda(\theta) = 1 - F(\theta)$ for some high types for which the profit constraint does not bind. Note that for this case, equation (36) can be written as (13).

Now we have two cases. With the solution $q^h(\theta)$ given by (13) it is the case that either

1. $c_\theta(q^h(\underline{\theta}), \underline{\theta}) < 0$ or
2. $c_\theta(q^h(\underline{\theta}), \underline{\theta}) \geq 0$

The first case implies that $q(\underline{\theta}) > k(\underline{\theta})$. Hence, for the first case, $\pi(\underline{\theta}) = 0$ and $\pi_\theta(\theta) > 0$ for (at least) θ close to $\underline{\theta}$ (see equation (5)). It follows from part 2 of lemma 4 that $c_\theta(q(\theta), \theta) \leq 0$ for all θ . Thus $\pi_\theta \geq 0$ for each $\theta > \underline{\theta}$ and the profit constraint $\pi(\theta) \geq 0$ does not bind for $\theta > \underline{\theta}$. Therefore the solution in equation (13) is the overall solution.

²⁴To be precise, this follows from the continuity of $\pi(\theta)$: As $\pi(\underline{\theta}) > 0$, profits have to be positive for some interval of low θ and consequently $\eta(\theta) = 0$ for those θ . This implies $\lambda(\theta) < 0$.

Finally, consider the virtual surplus in equation (11) with $\lambda(\theta) = 1 - F(\theta)$. Using an envelope argument, it is routine to derive that

$$\frac{dVV(\theta)}{d\theta} = -c_\theta \left(1 - \left(\frac{1-F}{f}\right)'\right) + \frac{1-F}{f} c_{\theta\theta} > 0 \quad (37)$$

Since the project is allocated to the firm with the highest VV , it is allocated to the firm with the highest θ . Thus $x_\theta(\theta) > 0$.

Now consider the second case in proposition 1 with $c_\theta(q^h(\underline{\theta}), \underline{\theta}) \geq 0$.

Lemma 8 *If $c_\theta(q^h(\underline{\theta}), \underline{\theta}) \geq 0$, then $q(\underline{\theta}) = k(\underline{\theta})$.*

Proof of lemma 8 If $c_\theta(q^h(\underline{\theta}), \underline{\theta}) = 0$, the lemma is true.

If $c_\theta(q(\underline{\theta}), \underline{\theta}) > 0$ then $\pi_\theta(\underline{\theta}) < 0$. Then $\pi(\underline{\theta}) = 0$ (lemma 5) implies that this violates the constraint that profits should be non-negative. In this case the solution cannot be given by equation (13) as the profit constraint is binding. Hence, the solution $q(\theta)$ is given by equation (30) where (see equation (9)) $\lambda(\theta)$ is given by equation (28). This solution cannot feature $c_\theta(q(\underline{\theta}), \underline{\theta}) > 0$ as this would lead to a violation of $\pi(\theta) \geq 0$ for θ close to $\underline{\theta}$.

The following argument shows that $c_\theta(q(\underline{\theta}), \underline{\theta}) < 0$ is not possible either. In this case $\pi_\theta(\underline{\theta}) > 0$. Then either (i) there exists $\theta' > \underline{\theta}$ such $\pi(\theta') = 0$ or (ii) $q(\theta) = q^h(\theta)$. Case (i) leads to a contradiction because of lemma 4. If (i) does not happen, then $\eta(\theta) = 0$ for all $\theta > \underline{\theta}$, which implies case (ii). However, case (ii) with $c_\theta(q(\underline{\theta}), \underline{\theta}) < 0$ contradicts the assumption in the lemma that $c_\theta(q^h(\underline{\theta}), \underline{\theta}) \geq 0$.

Thus we have $c_\theta(q(\underline{\theta}), \underline{\theta}) = 0$ or equivalently $q(\underline{\theta}) = k(\underline{\theta})$. *Q.E.D.*

Because of lemma 8, there is a largest $\theta_b \geq \underline{\theta}$ such that $q(\theta) = k(\theta)$ for all $\theta \in [\underline{\theta}, \theta_b]$. This θ_b is uniquely defined.

