Mix Stickiness under Asymmetric Cost Information

Antle, Rick; Bogetoft, Peter

Link to publication in CBS Research Portal

Mix Stickiness under Asymmetric Cost Information

Rick Antle, Peter Bogetoft

Mix Stickiness under Asymmetric Cost Information

Rick Antle
Yale School of Management

Peter Bogetoft
Copenhagen Business School

In most organizations, the agent has superior information about the relative costs of different products or activities. This handicaps the principal when he seeks to adjust the product mix.

Our main conceptual finding is that there is "mix stickiness". The agent's superior information about the relative costs of different products or activities leads to an advantage for the status quo; i.e., there is inertia (or stickiness) in the mix of products or activities pursued. The historical mix has an advantage because the asymmetric information about relative costs has less of an impact when the mix does not change. If the mix changes, the producer can extract information rents by claiming high costs on the least reduced or most expanded products. Only in the case of larger shifts in the environment will the mix change. Changing the mix comes with the advantage of making rationing cheaper.

Our analysis gives new insight into several business practices. It suggests, for example, that for moderate downsizing, a reinforcement approach of proportionally cutting all activities (lawn mowing) is optimal. When more dramatic downsizing is called for, a reorientation approach of eliminating certain activities (divesting) is optimal. It also suggests that outsourcing may come at the cost of reduced adaptability, and it may help explain why most, if not all, healthcare systems struggle to control cost

Key words: Production mix, Production scope, Asymmetric Information, Rationing, Downsizing, Outsourcing, Health care cost

1. Introduction

Choosing a product mix is one of the central management problems in a multiproduct firm, and there is a long history of studying and trying to improve the mix choice. One focus has been on formulating and solving multidimensional constrained optimization models of production. Such studies often assume that the product prices and costs are known and seek the mix that maximizes profit subject to technical constraints. Another focus has been on the development of cost information to support these decision models. Product costing begins with accumulating total costs, which

is assumed to be a relatively trivial exercise, and proceeds to allocate or assign costs to specific products or services, which is the difficult part. For example, activity-based costing systems seek to assign the pool of indirect costs to products or services in accordance with the activities that generate them.

Intuitively, it is not hard to see how better cost information can improve product mix decisions. Of course, installing and maintaining a costing system are themselves costly activities, and to decide whether or not to undertake them, we must know what would be optimal to do if these better cost estimates are not produced. This is the focus of our paper.

We explore optimal product mix decisions in an organizational context in which superior information about costs is held by members within the organization. We study a principal-agent model in which the agent produces two products (activities, services). The agent knows the unit costs of the different products. The principal, conversely, does not know the relative costs of the different products; he only knows the activities and the total costs in the last period. Hence, a distinguishing feature of our models is that the production and its total cost in the prior period are known to the principal, as it would be under a costing system that accumulated total costs but did not assign them to separate products. Further, the relative values of the products may change over time, raising the possibility that the mix should be changed and, therefore, that more detailed product cost information might well be useful.

To give an example, consider a university president considering the teaching and research activities of a department. He knows last year's mix of teaching and research activities, and he knows the total cost incurred by the department last year. Teaching has become more valuable this year, however, and he would like to increase teaching output by 20 %. One feasible plan is to increase the departments budget by 20 %. This would cover the departments increased cost of the additional teaching, even in the worst case in which teaching is the activity that causes all the costs. Of course, if teaching is not the sole cost driver, the 20 % budget increase will more than cover increased costs, and the department will enjoy a surplus of resources. The president would like to structure a more efficient budget if possible.

Although the president only knows last years total costs, the department might know more. Suppose it knows the separate costs of teaching and research. The president might want to structure some alternatives for the department and let it choose which one it wants. For example, he might offer the department a budget increase of 10% in exchange for a 20% increase in teaching and the same research as last year. This might well be attractive to the department if its teaching costs are relatively low. However, if teaching is the primary cost driver of the department, the department may prefer to take last year's budget and produce last year's teaching and research activities. If the value of teaching is relatively low, this might be a risk the president wants to take. However,

if the value of teaching is very high, it might be best for him just to give the 20% increase and ask the department to scale up both activities.

In our model, the optimal outcome balances the advantages of adjusting the mix to changing values and the costs of introducing such changes under asymmetric information. As it turns out, it will often be optimal to stick to the old mix. The agent's superior information leads to an advantage for the status quo; i.e., there is inertia (or stickiness) in the mix of products or activities pursued. The historical mix has an advantage because the asymmetric information about relative costs has less of an impact when the mix does not change. If the mix changes, the producer can extract information rents by "claiming" high costs on the least reduced or most expanded products.

Two aspects of the setting turn out to influence the stickiness of the mix decision. It is important how lucrative or profitable the setting is, i.e., how valuable the production is relative to the underlying costs to the agent and the agents ability to generate information rents. Likewise, it is important whether the agent can communicate with the principal.

Consider first the case of no communication. In a lucrative setting, where it is too costly (or not feasible) for the principal to reduce the level of some activities below their ideal levels (we call this "rationing"), the information asymmetry makes a proportional change strategy (second best) optimal. Proportional adjustments limit the agent's gain from cost shifting. In a less lucrative setting, where rationing is necessary, it may be better for the principal to simply bet on which product has the lowest cost, i.e. to reorient and abandon certain activities.

When communication is possible, the outcomes change. Communication presents the opportunity to engage in more nuanced behavior than the extremes of simple proportional adjustment and gambling on the lowest cost product. In a lucrative, no-rationing setting, the mix becomes less sticky since the parties can take advantage of adjusting the mix somewhat closer to the demand. In a less lucrative setting, communication presents an opportunity to control rents by using a more finely tuned approach that biases production toward the historical mix as opposed to simply gambling on one product or the other. In effect, stickiness is increased.

The outline of this paper is as follows. In Section 2, we discuss the literature that is most closely related to the specific problem we study. In Section 3, we introduce the basic assumptions about the costs and benefits of different activities. In Section 4, we consider the optimal strategy under perfect information. In Section 5, we consider optimal strategies when rationing is not possible, and in Section 6, we consider the optimal strategy when rationing is allowed. In Section 7, we discuss extensions of our model to include more general value functions and to dynamic contracting environments. All of the results are illustrated by numerical examples included in the different sections. Applications in a selection of business problems, viz. the downsizing of an organization, the outsourcing of production and the cost containment in health are discussed in Section 8. Final

remarks are provided in Section 9. In the Appendices, we provide more details about the numerical calculations and the extensions of the models.

2. Literature

Product mix decisions have been studied in the economics, operations research and accounting literature. In economics, the standard textbook treatment of supply begins by assuming that the firm is a single "entity" that maximizes profits by choosing its outputs and the inputs required to produce them (e.g., see Varian (2014)). The revenue and cost functions are both simple and known: market prices, assumed known, exist for all outputs and inputs; therefore, revenue and cost functions are simply the sums of prices times quantities. Classic comparative statistics analysis is then used to study how input and output decisions are related to both the various (known) prices and the characteristics of the production function.

The operations research literature also assumed the firm was a single entity that maximized profits and initially focused on two problems: 1. How to make optimal production decisions in the face of constraints; and 2. Whether it is possible to decompose the firm's optimization problem into smaller pieces that are easier to solve and, if so, how to perform this decomposition (e.g., see Dantzig and Wolfe (1960)).

The early accounting literature mirrors these two streams of economics and operations research literature, which is reflected in managerial accounting textbooks to this day. For example, one subject of interest is how to allocate joint costs in a multi-product environment. In activity-based costing, the focus is on more accurate assessments of incremental costs, one objective of which is to improve decisions about product mix (e.g., see Datar and Rajan (2014)).

There was, however, dissatisfaction with the view that firms act as monolithic profit maximizers. Firms, by definition, are non-market institutions, and understanding their activities should acknowledge the conditions that lead to their creation. In economics, March and Simon (1958), Williamson (1985) and others began to ask in a deeper way what determined the behavior of firms, which they viewed as consisting of many participants, each with their own information and incentives.

This focus on information and incentives has led to a vast literature on screening and signaling. In signaling models, better informed, economically advantaged agents take actions to signal their advantages, such as offering warranties. In screening models, less informed economic agents structure choices for the better informed that take (at least partial) advantage of their better information. We take a screening approach and, like much of the literature, rely on the revelation principle to focus on menus of contracts that induce truthful communication, albeit at the cost of underutilizing the information obtained.

4

The focus on information and incentives has been used to study a variety of problems, such as a government's problem with better informed taxpayers when supplying a public good or in regulating a firm[1], a principal's problem with an agent who might be a subordinate in a hierarchical organization[2] or a government's problem in procuring goods or services from a better informed defense contractor[3]. More recently, optimal screening contracts have been applied in a variety of supply chain contexts where the seller possesses private information about aspects such as marginal cost, risk attitude, or information regarding supply reliability; see, e.g., Kouvelis et al. (2006) for a comprehensive review.

In the accounting literature, the ability to analyze problems involving diversely-informed economic agents in models with endogenous determination of their incentives brought new life to the areas of budgeting and performance evaluation. The treatment of budgeting had traditionally focused on compiling and processing data. For example, textbooks have described in detail the process of moving from sales forecasts, to revenue and expense forecasts, to balance sheet forecasts and then to cash management. Although there was some recognition that this process was dependent on information supplied by self-interested individuals (e.g., see Horngren (1981)), this concern was not integrated into the overall decision-making on the product mix, which continued to be studied as a quest for obtaining the first-best, or full information solution.

In the presence of asymmetric information and incentive problems, we are led to a different view of budgeting and performance evaluation practices. For example, in the accounting literature, there are many studies of the principal-agent model in which the agent has superior information that allows him to extract informational rents. The principal's best response to the agent's superior information depends on how valuable the agent's production is. If production is very valuable, the principal may have to live with "organizational slack". If production is not too profitable, the principal's best response involves leaving some of what would otherwise be valuable production undone in order to economize on expected organizational slack (e.g., Antle et al. (1999), Antle and Eppen (1985), Antle and Fellingham (1997)).

Most analyses, whether in the economics or accounting literature, have restricted the dimensionality of the problems to gain tractability. In particular, private information is most frequently restricted to be of a single dimension, and the main results of this research stream have been a) the full-information, first-best solution usually cannot be implemented; and b) the good agent (e.g., the supplier with lower cost) is producing at the first-best level, while bad agents (e.g., suppliers with

---

[1] See Myerson (1979), Baron and Myerson (1982) and Laffont and Tirole (1986).

[2] Sappington (1983), Harris and Raviv (1996), M. Harris et al. (1982), Antle and Eppen (1985), and McAfee and McMillan (1988)

[3] See Rogerson (1994, 2003).

higher cost) produce less than the first-best level, which is often referred to as the "no distortion at the top" and "downward distortion at the bottom" idea[4].

In practice, however, these problems rarely involve only a single piece of information, and multi-dimensional private information severely complicates the analysis.[5]

In the setting of a multi-product firm facing consumers with (multi-dimensional) tastes for the various products, it has been shown, e.g., Armstrong (1996), that the optimal solution for the firm usually involves some rationing, i.e., excluding some consumers from its products in order to extract more revenue from the other higher value consumers. Likewise, it has been shown by Rochet and Chone (1998) that it is usually optimal to apply some level of "bunching", forcing some consumers with different tastes to choose the same bundle of products instead of using a more complete adaption of characteristics to tastes. The optimal construction of such bunches is supported by the "ironing" procedure in the one-dimensional case (Mussa and Rosen (1978)) and by its generalization, the "sweeping' procedure in the multiple-dimensional case (Rochet and Chone (1998)). A general overview of the multidimensional screening is provided in Rochet and Stole (2003). Although important insights have been derived about multiple-dimensional screening mechanisms with continues types, it has been hard to generate closed-form solutions to particular problems. It usually requires some simplifying assumptions like discrete type space, cf. Armstrong and Rochet (1999), before a case can be explicitly solved.

