A General Theory of Delegated Contracting and Auditing

Wolf Gick∗†

February 15, 2011

Abstract

This paper studies an adverse selection framework of a vertical hierarchy with a top principal, an intermediary with subcontracting power, and a productive agent who is of a continuum of types. The intermediary is hired to forward a screening contract to an interval of agent types. Compared to the literature, the paper differs by (1) using a continuous-type setup to determine the intermediary’s rent (loss of control, agency costs), and (2) by applying internal control or auditing to the continuous-type framework. The intermediary’s rent is determined endogenously. The setup is similar in flavor to auctions with costly participation where the auctioneer has discretion to exclude a nonzero measure of buyers. It also relates to the efficient tax literature under asymmetric information. The paper shows that auditing the intermediary always reduces her information rent and so renders delegated contracting more efficient.

JEL classification: D23, D73, D82, L51.

Keywords: Delegated Contracting, Vertical Hierarchies, Loss of Control, Agency Costs.

∗School of Economics, Georgia Institute of Technology, 221 Bobby Dodd Way, Atlanta, GA 30332, ph. (404) 894 4913, email: wolf.gick@gatech.edu
†Center for European Studies, Harvard University, ph. (617) 495 4303 x244, email: gick@fas.harvard.edu

I thank Jesse Bull, Bob Gibbons, Werner Güth, Dilip Mookherjee, Thomas Jeitschko, Simon Loertscher, Guillaume Roger, Emilson Silva and Jean Tirole for helpful discussions. An earlier version of the paper was given to the Research Seminar at LMU Munich, the ESI Spring 2010 Workshop at MPI Jena, the 2010 IIAC Conference at UBC, the 2010 Stony Brook Game Theory Festival, the Midwest Fall 2010 Economic Theory Meetings, and the School of Economics Seminar at Georgia Tech. All errors are mine.
1 Introduction

Williamson’s (1967) seminal paper on the origins of a “loss of control” across successive vertical layers of a hierarchy has opened several avenues for research. A first strand of literature, initiated by Calvo and Wellisz (1978), has analyzed under which conditions efficient performance monitoring will involve times when a productive worker will not be “checked” by the supervisor. Qian (1994), by adding a “loss of control” feature in the Calvo and Wellisz’s model, has reached a general result in a model where the span of control (i.e. the number of subordinates per supervisor) and wages are determined endogenously; higher effort levels and higher wages at upper tiers may reduce this loss. This, to a large extent, contrasts with Mirrlees’ (1976) findings, namely that “middle managers should get more than workers.”

While Williamson’s (1970) analysis on firm size was based on communication losses, the nature of the information loss across vertical layers has remained an unsolved issue in supervision models. It was Laffont (1990) who has shown that information rents accrue along at different stages of contracts, given what becomes observable along the timeline of the contracting game. When downstream players have an informational advantage, such rents are the result of limited contract design options of the top principal. Notably, this property is not restricted to settings with supervision and collusion.

An important step forward has been made by McAfee and McMillan (1995) who show that, if an intermediary is protected by limited liability and endowed with sub-contracting power, delegated contracting is plagued by a double-marginalization of rents. The intermediary is then in a position to extract information rents from both the downstream agent

---

1 Mirrlees (1976), p. 130.
2 Later contributions have shed important light on issues of adverse selection and hidden information in supervision settings. See e.g. Faure-Grimaud, Laffont and Martimort (2002 and 2003). Still, as Mookherjee (2006) observes in his overview paper, this literature for having been unable to generally explain why delegation may dominate centralization.
and the top principal.

The present paper develops a model in which the intermediary’s rent is determined endogenously, in a setting where the productive agent’s type space is continuous. The “loss of control” that I study is typical for two types of real-world organizations. First of all, it occurs as a control problems inside business firms. The literature on corporate re-engineering takes up this issue. Little has been said in the incentive and contract design literature about how to reduce this loss of control. Consider a multinational firm with branches in different countries. Clearly, headquarters cannot hire productive agents directly but delegate this task to intermediaries (branch managers), endowing the branch manager with sub-contracting power. As Horngren et al. (2003) have pointed out from a cost accounting perspective, decentralization in multinational companies typically comes with suboptimal decision making practices. Not surprisingly, management control systems and performance budgeting are a widespread form of control to restore congruence between top-layers and middle management of the hierarchy.