Since $\pi(\underline{\theta}) = 0$ and $\pi_\theta = 0$ for all $\theta \in [\underline{\theta}, \theta_b]$, we have $\pi(\theta) = 0$ for all $\theta \in [\underline{\theta}, \theta_b]$.

Uniqueness of θ_b as defined in (15) follows from the fact that the expression in equation (15) is strictly increasing in θ_b . Differentiating the expression with respect to θ_b and using assumption 2 we find:

$$\frac{1}{-c_{q\theta}} (c_{q\theta}^2 - c_{qq}c_{\theta\theta} - c_{q\theta}^2 \left(\frac{1-F}{f}\right)') > 0$$

We can only leave the interval $[\underline{\theta}, \theta_b]$ if $\pi_\theta(\theta_b + \varepsilon) > 0$ for $\varepsilon > 0$ small enough. Then $\pi(\theta) > 0$ for all $\theta > \theta_b$. If not, there would be $\theta' > \theta_b$ such that $\pi(\theta') = 0$ which contradicts lemma 4. Hence, $q(\theta) = q^h(\theta)$ for all $\theta > \theta_b$ and equation (15) makes sure that $q(\theta)$ is continuous.

As in the previous case, we have $q_\theta(\theta) > 0$ for $\theta > \theta_b$. For $\theta \in [\underline{\theta}, \theta_b]$ we have $q(\theta) = k(\theta)$ which is (strictly) increasing in θ if $c_{\theta\theta} > 0$. If $c_{\theta\theta} = 0$, quality is constant over the range $\theta \in [\underline{\theta}, \theta_b]$.

Finally, we show that $x_\theta(\theta) \geq 0$:

$$\frac{dVV(\theta)}{d\theta} = -c_\theta \left(1 - \left(\frac{\lambda(\theta)}{f}\right)'\right) + \frac{\lambda(\theta)}{f} c_{\theta\theta} \geq 0 \quad (38)$$

where the inequality is strict for $\theta > \theta_b$ and for $\theta \in [\underline{\theta}, \theta_b]$ if $c_{\theta\theta} > 0$.

Finally, lemma 9 establishes global incentive compatibility. *Q.E.D.*

Lemma 9 *The relaxed solution in proposition 1 is globally incentive compatible.*

Proof of lemma 9 The monotonicity of $x(\theta)$ and $q(\theta)$ together with $c_\theta \leq 0$ and $c_{q\theta} < 0$ imply that the local incentive compatibility constraint (SOC) is satisfied.

For global incentive compatibility we first show that no θ can profitably misrepresent as $\hat{\theta} > \theta$. This is true if

$$\pi(\theta) - \pi(\hat{\theta}) - x(\hat{\theta})[c(q(\hat{\theta}), \hat{\theta}) - c(q(\hat{\theta}), \theta)] \geq 0$$

Using (5), this can be rewritten as

$$\int_{\theta}^{\hat{\theta}} x(t)c_\theta(q(t), t) - x(\hat{\theta})c_\theta(q(\hat{\theta}), t) dt \geq 0$$

This last inequality can be rewritten as

$$\int_{\theta}^{\hat{\theta}} \int_t^{\hat{\theta}} x_\theta(s)c_\theta(q(s), t) + x(s)c_{q\theta}(q(s), t)q_\theta(s) ds dt \leq 0 \quad (39)$$

The second term of the integrand is negative by the monotonicity of $q(\theta)$ in proposition 1. Note that we saw in the proof of proposition 1 that $c_\theta(q(\theta), \theta) \leq 0$ for all types. Since $t \leq s$ and $c_{\theta\theta} \geq 0$, clearly $c_t(q(s), t) \leq 0$ in the first term of the integrand. As $x_\theta \geq 0$ in proposition 1, inequality (39) has to hold.