In our paper, the agent produces multiple products, and his type space is multiple-dimensional, viz. his unit costs of the different products. We do, however, have additional features that simplify the analysis. First, we assume an underlying linear cost function, and we make the assumption that the types are restricted to those that are able to produce the historic production bundle using the historic costs. Our type space is therefore simplified. Second, we focus on the choice of a production mix and largely avoid scale issues. Both features simplify the analysis by making it possible to explicitly identify the binding incentive compatibility constraints. The difficulty with general multiple-dimensional screening is that there is no exogenous type-ordering as in the single-dimensional case. Our assumptions help give us a partial ordering of types, which in turn makes it possible to derive several insights without too many complications. Our solutions have many of the usual features of the general screening problem. Certain types are rationed away entirely, and one of our contributions is to analyze which types of agents to ration away. In addition, we

---

[4] Note that the traditional focus has been on the distortions in the production scale coursed by asymmetric information. In this paper, we also stress the distortion in production scope that is necessitated by asymmetric information.

[5] Even more problematic is the fact that the results of multidimensional analyses deviate significantly from the one-dimensional results. For example, Holmstrom and Milgrom (1991) study a multi-task hidden information problem and demonstrate that incentives that are too strong on a more easily measureable task can undermine the production of a task that is more difficult to measure

have bunching, i.e., agents with different unit costs produce the same product bundle, but in our simplified model, we can determine the optimal bunching by a relatively simple trade-off between the extra costs of bunching and the benefits in terms of reduced rents to the other types much like in the single-dimensional case.

In addition to the above linkages to, in particular, the accounting literature, let us finally mention two other streams of literature that our work has partial links to.

The non-rationing outcomes in a multi-dimensional types model is technically related to Debreu (1951), who proposed evaluating the "dead loss" associated with a Pareto non-optimal situation of an economic system in terms of the so-called coefficient of resource utilization ρ. It measures the maximal proportional reduction in available resources that enables a system to generate the same levels of satisfaction. Debreu (1951) also developed a minimax representation of ρ. The equivalence between the proportional factor reduction story and the minimax story also plays a role in this paper. The same idea has also been used in the Operations Research (OR)-oriented literature on productivity analysis with considerable success. In the Data Envelopment Analysis (DEA) approach introduced by Charnes et al. (1978, 1979) (see Bogetoft and Otto (2011) for a recent review), inefficiencies relative to an empirical production frontier are measured by the maximal possible proportional reduction in inputs that are able to produce the same levels of outputs. The minimax interpretation of the inefficiency measure has also been discussed by Banker (1980) and Banker et al. (1989). This paper also has links to Bogetoft (2000), who investigated optimal contracting in multi-dimensional settings under more restrictive assumptions about the communications that can occur and without investigating the role of rationing.

Another stream of the literature worth noting in connection with the dynamic extension concerns dynamic pricing and learning. This topic has been studied by different scientific communities, including operations research, management science marketing, economics, econometrics, and computer science, cf. Boer (2015). At the general level, it is the combined application of two research fields: (1) statistical learning and in particular the estimation of demand functions, and (2) price optimization. Of course, the general idea of this literature is equally relevant to dynamic production planning and the learning of cost functions, as we investigate in this paper. The literature is quite technical in that it uses advanced methods from stochastic decision and control theory, for example, and the derived results are often linked specifically to the series of model assumption made. However, the general theme is the trade-offs between exploration (the estimation) and exploitation (making use of the info acquired). To improve learning and thereby support future revenue maximization, it is generally optimal to forgo some short-run profits to improve long-run prospects, much like the trade-offs in investment-consumptions problems. The dynamic learning literature is largely non-strategic. The idea is to extract information from "nature". However, there

is also more recent literature in which information is extracted from strategic agents. Reviews of this literature are given by Shen and Su (2007) and Gonsch et al. (2013). These studies typically have a game-theoretic flavor because both the firm and the strategic customers have a decision problem to solve, with contradicting interests. In our dynamic extensions, we encounter similar problems.

3.   The Setting

To model multiple dimensional production decisions in the simplest way possible, we consider a principal who delegates the production of two outputs to an agent. Let $y^o \in \mathbb{R}^2_+$ be the outputs produced in a previous period, and let $x^o \in \mathbb{R}_+$ be the corresponding total cost incurred. The costs x^o may not be the minimal cost necessary to produce y^o because the agent may have been inefficient or he may have earned information rents in the past. We assume that the principal knows

$$(x^o, y^o)$$

An alternative interpretation would be that (x^o, y^o) is a first project proposal from the agent in a budgeting or contract negotiation procedure and that we model the bargaining over project alterations (y^o) and payments (x^o).

   We assume that the underlying cost of producing $y \in \mathbb{R}^2_o$ is cy for some $c \in \mathbb{R}^2_+$. That is, we stick to the traditional assumption in cost accounting and assume a simple linear, additively separable cost function

$$c(y) = c_1 y_1 + c_2 y_2$$

The agent knows the unit cost vector c, but the principal does not. A priori, the principal only knows that $c \in \mathbb{R}^2_+$. Combined with the (x^o, y^o) information, he knows that[6]

$$c \in C^o = \{c \in \mathbb{R}^2_+ \mid c_1 y_1^o + c_2 y_2^o \leq x^o\}$$

Let the probability distribution $p(.)$ on C^o represent the principal's beliefs regarding the likelihood of the different cost functions.

   To model the demand side in the simplest way possible, we let the future value of production y be given by

$$V(y) = r \cdot 1_Y(y)$$

where $1_Y(.)$ is the indicator function giving 1 on Y and 0 otherwise. Hence, the value $V(y)$ is r if production y belongs to $Y \subseteq \mathbb{R}^2_+$ and 0 otherwise. A decline in demand can then be modelled simply as a decline in r or in Y. We can think of Y as the set of satisfactory outputs. We may naturally

---

[6] We discuss extensions of the analysis to more complex sets of possible units cost vectors C later.

assume that Y is freely disposable: $y \in Y, y' \geq y \implies y' \in Y$. The demand function is, of course, very simple, and we could easily introduce a more intricate impact of the scale of operations[7]. We shall not do so to a large extent in the main text, however, because our aim is to focus on scope effects as opposed to scale effects. More general demand functions are discussed in the Appendix.

The aim of the principal is to maximize profit. Equivalently, for any given revenue level r, the aim of the principal is to minimize the costs of inducing the agent to accept employment and select a production plan in Y.

Conversely, the agent seeks to maximize the (expected) surplus

$$b - cy$$

where y is the implemented production plan and $b \in \mathbb{R}$ is the budget or payment from the principal to the agent. The agent's reservation payoff is 0.

We further assume that the principal can observe and verify the output[8], y . The principal, however, cannot observe the (minimal) costs of producing y, i.e., cy. One way to think of this aspect is to assume that the actual, observable cost is $x = cy + s$ where $s \geq 0$ is a *slack* introduced by the agent. We assume that slack is as valuable to the agent as profit is. The surplus value to the agent is then[9] $b - (cy + s) + s = b - cy$.

An alternative interpretation of the initial information (x^o, y^o) is therefore that the agent - or another agent with the same cost structure - in the previous period accepted to produce y^o for a compensation of x^o. Historically, there may be slack in the compensation, which is why the initial information is $cy^o \leq x^o$ and not $cy^o = x^o$.

To formulate the principal's problem in general terms, we assume that the agent submits a full, possibly manipulated description of the unit cost vector $c \in C^o$ in response to the principal's offer of a production plan y and reimbursement plan b. A contract in this model is therefore a specification of the cost contingent production plan, y, and payment plan, b:

$$y[\cdot] : C^o \to Y \text{ and } b[\cdot] : C^o \to \mathbb{R}_o$$

[7] For example, we could work with a homothetic demand function

$$V(y) = \min\{\alpha \mid \alpha y \in Y\}$$

[8] Hence, there is no moral hazard involved in the implementation of a given production plan, and to save on notation, we act as though a chosen plan is simply implemented centrally; i.e., we leave out specifications of the penalties associated with deviations by the agent from a prescribed plan.

[9] It is simple to extend the analysis to situations where the agent's value of slack s is only a fraction γ of the value of profit $b - (cy + s)$; cf. Bogetoft (2000) for some initial results.

By revelation arguments, the use of such production and payment plans allows the best possible outcome in any organization that respects the privacy of the cost information. This is the second-best benchmark.

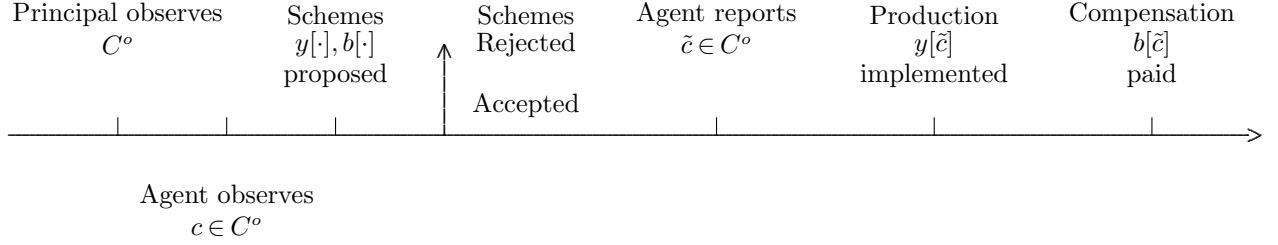The timing in the model is summarized in Fig. 1 below.

Figure 1   Timeline of Base Model

The principal's contract design problem can be formalized as

$$\max_{C,y[\cdot],b[\cdot]} \left[\sum_{c \in C}(V(y[c]) - b[c])p(c \mid C)\right] p(C)$$

$$\begin{array}{lll} \text{s.t.} & b[c] - cy[c] \geq 0 & \forall c \in C \quad \text{(IR)}\\ & b[c] - cy[c] \geq b[\tilde{c}] - cy[\tilde{c}] & \forall c, \tilde{c} \in C \ \text{(IC-}c,\tilde{c})\\ & b[c] \in \mathbb{R}, y[c] \in Y & \forall c \in C\\ & C \subseteq C^o & \end{array}$$

This is the principal's problem of determining which cost types C should produce and which production-compensation plans $(y[\cdot], b[\cdot])$ should be used to induce the agent to accept employment (IR) and report his cost truthfully (IC). Types that are not in C are screened out.

In the following sections, we investigate the solution to this problem. We do so in several steps. First, we consider the first-best solution. Next, we consider the solution when rationing is not an option, i.e., $C = C^o$. Finally, we consider optimal solutions when rationing is possible, $C \subseteq C^o$.

It is convenient to establish a bit more notation. Let the largest possible costs of a given output vector y that is consistent with information C about the possible unit cost vectors be denoted as c^{max}, i.e.,

$$c^{max}(y \mid C) = \max_{c \in C} cy$$

This is the maximal or worst case cost when the unit cost vector belongs to C. In particular, we have

$$c^{max}(y \mid C^o) = \max_{c} \ cy \qquad = \min_{\rho} \ \rho x^o = \quad x^o \max\{\tfrac{y_1}{y_1^o}, \tfrac{y_2}{y_2^o}\}$$
$$\text{s.t.} \quad cy^o \leq x^o \quad \text{s.t.} \ \rho y^o \geq y,$$

where we come from the first LP problem to the second LP problem by dualizing.