Secondly, the setting that I study is a widespread phenomenon in public agencies where intermediate layers of the hierarchy have sub-contracting power. In both cases there exist forms of internal control such as bookkeeping and performance budgeting systems that are designed to improve efficiency.

---

3See Brynjolfsson and Hitt (2000).
4Supplier contracts in multinational firms depend on stability, and designing new contracts will typically make it necessary to design new performance evaluation systems. See HBS (1982) on the Lincoln Electric Company case. I thank Christine Ries for pointing this out.
5Government auditing standards (see e.g. USGAO 2010) and management accounting systems aim at limit the discretion of intermediate managers or bureaucrats.
1.1 Relationship to the Literature

While focusing on typical problems of contract design and control in vertical hierarchies, the paper relates to the larger strand of delegation and supervision. Traditionally, most papers on delegation deal with some shifts in the relative advantage between centralized and decentralized organizational forms. As Che and Kim (2006) argue, delegation of contracting authority is difficult to justify in such settings. In turn, models that incorporate collusion such as Faure-Grimaud, Laffont and Martimort (2003), Mookherjee and Tsumagari (2004) and Celik (2009) do not necessarily show an advantage of decentralization in the case of collusion.\footnote{Che and Kim (2006) show in particular that delegation cannot be more justifiable in the presence of collusion than in its absence. My paper does not focus on a setting with two or more productive agents in the Marschak-Radner tradition such as Mookherjee and Tsumagari (2004) and Cella (2005).}

This paper is not on collusion, nor does it directly focus on a comparison of delegation which centralization. Given the setup, there is no general trade-off to expect from specialization and the reduction of information processing costs through having an intermediary offering the contract. Instead, the paper characterizes a delegation benchmark and so explains the emergence of a loss of control under delegated contracting in a one one-agent setting with a continuum of types. More generally, it shows that delegated contracting can be improved through internal control.

To my knowledge, there exists only one paper that extends the paper by McAfee and McMillan (1995) into a problem of delegated contracting with agency costs based on the task itself: Faure-Grimaud and Martimort (2001, FGM hereafter). It uses a discrete type setting in which the downstream agent can be of three possible types. Information, rents, and communication become intertwined in the following way: the intermediary is hired because of her ability to costlessly filter out an unwanted third type of agent before offering a Baron-Myerson (BM) contract to the two remaining types, which she still cannot
distinguish. While their key contribution is to shift the analysis from the simple moral hazard problem in Calvo and Wellisz to the task of forwarding a screening contract to the downstream (productive) agent, there are several issues that require more attention. In Gick (2008) I have extended the FGM setting by applying a viable auditing scheme to the contract, showing that the top principal can always reduce the loss of control that emerges from the intermediary’s discretion.

Different from this earlier research, the task of the present paper is to study a general framework in which a BM-style contract should be offered to a productive agent who is of a continuum of types. This paper studies internal control in a less restrictive framework. A theory of delegated contracting will lead to quite different results when analyzed in a continuous-type framework. Discrete-type settings may show extreme results when one more agent is added. A next advantage of an extension toward a setting with a continuum of agent types7 is that it permits to determine some key parameters endogenously. This not only relaxes the assumptions made in FGM, it also replaces the assumption of a risk-averse intermediary that drives a wedge between possible contract pairs in favor of a more intuitive treatment.

Lastly, the paper offers an auditing scheme for a continuous type setting. Auditing has so far been studied for productive agents at the bottom of a hierarchy, and are typically restricted to a treatment with two discrete types only. By applying auditing to an internal control between the two top layers of a hierarchy, the paper extends the scope of auditing to intermediate players.

The remainder of the paper is organized as follows. Section 2 presents the model, section 3 offers an auditing scheme, section 4 concludes.