To show that no θ gains by misrepresenting as $\hat{\theta} < \theta$ we use the following notation introduced in equation (3).

$$\pi(\hat{\theta}, \theta) = t(\hat{\theta}) - x(\hat{\theta})c(q(\hat{\theta}), \theta)$$

The idea is to define the following cost function

$$\tilde{c}(a, \theta) = \min\{c(q(a), a), c(q(a), \theta)\} \quad (40)$$

where $q(a)$ is the optimal quality schedule derived above. Next define

$$\tilde{\pi}(a, \theta) = t(a) - x(a)\tilde{c}(a, \theta) \quad (41)$$

The following inequalities show that the solution derived above satisfies IC globally as well:

$$\begin{aligned} & \pi(\hat{\theta}, \theta) - \pi(\theta, \theta) \\ \leq & \tilde{\pi}(\hat{\theta}, \theta) - \tilde{\pi}(\theta, \theta) \\ = & \int_{\theta}^{\hat{\theta}} \frac{\partial \tilde{\pi}(a, \theta)}{\partial a} da \\ = & \int_{\hat{\theta}}^{\theta} \left(\left. \frac{\partial \pi(a, \theta)}{\partial a} \right|_{\theta=a} - \frac{\partial \tilde{\pi}(a, \theta)}{\partial a} \right) da \end{aligned} \quad (42)$$

$$\begin{aligned} = & \int_{\hat{\theta}}^{\theta} x_{\theta}(a)(\tilde{c}(q(a), \theta) - c(q(a), a)) + x(a)(\tilde{c}_a(q(a), \theta) - c_q(q(a), a)q_{\theta}(a)) da \\ \leq & 0 \end{aligned} \quad (43)$$

where the first inequality follows from the definition of $\tilde{c}(\cdot)$ and the observation that $\tilde{\pi}(\theta, \theta) = \pi(\theta, \theta)$. Equation (42) follows because $\left. \frac{\partial \pi(a, \theta)}{\partial a} \right|_{\theta=a} = 0$ by the first order condition of truthful revelation. Equation (43) follows from the definitions of the derivatives of $\pi(a, \theta)$ and $\tilde{\pi}(a, \theta)$ w.r.t. a . The final inequality follows from the properties $x_a(a), q_a(a) \geq 0$ and the following three observations. First, by definition of $\tilde{c}(\cdot)$ we have

$$\tilde{c}(q(a), \theta) - c(q(a), a) \leq 0$$

Second, for values of a where $\tilde{c}(a, \theta) = c(q(a), \theta)$ we have

$$\tilde{c}_a(q(a), \theta) - c_q(q(a), a)q_a(a) = (c_q(q(a), \theta) - c_q(q(a), a))q_a(a) \leq 0$$

because $c_{q\theta} \leq 0$. Finally for values where $\tilde{c}(a, \theta) = c(q(a), a)$ we have

$$\tilde{c}_a(q(a), \theta) - c_q(q(a), a)q_a(a) = \left. \frac{\partial c(q(a), \theta)}{\partial \theta} \right|_{\theta=a} \leq 0$$

because in our solution $c_{\theta}(q(\theta), \theta) \leq 0$ for all θ .

Q.E.D.

Proof of lemma 6 We show that a menu featuring $\pi(\underline{\theta}) = 0$ is not optimal. In the WNM case, first best welfare is decreasing in type around $\underline{\theta}$. A standard envelope argument shows that this implies $c_{\theta}(q^{fb}(\underline{\theta}), \underline{\theta}) > 0$. Now suppose, $\pi(\underline{\theta}) = 0$. Then $\pi_{\theta}(\underline{\theta}) \geq 0$ which implies $c_{\theta}(q(\underline{\theta}), \underline{\theta}) \leq 0$ by (5). Therefore, $q(\underline{\theta}) > q^{fb}(\underline{\theta})$. But then a simple change in the menu would be beneficial and therefore the menu cannot be optimal: Change $q(\underline{\theta})$ to $q^{fb}(\underline{\theta})$ and adjust transfers such that $\pi(\underline{\theta})$ stays zero. As $\underline{\theta}$ has again zero profits his incentive compatibility does not change. Reducing quality will make $\underline{\theta}$'s menu point even less attractive for other types.²⁵

²⁵This argument can be made formal using the same dc expression as in the proof of lemma 7.

By the definition of $q^{fb}(\cdot)$, this change is beneficial.

Q.E.D.