In the benchmarking literature, the last formulation defines the constant returns to scale DEA cost function based on observation (x^o, y^o), cf. Bogetoft and Otto (2011)[10].

4. First Best

The first-best solution is easy to find. Given perfect information about the unit costs, $c \in C$, the principal finds the least costly point $\tilde{y}$ in Y, i.e., the point of tangency between the isocost line and the south-west boundary of Y, as illustrated in Fig. 2 below.
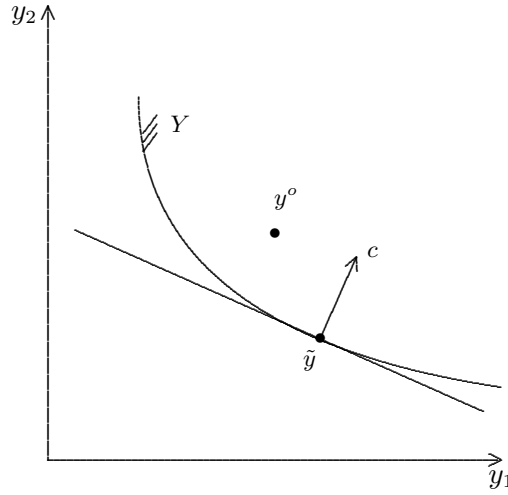
Figure 2   First best $\tilde{y}[c]$

If the cost of the tangency point, $c\tilde{y}$, is less than the value r, the production of y^* takes place; if not, no production is undertaken. Formally, we let

$$\tilde{y}[c] \in \arg\min_{y \in Y} cy$$

and the first-best solution is given by

$$y^{FB}[c] = \tilde{y}[c] \text{ and } b^{FB}[c] = cy^{FB}[c]$$

when $c\tilde{y}[c] \leq r$ and $y^{FB}[c] = 0$ and $b^{FB}[c] = 0$ otherwise. We can say that the principal implements the technically and allocatively efficient production plan for all possible cost types that are not rationed away. The cost types where the cheapest way to provide a satisfactory production plan leads to costs above the values they generate, r, are rationed away.

---

[10] One can use this analogy to extend our analysis. If more than one historical observation were available, we would use the DEA function based on multiple observations, and if there were also a fixed cost in the underlying true cost function, $\beta + cy$, where $\beta \geq 0$, we would obtain the so-called increasing returns to scale DEA cost function; further, cf. Bogetoft (1997) for an examination of these relationships.

5. No Rationing

5.1. Limited Communication

To gain intuition, and because it is of interest on its own, we initially assume that the agent cannot communicate detailed cost information to the principal. In this case, production and reimbursement plans must be unconditional

$$y[c] \equiv y \text{ and } b[c] \equiv b$$

In the limited communication set-up, the contract design problem is simplified. Because we can have only one payment level b and one production plan y, the IC constraints are trivially fulfilled, and we obtain

$$
\max_{C, y[\cdot], b[\cdot]} \left[\sum_{c \in C} (r - b) p(c \,|\, C) \right] p(C)
$$
$$
\text{s.t.} \quad b - cy \geq 0 \qquad\qquad \forall c \in C \text{ (IR)}
$$
$$
b \in \mathbb{R}, y \in Y
$$
$$
C \subseteq C^o
$$

and because we do not allow rationing, when we ignore rationing, $C = C^o$, the contract problem is simply

$$
\min_{y, b} b
$$
$$
\text{s.t.} \quad b - cy \geq 0 \quad \forall c \in C^o \text{ (IR)}
$$
$$
y \in Y, b \in \mathbb{R}_o
$$

Rewriting, the problem is

$$
\min_{y \in Y} \min_{b \geq cy, \forall c \in C^o} b = \min_{y \in Y} \max_{c \in C^o} cy = \min_{y \in Y} c^{max}(y \,|\, C^o) = \min_{y \in Y} \min_{\rho y^o \geq y} \rho x^o
$$

where the second equality uses the definition of c^{max}, and the third equality follows from the dual formulation of c^{max}. Therefore, we see that an optimal solution is to *proportionally downscale (upscale) y^o as much (little) as possible such that the rescaled plan belongs to Y*. We also summarize this concept in a proposition.

Proposition 1. *With no communication and no rationing, a solution to the principal's contracting problem is*

$$y = y^P = \rho^* y^o \text{ and } b = c^{max}(y^P \,|\, C^o) = \rho^* x^o$$

where $\rho^* = \min\{\rho \,|\, \rho y^o \in Y\}$ is the largest possible reduction or smallest possible expansion of y^o in Y. ☐

Proposition 1 demonstrates that the demand substitution possibilities, or the preferences of the principal, as modeled here by Y, need not influence the mix decision. *The asymmetric information*

leads to mix stickiness. It forces the principal to select the minimal acceptable production plan, which is proportional to the production plan of the previous period.

An illustration of the proportional adjustment solution is provided in Fig. 3. In the left panel, the value of the second output y_2 is low, whereas in the right panel, the value of the second output y_2 is high. Despite these differences in demand, the optimal solution is the same, namely, a proportional adjustment of the historical observation y^o down to the point y^P.
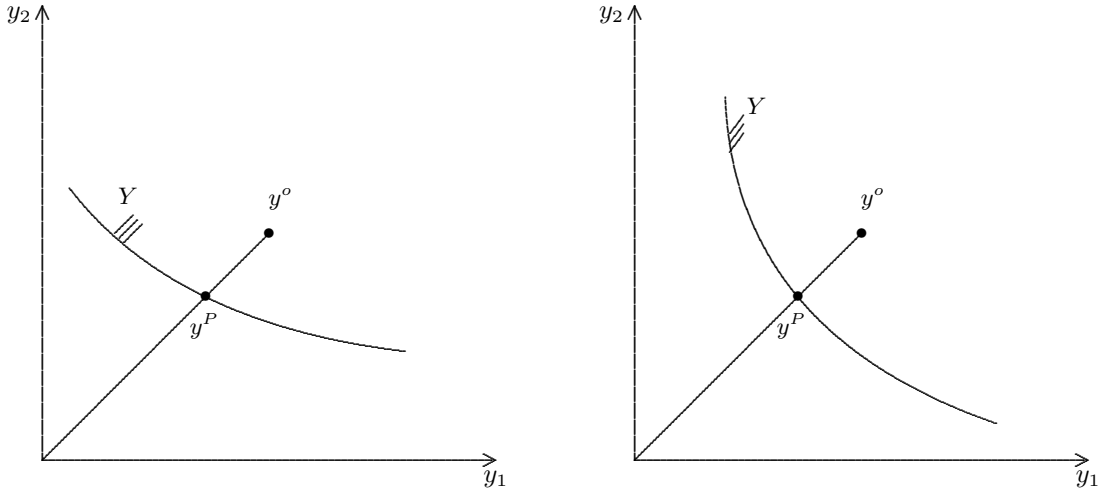
Figure 3   Proportional adjustments under different demand substitution possibilities

The optimality of the proportional adjustment policy is derived from the principal's lack of information about the cost function. Inspecting the proof, we see that the principal commands y^P because those are the outputs that are cheapest under the least favorable cost structure, consistent with the a priori information.

A simple reinterpretation of the proposition is also suggested by the alternative reformulation of the contract design problem.

$$\min_{y \in Y} \max_{c \in C^o} cy$$

A reinterpretation of the proposition is therefore as a bargaining problem: the principal bargains with an agent with considerable power. The agent can threaten to have any cost structure that is consistent with the a priori information C^o. Given any production plan $y \in Y$, the agent will of course claim to have the structure that maximizes the costs of producing y. The principal therefore picks $y \in Y$ to minimize the maximal costs of producing y; i.e., he picks y to solve the minimax program above.

Note that the agent's bargaining power here is embedded in the (IR) constraints, i.e., in the principal's desire to have acceptable production whatever the underlying cost structure is. In other

word, the agent gets bargaining power when the principal will not ration. This suggests that rationing may be useful in cases where the revenue r is not too large.

Proposition 1 is surprisingly simple. In the first-best world, the cost-minimal-production plan will intimately depend on the marginal costs, which must be balanced with marginal rates of substitution in the demand, as we illustrated in Fig. 2. One might expect a similar relationship between the expected marginal costs and the rates of substitution in demand in a second-best world without communication. However, because no rationing is possible, such considerations do not play any role. The principal will simply require all types to pick the proportional output and pay the agent the worst-case costs. In a one-dimensional context, this approach is similar to asking all types to produce the first-best outcome of the least productive type.

Proposition 1 can easily be generalized to cases with more than two outputs, and it does not assume convexity of Y[11].

5.1.1. Numerical example To illustrate the general formulations, it can be useful to look at a numerical example. To do so, we assume that

$$x^o = 1 \text{ and } y^o = (1,1)$$

This is without loss of generality as long as the original observation fulfills $x^o > 0, y_1^o > 0$, and $y_2^o > 0$ because in such cases, we can always simply change the measurement units for costs and products; i.e., we can simply measure inputs in x^o units and outputs in y_1^o and y_2^o units.

To begin, let us assume that future demand is $y^* = (2,0)$, i.e., the principal is only interested in the first output and he wants to double it. Formally, we assume that

$$Y = y^* + \mathbb{R}_+^2 \text{ and } r = 1$$

In the no-rationing case, the principal will therefore have to pay the agent

$$b = c^{max}(cy^* \mid c \in C^o) = 2$$

because the worst possible type when we have to produce $(2,0)$ is $c = (1,0)$ with corresponding cost 2. Note also, as suggested by Proposition 1, that in this case the principal might just as well ask the agent to produce $y^P = (2,2)$ because the worst case cost of $(2,2)$ is also 2. The situation is illustrated in Figure 4 below.

---

[11] Additionally, the proposition is valid if the set of cost functions are not necessarily linear. For example, it can be generalized to settings where the possible cost functions are the set of convex functions by following the analysis in Bogetoft (1997).
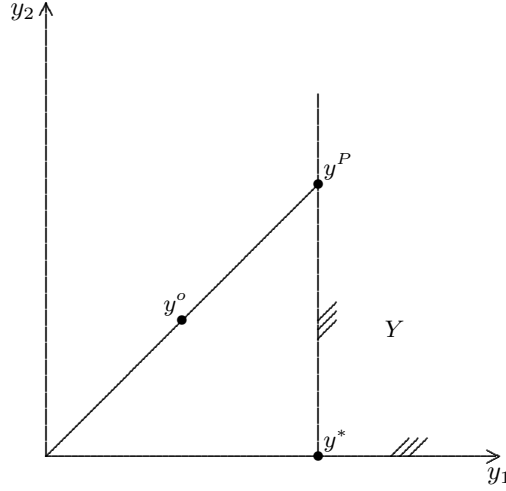
14

Figure 4   Proportional adjustments in numerical example

5.2.   Full Communication

Let us now analyze the non-rationing case in the most general situation of full communication possibilities. Because we want production for all possible cost types, the principal's problem of maximizing surplus becomes a problem of expected cost minimization.

$$
\begin{aligned}
&\min_{y[\cdot],b[\cdot]} \; \sum_{c\in C^o} b[c]p(c) \\
&\text{s.t.} \quad b[c]-cy[c]\geq 0 && \forall c\in C^o \quad (IR) \\
&\qquad\; b[c]-cy[c]\geq b[\tilde{c}]-cy[\tilde{c}] && \forall c,\tilde{c}\in C^o \;\; (\text{IC-}c,\tilde{c}) \\
&\qquad\; b[c]\in\mathbb{R}, y[c]\in Y && \forall c\in C^o
\end{aligned}
$$

It may be difficult to analytically solve the general full communication no-rationing problem. It is worthwhile to note, however, that we can approximate an optimal solution by solving discrete versions of the full communication contract design problem. In particular, if we substitute the cost type set C^o with a discrete subset $\bar{C}\subset C^o$, the contract design problem reduces to a simple *linear programming* problem.