7From a theory perspective, continuous-type models are seen as being based on a less restrictive setup, permitting a variety of types. More generally, continuous type model lead to quite distinct results compared to simple discrete-type settings. See e.g. the comparison by Armstrong and Rochet (1999) of discrete versus continuous type models of multidimensional screening.
2 Model

2.1 Players, preferences and payoffs

I use a continuous-type adverse selection environment. There are three players: a principal $P$, an intermediary $I$, and an agent $A$. The agent produces a quantity $q$ of output at a marginal cost $\theta$, which is his own private information or his type. $\theta$ is drawn from continuous distribution on the support $[\underline{\theta}, \bar{\theta}]$, with a c.d.f. of $F(\cdot)$ and a p.d.f. of $f(\cdot)$ that is positive for all $\theta$.

The distribution is common knowledge; I furthermore require that the monotone hazard rate condition holds for the distribution, that is, $\frac{d}{d\theta} \left( \frac{F(\theta)}{f(\theta)} \right)$ is assumed to be nonnegative, and the distribution is well defined and differentiable nearly everywhere over the entire interval. The agent is risk-neutral and has a utility function $U = t - \theta q$, where $t$ is the monetary transfer he receives from the intermediary. The agent accepts to produce as long as he gets his reservation utility exogenously normalized to zero.

The intermediary is hired to forward a screening contract to the productive agent. The rents for this contract are included in the budget $s$. Specifically, the principal requests the intermediary to offer this screening contract to all types of agents in the interval $[\underline{\theta}, \bar{\theta}]$, which would maximize his surplus. The intermediary is risk-neutral but protected by limited liability below zero wealth. She has preferences of $V = s - t$. In other words: subtracting the transfer to the agent, $t$, from the intermediary’s budget $s$ yields the intermediary’s income.

The principal has no access to a productive agent and thus cannot offer a contract himself. His gross surplus is $S(q)$, with $S(0) = 0$, $S'(q) > 0$ and $S''(q) < 0$. To ensure positive production levels, I assume that the Inada conditions hold, that is $R'(0) = +\infty$.

---

\footnote{As usual in this literature, I use male pronouns for principal and agent, and a female for the intermediary.}
and $R'(\infty) = 0$. The principal’s net surplus is simply his gross surplus minus the budget paid to the downstream hierarchy $S(q) - s$.

### 2.2 Timing

The contracting game has the following extensive form:

- $(t = 0)$ Agent learns its type $\theta$. Intermediary learns the agent’s type only if $\theta > \hat{\theta}$.
- $(t = 1)$ $P$ offers a Grand-Contract to $I$ specifying output targets and transfers
- $(t = 2)$ Intermediary accepts or rejects.
- $(t = 3)$ Subcontracting stage: intermediary offers a sub-contract to agent.
- $(t = 4)$ Agent accepts or rejects.
- $(t = 5)$ Production and transfers take place. Outputs are observed. Game ends.

### 2.3 Agent’s constraints

To find the optimal contract for the intermediary as a principal who optimally should forward a screening contract over the range $[\hat{\theta}, \theta]$ with $\hat{\theta} < \bar{\theta}$, I first reduce the setup to a two-layer vertical hierarchy consisting of a principal and the productive agent. This step is easy to motivate: as in FGM and related settings, the rent of the intermediary is a function of the rent of the productive agent. To find this rent, I characterize the optimal contract for the productive agent over the interval $[\hat{\theta}, \bar{\theta}]$.\footnote{To a large extent, my exposition and notation follows Laffont and Martimort (2002) (LM hereafter) at this point. Note that the contracting type space is not the entire type space, thus permitting the typical and “garbled” information structure as in FGM, in which the intermediary always has the option to misrepresent a type within $[\hat{\theta}, \bar{\theta}]$ as a type $\theta > \hat{\theta}$. For the continuous type case I reach $F(\theta)|_{\hat{\theta}} < F(\theta)|_{\bar{\theta}} = 1$.}
I now state monotonicity, participation and incentive compatibility constraints that characterize simple direct revelation mechanisms. Specifically, I restrict attention to direct revelation mechanisms \( q(\tilde{\theta}), t(\tilde{\theta}) \) for which

\[
t(\theta) - \theta q(\theta) \geq t(\tilde{\theta}) - \theta q(\tilde{\theta}).
\] (1)

2.3.1 Local incentive constraints

Truthful revelation contracts need to satisfy the following first order condition for the truthful response \( \tilde{\theta} \) to become optimal:

\[
\dot{t}(\tilde{\theta}) = \theta \dot{q}(\tilde{\theta}) = 0
\] (2)

Should this hold for all \( \theta \) in the type space, it must be the case that

\[
\dot{t}(\theta) = \theta \dot{q}(\theta) = 0,
\] (3)

or, in other words, for reporting any \( \theta \) different from the true type, it must not increase his rent.