Proof of proposition 2 Again the global IC constraint will be neglected first and checked ex post.

From lemma 6 we know that $\pi(\bar{\theta}), \pi(\underline{\theta}) > 0$ and therefore the transversality conditions $\lambda(\bar{\theta}) = \lambda(\underline{\theta}) = 0$ have to hold. Furthermore, the positive profit constraint is non binding and therefore $\eta(\bar{\theta}) = \eta(\underline{\theta}) = 0$. By (9) and the continuity of π , we have $\lambda(\theta) = 1 - F(\theta) > 0$ close to $\bar{\theta}$ and $\lambda(\theta) = -F(\theta) < 0$ close to $\underline{\theta}$. For these two expressions of $\lambda(\cdot)$, the monotone hazard rate assumption implies that the quality schedule determined in (30) is increasing in type, i.e. $q_\theta(\theta) > 0$.

Next we proof the existence of θ_1 and θ_2 . By continuity of $\lambda(\theta)$,²⁶ there exists an interval $[\tilde{\theta}_1, \tilde{\theta}_2]$ such that $\lambda_\theta(\theta) = -f(\theta) + \eta(\theta) > 0$ and thus $\pi(\theta) = 0$. Consequently, $\pi_\theta(\theta) = 0$ for all $\theta \in [\tilde{\theta}_1, \tilde{\theta}_2]$. Let θ_1 (θ_2) denote the lowest (highest) $\tilde{\theta}_1$ ($\tilde{\theta}_2$) such that this is true for all $\theta \in [\theta_1, \theta_2]$. By continuity of $q(\theta)$ it follows that $q^l(\theta_1) = k(\theta_1)$ and $q^h(\theta_2) = k(\theta_2)$.

As shown in the proof of proposition 1 the expression in equation (15) is strictly increasing in θ_b . This implies the uniqueness of $\theta_2 = \theta_b$. With a similar argument one shows that

$$S - c_q(k(\theta), \theta) - \frac{F(\theta)}{f(\theta)} c_{q\theta}(k(\theta), \theta)$$

is increasing in θ . This implies the uniqueness of θ_1 which solves

$$S - c_q(k(\theta_1), \theta_1) - \frac{F(\theta_1)}{f(\theta_1)} c_{q\theta}(k(\theta_1), \theta_1) = 0$$

Since

$$S - c_q(k(\theta), \theta) - \frac{F(\theta)}{f(\theta)} c_{q\theta}(k(\theta), \theta) > S - c_q(k(\theta), \theta) + \frac{(1 - F(\theta))}{f(\theta)} c_{q\theta}(k(\theta), \theta)$$

for all θ it follows that indeed $\theta_1 < \theta_2$.

By the uniqueness of θ_1 and θ_2 , c_θ is positive for $\theta < \theta_1$ and negative for $\theta > \theta_2$. Together with (5) this implies the sign of π_θ as stated in the proposition.

In (θ_1, θ_2) , there has to be a type with $\lambda(\theta) = 0$. From (30), this type produces his first best quality and as he is in the zero profit interval $q^{fb}(\theta) = k(\theta)$. The only type satisfying this conditions is the type with the lowest first best welfare θ_w . Note that all $\theta < (>) \theta_w$ have $c_\theta(q(\theta), \theta) \geq (\leq) 0$ and also $\lambda(\theta) < (>) 0$. Differentiating the virtual valuation with respect to θ

²⁶In particular, to connect $\lambda(\theta) < 0$ for small θ with $\lambda(\theta) > 0$ for high θ , we need $\lambda_\theta > 0$ over some range.

tells us the sign of x_θ :

$$\frac{dVV}{d\theta} = c_\theta(q(\theta), \theta) \left(-1 + \frac{\partial \lambda(\theta)/f(\theta)}{\partial \theta} \right) + \frac{\lambda(\theta)}{f(\theta)} c_{\theta\theta}(q(\theta), \theta) \quad (44)$$

From the paragraph above and the monotone hazard rate assumption, the virtual valuation, and therefore $x(\theta)$, has to be decreasing on $[\underline{\theta}, \theta_1]$ and increasing on $[\theta_2, \bar{\theta}]$. On (θ_1, θ_2) , c_θ is zero and as λ flips sign at θ_w the proposition follows.