The challenge facing the principal may be clearer if we reformulate it in terms of surplus or slack payments

$$
s[c]=b[c]-cy[c]
$$

on the one hand and direct production costs $cy[c]$ on the other hand. This leads to a full communication contracts design problem

$$
\begin{aligned}
&\min_{y[\cdot],s[\cdot]} \; \sum_{c\in C^o} cy[c]p(c)+\sum_{c\in C^o} s[c]p(c) \\
&\text{s.t.} \quad s[c]\geq 0 && \forall c\in C^o \quad (IR) \\
&\qquad\; s[c]\geq s(\tilde{c})+y[\tilde{c}](\tilde{c}-c) && \forall c,\tilde{c}\in C^o \;\; (IC) \\
&\qquad\; s[c]\in\mathbb{R}, y[c]\in Y && \forall c\in C^o
\end{aligned}
$$

15

The program shows the basic trade-off facing the principal. He wants to minimize production costs, i.e., $cy[c]$, when the costs are c. If this was his only concern, it would lead to the first-best solution. However, he must also pay the motivation cost or the cost of asymmetric information which is $s[c]$ for type c. Unfortunately, these cost elements are interrelated because the choice of production plan affects the ability to choose motivation payments, as emphasized by the (IC) constraints.

The (IC) constraints suggest that the more we vary the production plan, $y[c]$, the larger the agents ability to claim high motivation costs. Casually speaking, when we vary the production, the agent has more opportunities to shop around and find attractive cost types to imitate. From the point of view of motivation, therefore, we may not want to adapt production to the primary production costs; i.e., the outcome may not be allocatively efficient.

5.2.1. Numerical example (ctd) To analyze our numerical example in the case of full communication, we may assume that there are three special output profiles, namely, status quo $y^o = (1,1)$, specialization in output 1, $y^1 = (2,0)$ and specialization in output 2, $y^2 = (0,2)$, and that the set of acceptable output profiles is

$$Y = conv\{(2,0),(1,1),(0,2)\} + \mathbb{R}^2_+$$

That is, Y is the (smallest) convex subset of $\mathbb{R}^2$ that contains y^o, y^1 and y^2 and has the usual property of free disposability or more is weakly better than less, i.e., if y is acceptable, so is any y' with $y'_1 \geq y^*_1$ and $y'_2 \geq y^*_2$.

The first-best benchmark solution in this case is illustrated in the left panel of Fig. 5, and the second-best solution without rationing is shown in the right panel. We see that the second-best solution is biased toward the historical (proportional) outcome.
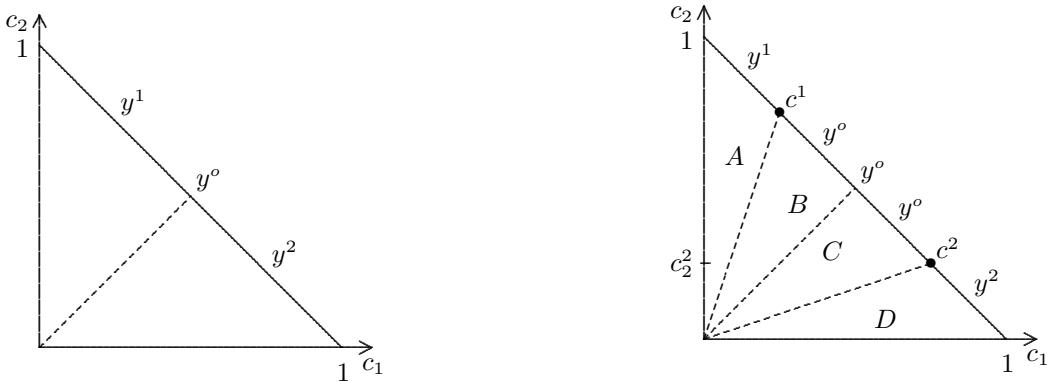

Figure 5   FB and SB outcomes without rationing

16

In the FB world, the historical outcome is only optimal for $c_1 = c_2$. As soon as one of the unit costs is larger than the other, it is better to change the output bundle such that we only produce the product with the smallest unit cost.

The second-best solution, however, extends the set of cost profiles, leading to the historical outcome to a subset $B \cup C$. To evaluate exactly how much to extend the use of the historical outcome, we need to make specific distributional assumptions. We do this below. Conceptually, however, the outcome is based on a common logic irrespective of the underlying probability distribution: *The optimal solution is biased toward the historical outcome*. This is mix stickiness. The principal does not adjust to the allocatively efficient solution of (almost) always choosing a reorientation of the production plan. The adaption of production to the marginal cost is sub-optimal. The principal asks a larger share of the agent types to imitate the worst-case type that produces y^o. The principal forgoes specializing to $y = (2,0)$ or $y = (0,2)$ for cost types in $B \cup C$ to save on the information rents that the cost types in A and D can extract. Note that the cost types c^1 on the border of A and B is the most costly type that is asked to produce y^1. This means that all types in A must be paid $c^1 y^1 = 2c_1^1$. Therefore, the smaller we make B, the more we must pay the types in A. Hence, we stick to the original mix to save on the information rents that the extreme types in A can earn. Likewise, we let types in C produce the historical mix because hereby, the types in D will receive a smaller payment, $c^2 y^2 = 2c_2^2$.

Our multi-dimensional case is therefore different from the traditional adverse selection model with one-dimensional types. The single dimensional rule of having "no distortion at the top" and "downward distortion at the bottom" is not directly applicable here since we have no scale issue and no exogenous ordering of the types. In our case, as it turns out, the bad types are located in the middle of the type space as opposed to on one side of the cost space in the one-dimensional case. The worst type, $c = (0.5, 0.5)$, is therefore not biased, and neither are the good low-cost types in A or D. It is the middle types in B and C that are biased. The intuition is, however, exactly like in the single dimensional case. We bias the production bundles for the middle types in B and C in order to reduce the information rents to the low-costs types in A or D [12].

The exact size of the different areas depends on the probability distribution of the cost types. To give an example, we may assume in the numerical example that the unit cost vectors are uniformly distributed on C^o. This assumption simplifies the calculations, and it allows us to investigate our examples without the compounding effects of the hazard rate, which we know from the one-dimensional case plays an important role when determining how aggressively to ration.

---

[12] Moreover, when we later introduce rationing, we will see that we actually ration more strongly in the middle areas, cf. Fig. 2 in order to make it less attractive to imitate the bad types. In this case, we can therefore say that the "no distortion at the top" and "downward distortion at the bottom" rules still apply.

Now, to determine the optimal split between C and D, we note that all types in C must be paid 1 and all types in D must be paid $2c_2^2$ because they are asked to produce $(0,2)$. To minimize to expected costs when $c_2 \leq c_1$, the principal therefore solves the following problem in the uniform case:

$$\min_{c_2^2} 1(\frac{1}{2} - c_2^2) + 2c_2^2(c_2^2)$$

where the first parenthesis is the probability of region C and the last is the probability of region D. It is easy to calculate these probabilities as 2 times the areas of the two triangles. Now, it follows that the optimal solution to this cost minimization problem is $c_2^2 = \frac{1}{4}$. By symmetry, we therefore have that the historical outcome in the uniform case is optimal for all c, such that $3 \geq \frac{c_2}{c_1} \geq \frac{1}{3}$. In other words, even with a full communication of the cost types, the original mix sticks with probability 50 % in a second-best solution without rationing and with uniformly distributed cost types.

6.   Rationing

We now introduce the possibility of rationing. Technically, this means that we allow the principal to acquire output $(0,0)$ at the cost of 0 and with associated revenue 0. Rationing may be attractive by limiting the information rents that an agent can extract by having particularly non-favorable unit costs, but it comes at the cost of giving up production in some cases.

Whether rationing is feasible depends on the context. A firm aiming to maximize its expected profit may have no problem forgoing production in some cases. A utility regulator, however, may be reluctant to drive some firms into bankruptcy to lower payments to other firms. Similarly, a non-profit organization may not want to maximize profit, and the demand may instead be determined by organizational preferences and donor restrictions.

Moreover, rationing may not be attractive even when it is feasible. This has to do with the profitability of the production plan, y. In a lucrative market where r is sufficiently high, rationing is not attractive because it is costly to forgo any production. In a less profitable setting where r is lower, it may be attractive to ration. As we will see below, there is a fundamental difference between the measures taken to lower the information rents in high r and low r settings.

6.1.   Limited Communication

With limited communication and the possibility of rationing, the contract design problem is, as explained above

$$\max_{C,y,b} \left[\sum_{c \in C} (r - b)p(c \,|\, C) \right] p(C)$$
$$\text{s.t.} \quad b - cy \geq 0 \quad \forall c \in C \qquad \text{(IR)}$$
$$b \in \mathbb{R}, y \in Y$$
$$C \subseteq C^o$$

18

We can rewrite this problem as

$$\max_{y \in Y} \max_{C \subseteq C^o} \max_{b \geq cy \, \forall c \in C} [r - b] \mathrm{Prob}(C)$$

and therefore as

$$\max_{y \in Y} \max_{C \subseteq C^o} [r - c^{max}(y \mid C)] \mathrm{Prob}(C)$$

For a specific production plan, $y = y^*$, the inner optimization problem becomes

$$\max_{C \subseteq C^o} [r - c^{max}(y^* \mid C)] \mathrm{Prob}(C)$$

We can simplify this equation further by noting that it is not all subsets C that we need to optimize over. It does not serve to ration away cost types with $cy^* < r$ unless the payment to the agent falls for the remaining types. Thus, given the structure of c^{max}, we see that that payment only falls to the extent that the largest possible value of cy^* falls. In other words, optimal rationing requires us to eliminate a subset $\{c \mid cy^* \geq \alpha\}$, as illustrated in Fig. 6 below. We see that the intended production plan y^* is the gradient of *the rationing hyperplane*. The more interested we are in the production of the first output, the more eager we are to eliminate unit cost vectors with high c_1, and vice versa.
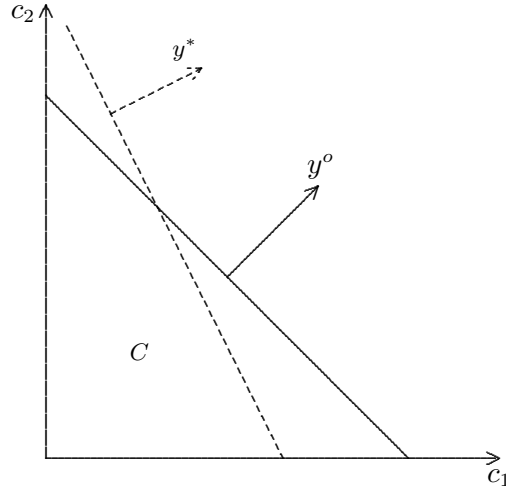

Figure 6   Optimal Rationing

Rationing using such hyperplanes means that we can express the cost profiles leading to production as

$$C = \{c \in C^o \mid cy^* \leq \alpha\}$$

and we can therefore reformulate the inner problem as

$$\max_{\alpha \in \mathbb{R}} (r - \alpha) \mathrm{Prob}\{c \in C^o \mid cy^* \leq \alpha\}$$

19

If we let $F(\cdot \,|\, y^*)$ denote the cumulative distribution of cy^*, i.e.,

$$F(\alpha \,|\, y^*) = \mathrm{Prob}\{c \in C^o \,|\, cy^* \leq \alpha\}$$

the principal's inner problem can be formulated as

$$\max_{\alpha \in \mathbb{R}} (r - \alpha) F(\alpha \,|\, y^*)$$

with the familiar first-order condition for an optimal rationing payment

$$\alpha(y^*) = r - \frac{F(\alpha \,|\, y^*)}{f(\alpha \,|\, y^*)}$$

where $f(\cdot)$ is the corresponding density function, $f(\alpha \,|\, y^*) = F'_\alpha(\alpha \,|\, y^*)$. We see that the principal foregoes production when the agent's cost is close to the revenue r, although such production is clearly profitable. He does so to lower the payment to the less costly agent types.