From the second-order condition,

\[
\dot{\ddot{t}}(\tilde{\theta})\bigg|_{\tilde{\theta} = \theta} - \theta \ddot{q}(\tilde{\theta})\bigg|_{\tilde{\theta} = \theta} \leq 0,
\] (4)

or

\[
\ddot{t}(\theta) - \theta \ddot{q}(\theta) \leq 0.
\] (5)

Differentiating (3) permits to rewrite (5) into

\[
-\dot{q}(\theta) \geq 0,
\] (5')

where (3) and (5') ensure that the agent does not misreport his type locally.
2.3.2 Global incentive constraints

It can be shown that the relevant local constraint (3) also holds globally, that is, for all types besides adjacent types, one can use (3) to replace the R.H.S. of (1) by:

\[
t(\theta) - t(\tilde{\theta}) = \int_{\tilde{\theta}}^{\theta} \tau \dot{q}(\tau) d\tau = \theta q(\theta) - \tilde{\theta} q(\tilde{\theta}) - \int_{\tilde{\theta}}^{\theta} q(\tau) d\tau
\]

(6)

Isolating the terms without tilde and \(\tau\) leads to:

\[
t(\theta) - \theta q(\theta) = t(\tilde{\theta}) = t(\tilde{\theta}) - \theta q(\tilde{\theta}) + (\theta - \tilde{\theta}) q(\tilde{\theta}) - \int_{\tilde{\theta}}^{\theta} q(\tau) d\tau
\]

(7)

The term right of the plus sign is nonnegative, which implies that the local incentive constraint (2) holds for all \(\theta\).

We now use rent notation for setting up the optimal contract, with \(U(\theta) = t(\theta) - \theta q(\theta)\). Substituting in (2), we have

\[
\dot{U} = -q(\theta).
\]

(IC)

The principal’s optimization problem can now be expressed over rents and outputs in the following form for the contracting space \([\theta, \tilde{\theta}]\):

\[
\max_{\{U(\cdot), q(\cdot)\}} \int_{\tilde{\theta}}^{\theta} \left( S(q(\theta)) - \theta q(\theta) - U(\theta) \right) f(\theta) d\theta
\]

(P)
s.t. (IC), a nondecreasing schedule of outputs (M) and a nonnegative rent for all agents (IR):

\[ \dot{U} = -q(\theta). \] \hspace{1cm} \text{(IC)}

\[ \dot{q}(\theta) \leq 0 \] \hspace{1cm} \text{(M)}

\[ U(\theta) \geq 0. \] \hspace{1cm} \text{(IR)}

Because of (IC), (IR) simplifies to

\[ U(\hat{\theta}) \geq 0, \] \hspace{1cm} \text{(IR)}

which, as in the discrete type case, becomes binding: the least efficient type of agent is given no rent. Because of the earlier stated monotone hazard rate property \( \frac{d}{d\theta} \left( \frac{F(\theta)}{f(\theta)} \right) \), condition (M) is always fulfilled by the optimal contract and will be slack in the optimum.

Solving (IC) yields

\[ U(\hat{\theta}) - U(\theta) = -\int_{\theta}^{\hat{\theta}} q(\tau)d\tau. \] \hspace{1cm} \text{(8)}

Last, because of (IR) \( U(\theta) = 0 \) and

\[ U(\theta) = \int_{\theta}^{\hat{\theta}} q(\tau)d\tau. \] \hspace{1cm} \text{(9)}

Replacing \( U(\theta) \) in (P) by its R.H.S. permits us to express the principal’s program in a reduced form:
\[
\max_{\{U(\cdot), q(\cdot)\}} \int_{\theta}^{\hat{\theta}} \left( S(q(\theta)) - \theta q(\theta) - \int_{\theta}^{\hat{\theta}} q(\tau) d\tau \right) f(\theta) d\theta. \tag{P'}
\]

Note that \( \int_{\theta}^{\hat{\theta}} \left( \int_{\theta}^{\hat{\theta}} q(\tau) d\tau \right) f(\theta) d\theta \) is the expected rent of the agent of type \( \theta \), or \( E(U(\theta)) \).