It was already mentioned that $q(\theta)$ is increasing for types with positive profits. Since $k(\theta)$ is non-decreasing, $q(\theta)$ is non-decreasing for all θ .

Finally, lemma 10 establishes global incentive compatibility. *Q.E.D.*

Lemma 10 *The relaxed solution in proposition 2 is globally incentive compatible.*

Proof of lemma 10 All $\theta \in [\theta_1, \theta_2]$ produce at $k(\theta)$ which is the quality level at which a type has lower cost than any other type. Since these types also have zero profits, no other type can profitably misrepresent as $\theta \in [\theta_1, \theta_2]$. For $\theta \geq \theta_w$ the menu is equivalent to the one described in proposition 1. Therefore, lemma 9 implies non-local IC on this part of the menu. The same proof as for lemma 9 with reversed signs implies that the menu for $\theta < \theta_w$ is non-locally IC.

What remains to be shown is that no type $\theta < \theta_w$ can profitably misrepresent as $\theta' > \theta_w$ (and the other way round). Take such a θ and observe that θ_2 has lower costs at $q(\theta')$:

$$c(q(\theta'), \theta_2) - c(q(\theta'), \theta) = \int_{\theta}^{\theta_2} c_\theta(q(\theta'), t) dt < 0 \quad (45)$$

The inequality follows from the fact that $k(\theta), k(\theta_2) < q(\theta')$ and $c_{q\theta} < 0$. Therefore, the integrand is negative over the whole range. Incentive compatibility for θ requires

$$\begin{aligned} \pi(\theta) &\geq \pi(\theta') + x(\theta')[c(q(\theta'), \theta') - c(q(\theta'), \theta)] \\ &= \underbrace{\pi(\theta') + x(\theta')[c(q(\theta'), \theta') - c(q(\theta'), \theta_2)]}_{\leq 0} + x(\theta')[c(q(\theta'), \theta_2) - c(q(\theta'), \theta)]. \end{aligned}$$

The first term in the last expression is negative because incentive compatibility between θ_2 and θ' is satisfied (see lemma 9 and recall that $\pi(\theta_2) = 0$). The second term is also negative because of equation (45). As $\pi(\theta) \geq 0$, the inequality above and therefore incentive compatibility holds.

The proof for $\theta > \theta_w$ and $\theta' < \theta_w$ works in the same way with θ_1 in place of θ_2 . *Q.E.D.*

Proof of proposition 3 Consider $\theta' = \underline{\theta}$. Define $\underline{W} = W^{fb}(\underline{\theta}) = W^{sb}(\underline{\theta})$. Since $\underline{\theta}$ produces his first best quality and first best welfare is decreasing at $\underline{\theta}$, there are types $\theta > \underline{\theta}$ with lower welfare than \underline{W} . By the definition of the (WNM)-case, $W^{fb}(\bar{\theta}) > \underline{W}$.

Taking these two points together and applying the intermediate value theorem yields the existence of a type θ'' such that $W^{sb}(\theta'') = \underline{W}$ and $W_{\theta}^{sb}(\theta'') > 0$.

$$\frac{dW^{sb}(\theta)}{d\theta} = (S - c_q(q(\theta), \theta))q_{\theta}(\theta) - c_{\theta}(q(\theta), \theta) = -\frac{\lambda(\theta)}{f(\theta)}c_{q\theta}(q(\theta), \theta)q_{\theta}(\theta) - c_{\theta}(q(\theta), \theta)$$

where the first order condition for $q(\cdot)$ is used for the second equality. From proposition 2 and its proof we know that λ and c_{θ} both change sign at θ_w and therefore $sign(\lambda(\theta)) = -sign(c_{\theta}(q(\theta), \theta))$. Consequently, $W_{\theta}^{sb}(\theta'') > 0$ implies $\lambda(\theta'') > 0$ and $c_{\theta}(q(\theta''), \theta'') < 0$.