With limited communication, a principal considering rationing must solve the following contract design problem:

$$\max_{y \in Y} (r - \alpha(y)) F(\alpha \,|\, y)$$

or to write it out in more detail,

$$\max_{y \in Y, \alpha \in \mathbb{R}} (r - \alpha) \mathrm{Prob}\{c \in C^o \,|\, cy \leq \alpha\}$$

because for every output profile, he must consider the associated optimal rationing. We summarize this argument in a Proposition.

Proposition 2. *With limited communication and the possibility of rationing, an optimal solution to the principal's contract design problem is found by solving*

$$\max_{y \in Y} \max_{\alpha} (r - \alpha) F(\alpha \,|\, y)$$

and making the following take-it-or-leave-it offer to the agent

$$y[c] = y^* \quad and \quad b[c] = \min\{\alpha^*, r\} \;\; \forall c \in C = \{c \in C^o \,|\, cy \leq \alpha^*\}$$

where α^* is a solution to

$$\alpha = r - \frac{F(\alpha \,|\, y^*)}{f(\alpha \,|\, y^*)}$$

with F and f the distribution and density function of cy^*, respectively.

$\square$

One way to look at this problem is as a choice between different projects corresponding to different output vectors $y \in Y$, all of which have uncertain costs. A similar problem of choosing

between mutually exclusive investment opportunities was studied in Antle et al. (1999). It is showed here that the optimal strategy is a handicapping scheme, where for each investment opportunity, the principal specifies his maximal willingness to pay and ask the agent to choose which project, if any, to implement. The agent will then choose the project with the largest slack, i.e., the largest difference between the actual cost and the principal's willingness to pay. We return to this analogy when we consider the general case with full communication possibilities below.

For now, let us note that the solution to the above problem can be constructed from the following equivalent problem:

$$\max_{y \in Y} \max_{\alpha \in \mathbb{R}} (r - \alpha) \mathrm{Prob}\{c \in C^o \mid cy \leq \alpha\}$$

Hence, we can solve the problem in two steps. First, we find the optimal rationing payment $\alpha(y)$ for all possible $y \in Y$. That is, we solve the inner optimization problem. Based on this result, we can find the expected profit to the principal should he choose to implement the different production mixes; i.e., we solve the outer optimization problem. In this stage, the optimal y can therefore be chosen. We illustrate this concept below.

6.1.1. Numerical example (ctd)

Consider again the uniform case where

$$Y = y^* + \mathbb{R}_+^2 \text{ and } r = 1$$

Initially, we assume that $y^* = (2, 0)$

With rationing, the principal offers α, and the agent will accept to produce $y^* = (2, 0)$ as long as $2c_1 \leq \alpha$. The probability of this is

$$F(\alpha) = \mathrm{Prob}\{2c_1 \leq \alpha\} = \mathrm{Prob}\{c_1 \leq \frac{\alpha}{2}\} = 1 - (1 - \frac{\alpha}{2})^2 = \alpha - \frac{\alpha^2}{4}$$

The corresponding density is

$$f(\alpha) = 1 - \frac{\alpha}{2} \qquad \alpha \leq 2$$

Now, inserting into the first-order condition

$$\alpha(y^*) = r - \frac{F(\alpha \mid y^*)}{f(\alpha \mid y^*)}$$

we obtain

$$\alpha = 1 - \frac{\alpha - \frac{\alpha^2}{4}}{1 - \frac{\alpha}{2}}$$

and by solving for α, we find that the optimal payment is $\alpha = 0.46$, leading to a probability of acceptance $\mathrm{Prob}(C^*) = 0.41$ and an expected profit of 0.22.

We can generalize the above analysis to cover an arbitrary output vector y^*. The analysis is a little more cumbersome for an arbitrary y^* but the general principle is the same as in the specific case of $y^* = (0, 2)$. A detailed analysis is shown in the appendix.

Table 1 below shows the outcome for selected values of the parameters in the case of $y_1^* \geq y_2^*$ for two values of r, 1 and 2.

When $r = 1$ and $y^* = (1, 1)$, the principal is willing to pay $\frac{2}{3} \simeq 0.67$. The probability of production in this case is $\frac{4}{9} \simeq 0.44$. It is interesting to compare the result to a situation where there is only one product with value 1 and uniform costs on $[0, 1]$; in this case, the optimal cut off is found by solving $\max_\alpha (1 - \alpha)\alpha$, leading to an optimal value of $\alpha = \frac{1}{2}$. The willingness to pay more in the two-dimensional case compared with the one-dimensional case is explained by the fact that large values for total cost are more likely in the two-dimensional case. Rationing is therefore more expensive because the principal will lose more production for a given cost decrease than in the one-dimensional case. Indeed, in terms of probabilities, he also rations away more production in the two-dimensional case, where the probability of production is $\frac{4}{9}$ compared with $\frac{1}{2}$ in the one-dimensional case.

We see also that for $r = 1$, it is more profitable to implement the specialized solution $y^* = (2, 0)$ than the historical plan $y^* = (1, 1)$. This result changes, however, in the $r = 2$ case. In fact, detailed analysis of the intermediate cases shows that the change happens for $r = 1.4$. Hence, for small values of r, it is more profitable to implement the specialized solution, while for larger values of r, it is more profitable to implement the historical mix[13].

r	y_1^*	y_2^*	α	$P(C^*)$	Π^{max}	Π^{NoRat}
1	2	0	0.46	0.41	0.22	-1.00
1	1.8	0.2	0.52	0.43	0.21	-0.80
1	1.6	0.4	0.58	0.45	0.19	-0.60
1	1.4	0.6	0.64	0.49	0.16	-0.40
1	1.2	0.8	0.67	0.46	0.15	-0.20
1	1.00	1.00	0.67	0.44	0.15	0.00
2	2.0	0.0	0.85	0.67	0.77	0.00
2	1.8	0.2	0.88	0.71	0.79	0.20
2	1.6	0.4	0.92	0.76	0.82	0.40
2	1.4	0.6	0.96	0.82	0.86	0.60
2	1.2	0.8	0.99	0.90	0.92	0.80
2	1.00	1.00	1.00	1.00	1.00	1.00

Table 1 **Optimal rationing when $y_1^* \geq y_2^*$**

The numerical example shows that

---

[13] In this example, we have considered the simple case of only two products because the result can be calculated directly using basic calculus. However, generalizations to more dimensions and to cases where we have more observations are a priori relatively simple because one can use numerical codes such as the so-called "qhull" program to find the volume of any convex hull of a set of points in n dimensional Euclidean space.

- The larger the value r is, the less we ration.

- The more similar the desired product mix is to the original mix, the less we ration.

- The more similar the desired product mix is to the original mix, the larger the profit, unless we need to ration extensively; in the latter case, the more extreme mixes are the cheapest.

- Rationing is superior to no rationing unless the value r is high and the requested mix is close to the original mix.

The first observation is in accordance with the simple case of rationing in one dimension. When we lose a large value r when no production takes place, it becomes less attractive to ration.

The second observation follows from the same logic: when the desired output profile is more extreme, we can reduce the payment more aggressively without reducing the probability of production too dramatically. Hence, from a rationing perspective, the more extreme output profiles y^* are attractive. More precisely, the distribution of cy^* is more left-skewed when y^* is less extreme and more right-skewed when y^* is more extreme compared with y^o. It is expensive to ration in a left-skewed distribution because we lose substantial production for just a small decrease in payment. In a right-skewed distribution, the production loss is smaller compared with the reduction in payment. It is interesting to note that when the value of the production, r, is small, such that we are forced to ration more aggressively, the extreme production plan is better than the less extreme production plan.

The third observation shows that although the extreme output profiles allow us to use the rationing instrument more effectively, the highest profits are generated when the desired future mix is more similar to the historical mix. We can understand this result by noting that it is the cost of the original output profile that we know the best.

To illustrate the latter two observations, we can, for example, compare the distribution of cy^* when $y^* = (1,1)$ and $y^* = (2,0)$. In the first case, we find that the cumulative distribution function of cy^* is $F(z) = z^2$, such that the density is $f(z) = 2z$ on $[0,1]$. In the latter case, where $y^* = (2,0)$, the cumulative distribution function of cy^* is $F(z) = 1 - (1 - \frac{z}{2})^2$, such that the density is $f(z) = 1 - \frac{z}{2}$ on $[0,2]$. The densities are illustrated in Fig. 7.

Implementing an extreme production profile is similar to betting that the most reduced output is the most costly. In other words, by using the old observation, we do not know which of the two products is the cheapest to produce, but when we have to ration sufficiently harshly, we simply bet against one of the products. Our results therefore suggest that *the more competitive the environment* is; i.e., the less the profit r is, *the more the principal must take a risk by going extreme rather than imitating past behavior*. This result also suggests that a lack of competition may stall product development and specialization. Note, however, that this delay is not driven by competition for consumers; it is driven solely by the asymmetric cost information.
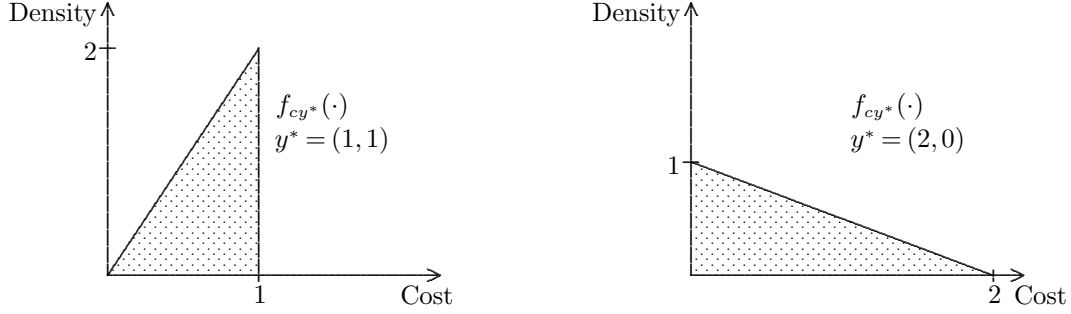
Figure 7   Densities of cy^* in uniform example