Using integration by parts, this expression simplifies into
\[
\int_{\theta}^{\hat{\theta}} \left( \int_{\theta}^{\hat{\theta}} q(\tau) d\tau \right) f(\theta) d\theta = \left. \left( \int_{\theta}^{\hat{\theta}} q(\tau) d\tau \right) dF(\theta) \right|_{\theta}^{\hat{\theta}} - \int_{\theta}^{\hat{\theta}} F(\theta)(-q(\theta)) d\theta
\]
\[
\int_{\theta}^{\hat{\theta}} \left( \int_{\theta}^{\hat{\theta}} q(\tau) d\tau \right) f(\theta) d\theta = \int_{\theta}^{\hat{\theta}} F(\theta) q(\theta) d\theta,
\]
which yields
\[
U(\theta) = \int_{\theta}^{\hat{\theta}} \left( \int_{\theta}^{\hat{\theta}} q(\tau) d\tau \right) f(\theta) d\theta = \int_{\theta}^{\hat{\theta}} \frac{F(\theta)}{f(\theta)} q(\theta) f\theta d\theta. \tag{10} \]

Lastly, the closed-form expression of the principal’s program is reached:
\[
\max_{\{U(\cdot), q(\cdot)\}} \int_{\theta}^{\hat{\theta}} \left( S(q(\theta)) - \left( \theta + \frac{F(\theta)}{f(\theta)} \right) q(\theta) q(\theta) \right) f(\theta) d\theta. \tag{P''}
\]

**Lemma (LM, 2002).** The program’s solution entails:

\[
S'(q^{SB}(\theta)) = \theta + \frac{F(\theta)}{f(\theta)}, \tag{10}
\]

implying no downward distortion for the most efficient type \( \theta \) and a decreasing schedule of outputs \( q \) for all other types. This output distortion is second-best optimal.

---

\(^{10}\)Note in particular that this holds because the density at the left margin must be zero, or \( F(\theta) = 0 \), and that the IC constraint was obtained by differentiating \( \frac{d}{d\theta} \left( \int_{\theta}^{\hat{\theta}} q(\tau) d\tau \right) = -q(\theta) \).
2.4 Grand Contract

After finding the optimal contract in a two-player adverse selection game for the contracting space of $[\theta, \hat{\theta}]$, I now characterize the vertical hierarchy through the optimal Grand Contract, which is carried out sequentially. The goal is to find the optimal solution and to check if additional distortions become optimal when an intermediary is added to the hierarchy.

**Definition 1 Grand-Contract.** A Grand-Contract $GC = s(q)$ is a direct truthful mechanism that satisfies all rents of all players in the hierarchy, with the revelation principle applying to the sub-contracting stage.

Without loss of generality, I restrict attention to direct truthful mechanisms $(t, q)$ between $I$ and $A$ that include budgeting between $P$ and $I$ with $I$ receiving a payment $s - t$. To see this, define the contract from the intermediary’s perspective, replacing transfers as functions of output targets $q$, and expressing the expected utility of the intermediary when designing the screening contract for the agent:

$$
\max_{\{U, q\}} \int_\theta^{\hat{\theta}} V(s(q)) - \theta q - U(\theta) f(\theta) d\theta.
$$

(\text{GC}_{SC})

2.4.1 Delegation Proofness

Should the intermediary be incentivized to offer a screening contract for all types in $\frac{d}{d\theta} \left( \frac{F(\theta)}{F(\hat{\theta})} \right)$, the optimal sub-contract needs to implement optimal output targets $q(\theta)$, fulfilling the following incentive constraints between intermediary and agent:

$$
s(q(\theta)) - \theta(q(\theta)) \geq s(q(\theta')) - \theta(q(\theta')) \text{ for all } \theta \leq \theta'.
$$