The virtual valuation can be written as

$$VV(\theta) = W^{sb}(\theta) + \frac{\lambda(\theta)}{f(\theta)}c_{\theta}(q(\theta), \theta)$$

and thus $VV(\theta) \leq W^{sb}(\theta)$ since λ and c_{θ} have opposite signs and the inequality is strict if $\lambda(\theta), c_{\theta}(q(\theta), \theta) \neq 0$.

If $c_{\theta}(q(\theta''), \theta'') < 0$ it then follows that $VV(\underline{\theta}) > VV(\theta'')$. By continuity, there exist types θ that yield strictly higher welfare than $\underline{\theta}$ but still lose from $\underline{\theta}$ in the procurement.

Now consider the case where $\theta'' \in (\theta_1, \theta_2)$ such that $c_{\theta}(q(\theta''), \theta'') = 0$. In this case there are types close to $\underline{\theta}$ that lose from types close to θ'' although the former yield higher (second best) welfare W^{sb} . *Q.E.D.*

Proof of proposition 4: If the scoring rule implements the optimal mechanism it has to hold that $bid(\theta') = bid(\theta'')$ whenever $VV(\theta') = VV(\theta'')$ under the optimal mechanism.

Now take θ_1 and θ_2 as defined in proposition 2. Because $bid_{\theta}(\theta) = -c_{\theta}(q(\theta), \theta)$ and all $\theta \in (\theta_1, \theta_2)$ have $q(\theta) = k(\theta)$, it follows that $bid(\theta_1) = bid(\theta_2)$. As virtual valuation and bids are continuous in type, this implies that $VV(\theta_1) = VV(\theta_2)$ has to hold if the scoring rule implements the optimal mechanism: Otherwise, types slightly below θ_1 and slightly above θ_2 have the same bid but different virtual valuations. Since $q(\theta_i) = k(\theta_i)$, the virtual valuation for θ_i is $Sk(\theta_i) - c(k(\theta_i), \theta_i)$ for $i = 1, 2$. Consequently, the following equation has to hold if the scoring rule implements the optimal mechanism:

$$\int_{\theta_1}^{\theta_2} \frac{d\{Sk(\theta) - c(k(\theta), \theta)\}}{d\theta} d\theta = 0$$

This can be rewritten as

$$\int_{\theta_1}^{\theta_2} \frac{(S - c_q(k(\theta), \theta))c_{\theta\theta}(k(\theta), \theta)}{-c_{q\theta}(k(\theta), \theta)} d\theta = 0.$$

Note that this equation uniquely pins down θ_2 for a given θ_1 .²⁷ Furthermore, it does so independent of the distribution of types. However, θ_2 is defined by the equation $S - c_q(k(\theta), \theta) + \frac{1-F(\theta)}{f(\theta)c_q k(\theta), \theta}$ which depends on $f(\theta_2)$. Hence, slightly perturbing f around θ_2 changes θ_2 but not the equation above. Consequently, a scoring rule auction cannot implement the optimal mechanism in a generic sense. *Q.E.D.*

Proof of proposition 5 As shown in the proof of lemma 9, incentive compatibility between θ and $\hat{\theta}$ boils down to the inequality

$$\int_{\theta}^{\hat{\theta}} \int_t^{\hat{\theta}} x_{\theta}(s)c_{\theta}(q(s), t) + x(s)c_{\theta q}(q(s), t)q_{\theta}(s) ds dt \leq 0.$$

Now note that $c_{\theta\theta} = 0$ implies

$$x_{\theta}(s)c_{\theta}(q(s), t) + x(s)c_{\theta q}(q(s), t)q_{\theta}(s) = x_{\theta}(s)c_{\theta}(q(s), s) + x(s)c_{\theta q}(q(s), s)q_{\theta}(s).$$

But then global incentive compatibility has to be satisfied as $x_s(s)c_{\theta}(q(s), s) + x(s)c_{\theta q}(q(s), s)q_s(s) \leq 0$ by the local second order condition. *Q.E.D.*

²⁷The reason is that the integrand is negative around θ_1 , positive around θ_2 and changes sign only at one type which is between θ_1 and θ_2 . This follows from lemma 1.