Another way to illustrate this result and to gain insight into the rationing advantage of more extreme types is to examine the probability of production $F(\alpha \mid y^*)$ as a function of the payment threshold α. In Fig. 8, we have depicted the probability of production for two possible product mixes, namely, $y^* = (1,1)$ and $y^* = (2,0)$. We see that the original observation $(1,1)$ has advantages when we need a high probability of acceptance, while the extreme profile is advantageous when we need more aggressive rationing.
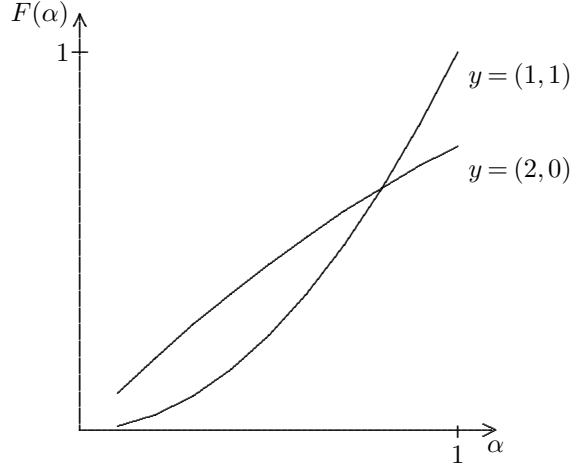
Figure 8   Rationing effect with different mixes

The examples in Table 1 also illustrate the optimal solutions when

$$Y = \{(2,0), (1.8, 0.2), (1.6, 0.4), \ldots, (1,1)\}$$

and we seek a solution to the contract under limited communication. This problem corresponds to a situation with perfect substitution between the products, as long as we have at least 1 unit of the first product. We see that when $r = 1$, it is optimal to choose $\alpha = 0.46$ and implement $y = (2,0)$. If $r = 2$, it is optimal to choose $\alpha = 1$ and implement $y = (1,1)$. Hence, *the need to ration may make the it attractive to adjust the demand mix rather drastically.*

6.2. Full Communication

The case of a general set of acceptable outputs Y and full communication possibilities leads to the following simple generalization of the full communication, no-rationing problem:

$$\min_{y[\cdot],b[\cdot],C} \left(r - \sum_{c \in C^o} b[c]p(c \mid C)\right) p(C)$$

$$\text{s.t.} \quad b[c] - cy[c] \geq 0 \qquad\qquad \forall c \in C \quad \text{(IR)}$$

$$\qquad\quad b[c] - cy[c] \geq b[\tilde{c}] - cy[\tilde{c}] \qquad \forall c, \tilde{c} \in C \ \text{(IC-}c,\tilde{c})$$

$$\qquad\quad b[c] \in \mathbb{R}, y[c] \in Y \qquad\qquad \forall c \in C$$

$$\qquad\quad C \subseteq C^o$$

The interpretation of this program was given in Section 3. In particular, we recall that whatever can be accomplished in any organization can be accomplished in a full communication set-up where the principal is able to commit and the agent is given incentives to tell the truth.

As in the case of no rationing, it is difficult to derive general analytical solutions for this problem. In the next section, however, we solve the problem for a more specific Y set. This provides useful insight into the forces that determine the optimal rationing strategy.

6.2.1. Numerical example (ctd) Let us now consider the optimal solution with rationing and full communication. As above, we assume that there are three special output profiles, namely, status quo $y^o = (1,1)$, specialization in output 1, $y^1 = (2,0)$ and specialization in output 2, $y^2 = (2,0)$, and that

$$Y = conv\{(2,0),(1,1),(0,2)\} + \mathbb{R}^2_+$$

The rationing framework is illustrated in Fig. 9. If the agent reports a cost in A, B or C he is instructed to implement y^2, y^o or y^1, respectively. The northeast boundary of the sets A, B and C contains the cost types being paid exactly the cost, i.e., the types making no rents. To understand this result, note that if any rents were earned here, points northeast but close to the frontier could make a positive rent by understating their costs, which conflicts with the IR constraints. To understand the shape of the zero rents lines, simply note that the cost of producing (0,2) does not depend on c_1. Hence, the rationing line for set A must be horizontal. The vertical rationing line for set C is determined by similar logic. Finally, the rationing line for set B must have $(1,1)$ as a gradient because it is composed of the cost types for which $c_1 \cdot 1 + c_2 \cdot 1 = $ constant. Note also that the northeast boundaries of A, B and C must be connected because otherwise, cost types close to the corner points could gain from manipulating their reports.

To determine the specific location of the areas, we once again need to make distributional assumptions. Assuming uniform unit costs, there should be symmetry between areas A and C. Let α be the payment to the agent in these areas and let β be his payment in areas B. The choice of an optimal rationing structure therefore boils down to the choice of three parameters, α, β and $k \leq 0.5$,
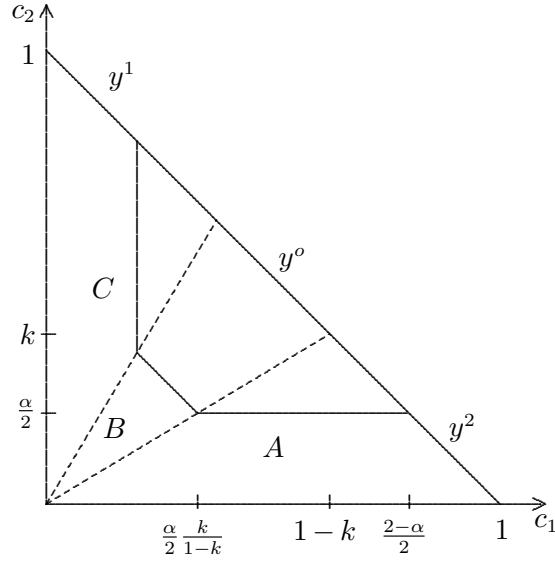
Figure 9   **Optimal SB cut-off in uniform example with rationing**

as also illustrated in Fig. 9. Note that the sufficiency of these three parameters uses the assumption of uniformly distributed unit costs.

The payment β in B should exactly cover the cost of $(1,1)$ for the cost type on the corner where A and B meet. The coordinates of this point are $(\frac{\alpha}{2}\frac{1-k}{k}, \frac{\alpha}{2})$ such that

$$\beta - \frac{\alpha}{2}\frac{1-k}{k} - \frac{\alpha}{2} = 0$$

i.e.,

$$\beta = \frac{\alpha}{2}\frac{1}{k}$$

The profit-maximizing principal will therefore choose α and k to maximize

$$\Pi = (r - \alpha) \cdot 2 \cdot \mathrm{Prob}(A) + (r - \frac{\alpha}{2}\frac{1}{k}) \cdot \mathrm{Prob}(B)$$

We can now derive expressions for the probabilities and based thereon, we can maximize the expected profit. Because the derivations are straightforward but tedious, we provide the details in the appendix. We show here that if we define

$$H = 9 - r - \sqrt{36 - 18r + r^2}$$

the solution can be written as

$$k = 1 - \frac{1}{6}H$$

$$a = \frac{1}{45}(-24 + 36r + 20H - \frac{14}{3}rH - 6H^2 + \frac{2}{9}rH^2 + \frac{7}{9}H^3 - \frac{2}{27}rH^3 - \frac{1}{27}H^4)$$

26

when $r \leq 1.75$ and

$$k = \frac{1}{4}$$
$$a = \frac{1}{2}$$

when $r \geq 1.75$. The latter corresponds to the optimal non-rationing solution from above.

Some numerical results are given in Table 2. Recall that $P(C^*)$ is the probability of production. In the two-dimensional case, this is equal to $2P(A) + P(B)$, cf. Fig.9. We see that for small values of r, where we need to ration more harshly, the optimal strategy is to primarily choose the extreme alternative production plans $(0,2)$ or $(2,0)$, depending on which product the agent declares to be the least costly. When the value of production r increases such that rationing is more expensive, the old production plan $(1,1)$ is picked more often, namely, when the relative costs of the two products do not differ too greatly.

r	H	k	α	$P(A)$	$P(B)$	$P(C^*)$	Π
0.25	3.13	0.48	0.12	0.11	0.00	0.23	0.03
0.50	3.28	0.45	0.23	0.20	0.01	0.41	0.11
0.75	3.45	0.43	0.33	0.27	0.02	0.56	0.23
1.00	3.64	0.39	0.42	0.31	0.06	0.67	0.39
1.25	3.87	0.36	0.48	0.32	0.13	0.77	0.57
1.50	4.15	0.31	0.51	0.30	0.26	0.86	0.77
1.75	4.50	0.25	0.50	0.25	0.50	1.00	1.00
2.00	5.00	0.25	0.50	0.25	0.50	1.00	1.25

Table 2   Optimal rationing in two-dimension uniform case

The optimal solutions for $r = 0.5, 1, 1.5, 2$ are illustrated in Fig. 10 below. *The solutions clearly illustrate the tendency to choose extreme solutions, a reorientation strategy, when harsh rationing is necessary, and the tendency to make excessive use of the historical solution, a reinforcement strategy, when rationing is limited.* Recall that the first-best solution only uses the historical solution for cost types with $c_1 = c_2$, i.e., with probability 0. In all of the second-best solutions, the historical solution is used with a positive probability, namely, the probability $\text{Prob}(B)$. We see that when we do not ration, we use the historical outcome with probability $\frac{1}{2}$. The probability of using the historical outcome, given that production is not rationed away, $\text{Prob}(B \mid C^*)$, declines as we ration more and more. This result is clear from Table 2. When, for example, $r = 0.5$ such that production only takes place with probability 0.41, the chance of picking the historical outcome is only slightly more than 2%, namely, $\frac{0.01}{0.41}$.
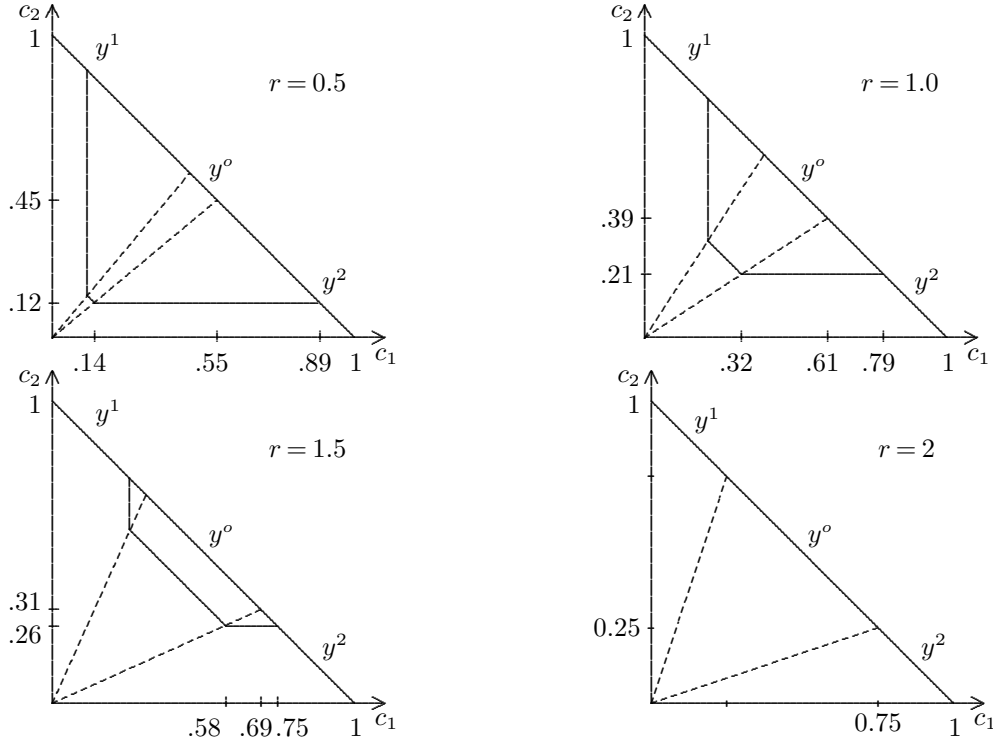
Figure 10   **SB outcomes with rationing**

7.   Extensions

The analysis above may be extended in several ways. We will comment briefly here on a generalization of the framework and on some extensions to dynamic contracting environments. The first is important in attempts to change one or more of our assumptions. We suggest a general framework that allows for numerous such adjustments. The second is relevant since many contracting environments are dynamic. Moreover, it can serve to endogenize the initial information (x^o, y^o) in a context with a strategic agent. This is particularly clear when you assume limited communication. In such cases, the information obtained from a first period is precisely that some costs x^o suffices to produce some output vector y^o. More details about the extensions are provided in the Appendix.