This implies that for more efficient agent types, $I$ needs to receive an at least weakly higher payment scheme, and that this must hold for all output targets in the contract.
Similarly to the *Revelation Principle* for the simple two-player Principal-Agent model, I now apply the *Delegation Proofness Principle* that implements a truth-telling contract across the hierarchy, with optimal output targets and information rents paid to each player: take any GC such that the optimal sub-contract recommends production $\tilde{q}(\theta)$ and a production target of zero for all types of agents with higher marginal costs than envisaged by the principal. It must then be possible to establish a delegation proof contract as a direct mechanism defining output targets $q = \tilde{q}(\theta)$ and budgets $s = s(\tilde{q}(\theta))$ such that the intermediary truthfully reveals the agent’s type to the principal. If not, the envisaged output target $\tilde{q}(\theta)$ would not have been optimal in the first place. This can be summarized in the following statement using contracts as pairs of output targets and budgets.

**Definition 2 Delegation Proof Grand-Contract.** A delegation-proof Grand-Contract for a continuum of agent types consists of a menu $\{(s, q)\}$ for all agent types, satisfying the intermediary’s incentive constraints for each intermediary-agent coalition with $t = \theta q$ being the transfer paid to the agent, and $s - t$ the part of the budget that $I$ keeps for herself:

$$s - \theta q \geq s' - \theta(q'),$$

where $s' > s$ and $q' > q$ denote budgets and outputs that are suboptimal for types $\theta' > \theta$.

Furthermore, the intermediary is protected by limited liability against outcomes below zero, with her rent being $V(\theta)$ being

$$V(\theta) \geq 0. \quad (\text{LL})$$

For an agent type of $\hat{\theta}$ there is no rent to be offered by $I$. That is, would $P$ want that $I$ offers a contract only to the $\hat{\theta}$-type of agent, there would be no need to include any rent. For all lower types, the intermediary’s rent may become positive.
2.4.2 Downward Incentive Constraint of Intermediation

In a (real) two-type setting as in FGM, the intermediary’s rent follows from her option to “gamble” and to offer a contract to the most efficient type only. Specifically, she can offer a shutdown contract to the most efficient agent in the discrete type setting, reaping the entire information rent of this type with a probability that is exogenously given. If the intermediary loses the gamble, no contract exists and no production occurs. If she succeeds, she pockets the entire rent she is hired to include in the contract design.

The results of the continuous type setting differ from FGM in two important aspects. First, the intermediary cannot offer a contract to a close to zero mass of agent types on the left side. Second, since the mass between the most and the least efficient agent type is positive everywhere for any nonzero subset of types, the rent analysis for the intermediary is built on an entirely different concept.

The optimal contract can be characterized using the following steps. Observe first, the intermediary requires a strictly positive rent for truthfully forwarding a BM-style contract to all agent types in the interval $[\theta, \hat{\theta}]$ except for the highest type. If not, the contract would not be delegation proof and the intermediary could do better to cut off a positive mass of agent types and offer a BM-style screening contract to a subinterval from the most efficient type up to a cutoff type $[\theta, \theta^C]$ that maximizes her rent.

I characterize the origin of the intermediary’s rent (loss of control, agency cost) for any $\theta^C \in [\theta, \hat{\theta}]$:

$$E(V_{\theta^C}) = \hat{\theta} \int U(\theta)d\theta - \int_{\theta^C}^{\hat{\theta}} U(\theta)d\theta - \int_\theta^{\theta^C} U(\theta)d\theta$$  \hspace{1cm} (11)
Proposition 3 The intermediary chooses a cutoff value $\theta^C \in [\underline{\theta}, \hat{\theta}]$ that maximizes her rent $E(V^*_\theta^C)$. The principal includes $E(V^*_\theta^C)$ in the Grand Contract to reach delegation proofness. Given single-peaked densities that satisfy monotone hazard rate property, there is always an interior solution for $E(V^*_\theta^C)$.

The above illustration (Fig. 1) sketches the optimal cutoff that the intermediary chooses to maximize her rent. The findings are summarized in the following proposition.