7.1.   General framework

We have assumed that all production plans $y \in Y$ have the same value to the principal. We have done so to emphasize the mix bias introduced to save information rents as opposed to the scale bias in traditional one-dimension adverse selection models. However, we might easily extend the framework to cases with different values associated with different production plans.

We might, for example, assume that there are m possible production levels, $y^1, \ldots, y^m$ with corresponding values $V^1, \ldots, V^m$ to the principal. At the abstract level, we can, without loss of generality, think of the resulting problem as a revelation game where the agent reveals his type c, which in turn determines the production plan y^i and the payment level x^i. The structure of

this problem is equivalent to the mutually exclusive "investment projects" problem in Antle et al. (1999). It was shown here that it suffices to select m payments levels $x^i, i = 1, \ldots, m$, one for each of the discrete alternative $y^i, i = 1, \ldots, m$, and to offer the agent a menu of contracts

"I will pay you x^i if you produce y^{i}", $i = 1, \ldots, m$

and let the agent pick the project giving him the most slack, $x^i - cy^i$. In theory, therefore, the problem is simple.

In practice, however, it is difficult to find closed form expressions for the probabilities of the different choices and therefore to actually solve for the optimal payment levels. The challenge is to add an additional structure that 1) makes sense in applications, and 2) facilitates the calculations of probabilities such that the principal's problem can be at least partially solved. The analyses in this paper illustrate this process, and we suggest that other interesting insights can be derived using a similar process on other settings with an additional structure.

7.2. Dynamic model with far-sighted agents

One application of this process is actually to the analysis of dynamic versions of the problems studied in this paper. With rational, far-sighted agents, a multi-period model looks exactly the same. Consider the case of two periods. Assume that there are m possible production levels in Period 1, $y^1, \ldots, y^m$ with corresponding values $V^1, \ldots, V^m$ to the principal, and n possible production levels in Period 2, $z^1, \ldots, z^n$ with corresponding values $W^1, \ldots, W^n$ to the principal. Again using Antle et al. (1999), we can think of the mutually exclusive projects as combinations of Period 1 and Period 2 production plans, (y^i, z^j), and it suffices to consider a set of take-it-or-leave-it offers such as

"I will pay you x^{ij} if you produce (y^i, z^j)", $i = 1, \ldots, m, j = 1, \ldots, n$

Using such a framework, one can show several interesting features of a dynamic solution. One question is how Period 1 behavior is modified by the existence of future production plans. Technically, the problem looks like the single period problem, but an important difference is that *rationing is now targeted to keep agents that are low costs not only in the production of Period 1 outputs but in the production of Period 2 outputs as well*. In a dynamic framework, the agent's choice of Period 1 production reveals information about his type. In this sense, Period 1 also serves to *explore* the type space, while Period 2 *exploits* the information from Period 1. The contract menu defines an information system. The exploration again depends on the importance of Period 1 and Period 2 production. If the outputs in Period 2 are much larger than those in Period 1, the Period 2 adaption

of types to production will dominate. The hyperplanes will be almost similar in the dynamic model and a single Period 2 model. As a consequence, Period 1 mainly serves to explore the type space. If, in contrast, Period 1 production is dominating, the exploration of type space to serve Period 2 will be downplayed, and the hyperplanes will look more like the hyperplanes in a single Period 1 model.

7.3. The impact of having myopic or limited liability agents

Of course, the above extension to multiple periods only works because the agent and the principal commit to a long-term contract. If they are myopic, e.g., because the agent faces a liquidity constraint in each period (instead of simply one individually rationality constraint covering both periods), or because the principal cannot commit to long-term payment plans and the agent therefore will fear a ratchet effect, the solutions are more complicated.

Some initial analysis of a dynamic framework with myopic and liquidity constrained agents is provided in the Appendix. We assume that the agent will accept producing in a given period if and only if he earns a non-negative slack. To limit the possible outcomes, we also follow Tirole (2016) and make one additional assumption, namely, that exit is absorbing. This means that an agent who does not produce in Period 1 also cannot produce in Period 2[14]. There can be many reasons for this. A simple reason is that restarting a production that has been halted takes time or involves prohibitive start-up costs, e.g., because it keeps a production to maintain the physical equipment, the skillsets fresh, the sales channels open, and so on. In the extreme, the agent may be out of business after a period of no production. A franchised dealer may lose its franchise rights, and in a regulatory set-up, a network company may lose its service license if it fails to perform in a period. With these additional assumptions, we can now provide first-order characterizations of the optimal contracts, much like in a single period, one-dimensional screening problem.

It is interesting to note that with limited communication, myopic agents and absorbing exit, the Period 2 problem is exactly like the problem that we have studied above. The starting point in Period 2 is the information that $x^o = x^1$ suffices to produce $y^o = y^1$. Hence, in a dynamic setting, we can think of the initial information (x^o, y^o) as being endogenously generated by the outcome in Period 1.

The more specific outcomes depend on the details of the situation. In the case of a constant production $y = z$, for example, where one period is more lucrative than the other, $V \neq W$, the principal in Period 1 takes into account the types of agents he would like to have in Period 2. *If demand becomes more lucrative over time, $V \leq W$, he rations less harshly in Period 1. If demand*

---

[14] One effect of this is that the ratchet effect is removed - the agent cannot under-produce in the first period to strike a better deal in the second period

becomes less lucrative over time, $V \geq W$, the rationing in Period 1 is not affected. There is no reason to ration in Period 1 to prepare for a less lucrative Period 2; whatever rationing is called for can just as well be implemented directly in Period 2.

8.   Some Applications

8.1.   Downsizing

Downsizing is usually understood as a permanent personnel reduction in an organization in an attempt to improve the organization's efficiency and/or effectiveness. Originally, downsizing was probably mainly targeted at blue-collar workers as the manufacturing base declined. With advancing technology and offshore outsourcing, however, white-collar workers, primarily in private companies, were downsized. In the 21st century, highly educated gold-collar workers, including attorneys, physicians, scientists and investment bankers, also found themselves among the ranks of the downsized. Similarly, as governments have implemented budget-cutting initiatives, teachers, police, fire, and healthcare employees have found that they are not safe from downsizing either; see Martin and Davis (2013)

Several downsizing strategies have been described in the literature. Three such strategies are workforce reductions, organizational redesigns and systematic strategies; cf. Cameron et al. (1991, 1993) and Cameron (1994). These strategies differ in the depth and breadth of the downsizing involved. For the purposes of this paper, it suffices to simply think of two archetypical approaches, reinforcement and reorientation. These approaches were originally developed and empirically tested by Freeman and Cameron (1993) and Freeman (1994), who labeled the concepts "convergence" and "reorientation". Reinforcement (convergence) embraces the notion of an evolutionary, incremental, and gradual change model. As a consequence, reinforced or converged downsizing would be implemented on a smaller scale and as part of a process aimed at reinforcing an organization's mission, strategy, systems, and structure. As an efficiency-enhancing strategy, the focus is on doing things better. Reorientation, in contrast, encompasses the notion of a revolutionary and discontinuous change model. Thus, reoriented downsizing would be implemented on a larger scale, with major redefinitions of the organization's mission, strategy, and structure Cameron et al. (1993). The focus is on doing the right things. We can also think of the two strategies as downscaling and downscoping, respectively; cf., e.g., Dewitt (1998).

In an organizational context, we can interpret the no-rationing results illustrated in Fig. 3 as a reinforcement approach of proportionally cutting all activities (lawn mowing). When organizations are asked to downsize modestly, it is quite common for all activities to be cut by the same

share. This approach may seem to be counter-intuitive why should the demand or preferences not affect the decisions? Proposition 1 shows that such bureaucratic behavior may not be sub-optimal by giving insufficient attention to output preferences. When no rationing is allowed, and communication is limited, the reinforcement approach is the optimal solution for all cost types. When communication is possible, the proportional adjustments are still used much more under asymmetric information than in a first-best world. Finally, when rationing occurs and we have full communication possibilities, a lawn mowing approach is still optimal more often than in a first-best world. When more dramatic downsizing is called for, a reorientation approach of elimi-nating certain activities (divesting) is often optimal. This result is clearest when communication is limited, such that a single production plan must be adopted for all cost types. In such cases, a reorientation strategy is superior to a reinforcement strategy as soon as sufficient rationing is called for. Additionally, the probability of a reorientation strategy being optimal increases as changes in demand force the principal to ration more harshly. Fig. 11 illustrates the areas of the cost space that lead to no production, to production according to a reinforcement strategy (lawn mowing) and to production according to a reorientation strategy.
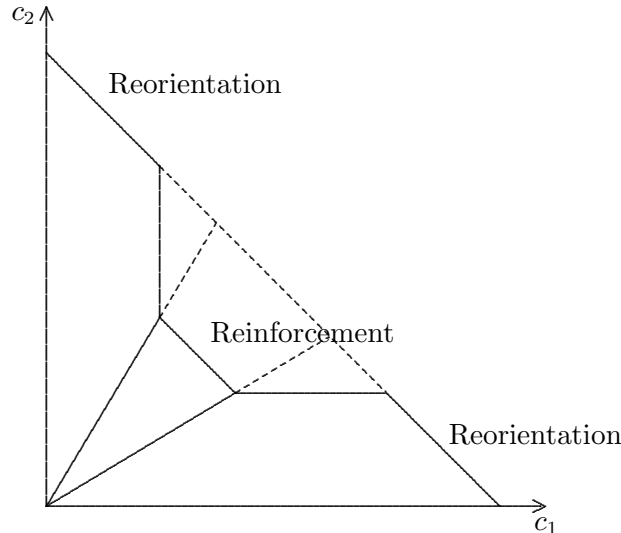

Figure 11   Optimal rightsizing strategies under asymmetric information

8.2. The nature of increasing health costs.

Our analysis may also help understand why most, if not all, healthcare systems have struggled to control costs over the last several decades. The delivery of health care, especially in the United States, has been remarkably unyielding to attempts to achieve efficiency. Legions of MBAs and oth-ers who trained in traditional production and cost-control methods have worked in the healthcare sector for some time now, but the costs are widely viewed as increasing at unacceptable rates.

There is no shortage of things that have been blamed for this trend, including gross inefficiency, poor costing systems, a cost-reimbursement payment system, and the separation between the consumers of medical services from those who pay for it (see, for example, Kaplan and Porter (2011) and Porter and Lee (2013). Our analysis suggests another consideration to throw on this pile: the relatively better cost information possessed by the providers in an environment with an almost unimaginable scope of treatments and changing values of those treatments. It also suggests a serious shortcoming of proposed improvements that are confined to the cost side of production and suppress consideration of possible changes in what is produced. Furthermore, proposed improvements that do not involve rationing will have serious limitations in addressing asymmetric information problems.