Proposition 4 Because of her subcontracting power, the intermediary can offer any sub-contract to $A$ over a subset of types. For any leftbound subset of the contracting space, $[\underline{\theta}, \theta^C]$, with $\theta^C < \hat{\theta}$ she can reap a strictly positive rent. If the contract was accepted, she was able to pocket parts of the information rent included in the budget for herself. If not, no contract exists and the intermediary misreports the type being in the interval $[\hat{\theta}, \bar{\theta}]$.
As a last exercise, I proceed similar to Section 2.3 where we derived \((P'')\) and first set up the Grand-Contract over rents and outputs:

\[
\max_{\{U(\cdot), V(\cdot), q(\cdot)\}} \int_{\theta}^{\hat{\theta}} \left( S(q(\theta)) - \theta q(\theta) - U(\theta) - V(\theta) \right) f(\theta) d\theta. \tag{GC}
\]

**Proposition 5 Optimal Delegation-Proof Grand-Contract.** The optimal contract entails additional distortions compared to second-best outputs, with

\[
S'_{GC}(q^{SB}(\theta)) = \theta + \frac{F(\theta)}{f(\theta)} + \frac{F(\theta)}{f(\theta)} \Bigg|_{\theta}^{\hat{\theta}} - \frac{F(\theta)}{f(\theta)} \Bigg|_{\theta}^{\theta_{GC}}. \tag{12}
\]

Note that the last three parts of the sum follow from the additional rent that the delegation proof Grand-Contract includes. This increases the slope of the principal’s surplus function and reduces efficiency. The proof is given in the appendix.

### 3 Auditing

This section shows that it is possible to audit a contract under a continuous type setting, to restore efficiency to some degree. I assume that the principal has access to a costly audit technology and can commit to audit the intermediary whenever a contract exists.

The intermediary is now imposed a penalty \(P^s\) with probability \(\varnothing\) if the examination of the written subcontract detects an irregularity in the contract offer, while with probability \(1 - \varnothing\) she keeps her information rent as before. This simple control scheme does not need to involve probabilistic auditing; as long as the principal commands a costly but fully revealing examination technology, it is sufficient for the principal to examine the contract whenever
output was observable. The argument is similar as in Gick (2008) where I have provided an auditing scheme to the FGM contract.

$$\varphi \left( E(V(\theta)^*) - P^* \right) + (1 - \varphi)E(V(\theta)^*)$$

With endogenous punishment, \( E(V(\theta)) \) reduces to

$$E(V(\theta)^A) = (1 - \varphi) \left( \int_{\theta}^{\theta} U(\theta) d\theta - \int_{\theta}^{\theta} V(\theta) d\theta - \int_{\theta}^{\theta} U(\theta) d\theta \right) < E(V(\theta)^*). \quad (13)$$

The principal's problem under auditing changes to:

$$\max_{\{U(\cdot),V(\cdot),q(\cdot),c(\varphi)\}} \int_{\theta}^{\theta} (S(q(\theta)) - \theta q(\theta) - U(\theta) - V(\theta)) f(\theta) d\theta$$

**Proposition 6** Under auditing, the optimal downward distortion in the solution to the delegation proof Grand-Contract is reduced. The principal chooses an audit probability \( \varphi \) equal to the marginal cost of auditing and the optimal Grand-Contract now entails:

$$S'_A(q^{SB}(\theta)) = \theta + \frac{F(\theta)}{f(\theta)} + (1 - \varphi) \left( \frac{F(\theta)}{f(\theta)} \bigg|_{\theta^C}^{\theta} - \frac{F(\theta)}{f(\theta)} \bigg|_{\theta}^{\theta^C} \right) < S'_G(q^{SB}(\theta)). \quad (14)$$

The proof is analog to the one of **Proposition 5** (utilizing (14)) and omitted.

As a result, **Proposition 6** shows that auditing increases efficiency and reduces the optimal distortion in the delegation proof Grand-Contract. It decreases the slope of the surplus function for all \( \theta \) and improves efficiency of the Grand Contract.
4 Concluding remarks

The paper has presented a generalized treatment of delegated contracting and internal control. The central findings of the paper are the following. First, it extends the findings of the previous literature on delegated contracting and organizational diseconomies of scale in the following three ways:

- The origins of agency costs or a loss of control are no longer directly based on the span of control as in FGM, that is, the type difference between the highest and lowest cost type. Instead, as this paper has shown, in a continuous type setting the intermediary can only grasp parts of the rent that the top principal includes to be forwarded to the productive agent.\textsuperscript{11}

- The continuous-type model furthermore shows that the intermediary will always cut off high-cost types of agents from receiving a contract. While this option of the intermediary is the source of inefficiency, she does not have an incentive to exclude efficient types in the regime. Note also that the paper endogenously derives the rent of the intermediary.

- There exists a simple auditing scheme that always reduces inefficiency. Auditing schemes of this kind are typically used together with budgeting techniques in firms to enhance efficiency. The novel contribution of this paper is that it offers an auditing scheme which works with continuous types of agents. Adverse selection models with an auditing scheme for the productive types typically restrict their attention to discrete types only.

The results shed new light on the specific issues of agency costs that typically arise in large multinational firms and in public bureaucracies. Seen in a continuous type framework, the nature of the “loss of control” is quite different, and the setting itself limits the options

\textsuperscript{11}In FGM, the rent is the prior of the efficient agent’s type times the entire virtual cost $\Delta \theta \bar{q}$. In this way, the findings of my paper show that vertical hierarchies, in a more general setting, show a lower loss of control compared to the result of discrete type models.
of the intermediary of requiring rents up to the level of the downstream agent’s rent. That is, a top principal should be aware of losses in the hierarchy, but adding a next layer to a hierarchy does clearly not imply doubling information rents as modeled in the literature. In addition, the paper has shown that even if the top principal has no access to the bottom tier, he generally has some leeway to reduce the loss through auditing.

As already stressed, the present paper is not about collusion. Still, collusion constitutes a borderline case that may be worth extending: the intermediary offers a take-it-or-leave-it subcontract and never leaves any rent to the productive agent beyond what a standard screening contract would encompass. Collusion does not arise in the model because the intermediary does not observe a partition of the contracting space that would endow him with additional information. Extending the intermediary’s observation space into the contracting space and so open the setting toward collusion with a subset of agent types may be one worthwhile extension.\(^{12}\)

Two last comparisons and extensions may be worth exploring. First, the findings of this paper concerning the intermediary’s rent are reminiscent of the literature on optimal income tax schedules under asymmetric information. Mirrlees (1971 and 1997), as well as Seade (1982) and Guesnerie and Seade (1982) show properties of an optimal tax for different, continuous income “types” that resembles the endogenously determined information rent.\(^{13}\) Second, and as mentioned at the beginning, the setup between top principal, intermediary and agent is similar to an auction environment with costly participation involving with a seller, auctioneer and a continuum of buyers. If the auctioneer has discretion to exclude a nonzero subset of buyer types, e.g. by designing access in a way that the seller would find suboptimal, the optimal contract will require additional information rents for the

\(^{12}\)Celik’s (2008) paper is, to my knowledge, the only contribution so far using a continuum of agent types in a setting with principal, insurer and agent where the principal facing a budget constraint. The vertical structure is different to mine, and there is no endogenously determined information rent of the insurer.

\(^{13}\)I thank Emilson Silva for pointing this out.
auctioneer.\textsuperscript{14} This, as well, is left for future research.

5 Bibliography


\textsuperscript{14}See Menezes and Monteiro (2000) for an overview as well as McAfee and McMillan (1987a,b). Celik and Yilankaya (2009) study an optimal auction with stochastic bidder participation and endogenous cutoffs.


6 Appendix

Proof of Proposition 5.

Re-expressing $E(V(\theta))$ from (11) into

$$
\int_{\theta}^{\hat{\theta}} U(\theta)d\theta - \int_{\theta}^{\hat{\theta}} U(\theta)d\theta - \int_{\theta}^{\hat{\theta}} U(\theta)d\theta = \int_{\theta}^{\hat{\theta}} \frac{F(\theta)}{f(\theta)} q(\theta) f\theta d\theta - \int_{\theta}^{\hat{\theta}} \frac{F(\theta)}{f(\theta)} q(\theta) f\theta d\theta - \int_{\theta}^{\hat{\theta}} \frac{F(\theta)}{f(\theta)} q(\theta) f\theta d\theta
$$

which gives the required result.  

\[\Box\]