Proposition 1 shows that the cheapest mix is the historical one and that any deviation from this can be used by the agent to extract information rents. The agent can claim high unit costs on the most expanded / least contracted output in the new service profile. A new treatment corresponds to an output that has never been tried before. The unit costs of this service is therefore unconstrained in our model, and a non-rationing principal can be forced to pay excessive costs. Even if no new treatments are introduced, a change in the mix leads to increases in costs. The minimal payment necessary to ensure the production of output vector $(2,0)$ in our first example was 2 when the mix of $(1,1)$ historically had been produced at a cost of 1.

Demand changes may make ignoring rationing even more costly. This is a consequence of the details of the informational problem, and in particular how the historical data delimit the possible cost types. In a new mix, rationing is more efficient because less production has to be forgone to lower payments to the remaining cost types. *Hence, although changes in health demand drive up costs when rationing is not allowed, the same changes may actually help improve the social values when rationing is allowed.*

Our analysis therefore suggests that an intrinsic reason for increasing healthcare costs is the unwillingness to ration.

9. Final Remarks

In this paper, we have analyzed a simple multiple dimensional adverse selection problem.

We have seen that with asymmetric information the optimal solution will generally be biased towards the historical mix.

The advantage of proportional adjustments, i.e., of only rescaling the historical mix, is easiest to understand when the principal will not or cannot ration. In such cases, the agent has considerable bargaining power. He can essentially claim that the least reduced or the most expanded production

is the main cost driver. Consequently, the cost function facing the principal is a Leontief-like cost function, and the cost minimal production plan is always proportional to the historical plan. The lawn mowing limits the agent's ability to extract rents by claiming high costs for the least reduced activities.

When rationing is introduced, the virtue of the proportional approach is reduced because the cost of rationing, namely, the production forgone for some types to reduce the cost of other types, is relatively high when the production mix is similar to the historical mix. Rationing is much more efficient for extreme mixes; therefore, the virtue of the status quo mix is gradually reduced as more and more rationing is called for. An extreme mix makes rationing to save information rents more effective.

The mix stickiness and the value of reorientation when severe rationing is called for give new insight into several business practices. We have discussed the tendency to use across-the-board cutbacks with no clear priorities when downsizing as one example, and the intrinsic nature of the increasing health costs as another example. Many other applications seem possible. One could for example use the framework to rationalize outsourcing of simple scaleable activities and insourcing of services that vary in scope instead of scale.

The analysis is this paper can be extended in several ways.

We have briefly discussed the introduction of more general value functions as well as dynamic extensions in the paper. It seems that in particular the analysis of dynamics and the interaction between exploration of the multiple-dimensional type space in early periods and exploitation of the information acquired in later periods would be interesting.

Another extension could be to introduce more historical information, e.g., in the form of several observations of production and costs in previous periods. Likewise, one could allow the underlying cost function to be more complex, e.g., affine or convex, compared with the linear form assumed here. We conjecture that the historical outcomes would again be overrepresented in the solution as long as we do not ration too harshly[15].

Let us finally note that the choice of a rescaling and rescoping strategy of course will depend on numerous factors in real applications. In particular, the market details would normally be an important determinant. We have modelled the market side, or the demand side, as simply as possible in this paper, and it might be worthwhile to investigate more elaborate models. Still, our

---

[15] A simple extension of the example would be to analyze cases where the outputs are not pure substitutes, as we have assumed above. The perfect substitution case was chosen because it gives the least mix stickiness in a first-best world. A simple possibility is to let

$$Y = conv\{k(2,0), (1,1), k(0,2)\} + \mathbb{R}_+^2$$

When $k = 1$, we have the case of perfect substitutes above, and when $k \to \infty$, we have a case of perfect complements. It is easy to see that with values of $k > 1$, the historical outcome gets more attractive, and there would be some organizational inertia even in a first-best world. The main conclusions would, however, not be altered.

results also suggest that the details of the market may not matter. In fact, this was the puzzle that initiated this research.

References

Antle, R., P. Bogetoft, A. Stark. 1999. Selection from many investment with managerial private information. *Contemporary Accounting Research* **16**(3) 397–418.

Antle, R., G.D. Eppen. 1985. Capital rationing and organizational slack in capital budgeting. *Management Science* **31** 163–174.

Antle, R., J. Fellingham. 1997. Models of capital investments with private information and incentives: a selective review. *Journal of Business Finance and Accounting* **24** 887–908.

Armstrong, M, J Rochet. 1999. Multi-dimensional screening: A user's guide. *European Economic Review* **43** 959–979.

Armstrong, Mark. 1996. Multiproduct nonlinear pricing. *Econometrica* **64**(1) 51–75.

Banker, R. D. 1980. A game theoretic approach to measuring efficiency. *European Journal of Operational Research* **5** 262–268.

Banker, R. D., A. Charnes, W. W. Cooper, R. Clarke. 1989. Constrained game formulations and interpretations for data envelopment analysis. *European Journal of Operational Research* **40** 299–308.

Baron, D.P., R.B. Myerson. 1982. Regulating a monopolist with unknown costs. *Econometrica* **50**(4) 911–930.

Boer, Arnoud von den. 2015. Review: Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science* **20** 1–18.

Bogetoft, P. 1997. DEA-based yardstick competition: The optimality of best practice regulation. *Annals of Operations Research* **73** 277–298.

Bogetoft, P. 2000. DEA and activity planning under asymmetric information. *Journal of Productivity Analysis* **13** 7–48.

Bogetoft, P., L. Otto. 2011. *Benchmarking with DEA, SFA, and R*. Springer, New York.

Cameron, K. S. 1994. Strategies for successful organizational downsizing. *Human Resource Management* **33** 189–211.

Cameron, K. S., S. J. Freeman, A. K. Mishra. 1991. Best practices in white-collar downsizing: managing contradiction. *Academy of Management Executive* **5**(3) 57–73.

Cameron, K. S., S. J. Freeman, A. K. Mishra. 1993. Downsizing and redesigning organizations. W Huber, G Glick, ed., *Organizational Change and Redesign*. Oxford University Press, New York, 19–63.

Charnes, A., W. W. Cooper, E. Rhodes. 1979. Short Communication: Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* **3** 339.

Charnes, A., William W. Cooper, E. Rhodes. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** 429–444.

Dantzig, G.B., P Wolfe. 1960. Decomposition principle for linear programs. *Operations Research* **8** 101–111.

Datar, Srikant M., Madhav V. Rajan. 2014. *Managerial Accounting: Decision Making and Motivating Performance*. Pearson.

Debreu, G. 1951. The coefficient of resource utilization. *Econometrica* **19**(3) 273–292.

Dewitt, R-L. 1998. Firm, industry, and strategy influences on choice of downsizing approach. *Strategic Management Journal* **19** 59–79.

Freeman, S. J., K. S. Cameron. 1993. Organizational downsizing: A convergence and reorientation framework. *Organizational Science* **4** 10–29.

Freeman, S.J. 1994. Organizational downsizing as convergence or reorientation: implications for human resource management. *Human Resource Management* **33**(2) 213–238.

Gonsch, J., R. Klein, M. Neugebauer, C. Steinhardt. 2013. Dynamic pricing with strategic customers. *Journal of Business Economics* **83** 505–549.

Harris, M., A. Raviv. 1996. The capital budgeting process, incentives and information. *Journal of Finance* **51** 1139–1174.

Holmstrom, B., P. Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, Organization* **7** 24–52.

Horngren, C.T. 1981. *Management Accounting*. Prentice Hall.

Kaplan, R. S., M. E. Porter. 2011. How to solve the cost crisis in health care. *Harvard Business Review* **89**(9) 47–64.

Kouvelis, P., C. Chambers, H. Wang. 2006. Supply chain management research and production and operations management: Review, trends, and opportunities. *Production and Operations Managemen* **15**(3) 449–469.

Laffont, Jean-Jacques, Jean Tirole. 1986. Using cost observation to regulate firms. *Journal of Political Economy* **94**(3) 614–641.

M. Harris, M., C. H. Kriebel, A. Raviv. 1982. Asymmetric information, incentives and intrafirm resource allocation. *Management Science* **28**(6) 589–603.

March, J., H.A. Simon. 1958. *Organizations, New York: John Wiley and Sons*. John Wiley and Sons.

Martin, W. M., A. C. Davis. 2013. Alternatives to downsizing: An organizational innovation approach. *International Journal of Business and Social Research* **3**(7) 19–27.

McAfee, R. P., J. McMillan. 1988. Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory* **46**(2) 335–354.

Mussa, M, S Rosen. 1978. Monopoly and product quality. *Journal of Economic Theory* **18** 301–317.

Myerson, Roger B. 1979. Incentives compatibility and the bargaining problem. *Econometrica* **47** 61–73.

Porter, M. E., T. H. Lee. 2013. The strategy that will fix health care. *Harvard Business Review* **91**(10) 50–70.

Rochet, J.-C., P. Chone. 1998. Ironing, sweeping and multidimensional screening. *Econometrica* **66**(4) 783–826.

Rochet, J.-C., L. A. Stole. 2003. The economics of multidimensional screening. M Dewatripont, L P Hansen Hansen, S J Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, vol. III. Cambridge University Press, 150–197.

Rogerson, W.P. 1994. Economic incentives and the defense procurement process. *Journal of Economic Perspectives* **8**(4) 65–90.

Rogerson, W.P. 2003. Simple menus of contracts in cost-based procurement and regulation. *American Economic Review* **93**(3) 919–926.

Sappington, D.E.M. 1983. Limited liability contracts between principal and agent. *Journal of Economic Theory* **29**(1) 1–21.

Shen, Z.M., X. Su. 2007. Customer behavior modeling in revenue management and auctions: a review and new research opportunities. *Prod. Oper. Manage.* **16** 713–728.

Tirole, Jean. 2016. From bottom of the barrel to cream of the crop: Sequential screening with positive selection. *Econometrica* **84**(4) 1291–1343.

Varian, Hal R. 2014. *Intermediate Microeconomics*. Norton, New York.

Williamson, Oliver. 1985. *The Economic Institutions of Capitalism*. The Free Press.

Appendix A: Detailed analysis of the uniform case

In this appendix, we give the analytical details of our analysis of the uniform example.

Limited communication and rationing

We here analyze the numerical example in the case where we seek to implement a specific but arbitrary production plan y^* and allow rationing.

Consider first a production plan that is proportional to the original plan y^o

$$\exists k \geq 0 : y^* = k y^o$$

In this case, the principal must pay k if he does not ration, which happens when $\alpha \geq k$. Effective rationing takes place when $k > \alpha \geq 0$, in which case the principal pays the agent α and obtains production with probability $\frac{\alpha^2}{k^2}$. With rationing in place, the principal's optimization problem therefore becomes

$$\max_{\alpha \in \mathbb{R}} (r^* - \alpha) \frac{\alpha^2}{k^2}$$

which is optimized for $\alpha = \frac{2}{3} r$ if $\frac{r}{k} \leq \frac{3}{2}$ and k otherwise. Inserting this into the profit formula and comparing it with the profit under no rationing, we have the maximal possible profit

$$\Pi^{\text{max}} = \begin{cases} \frac{4}{27} \frac{r^3}{k^2} & \text{if } \frac{r}{k} \leq \frac{3}{2} \\ r - k & \text{otherwise} \end{cases}$$

$$\Pi^{\text{max}} = \max\{r - k, \frac{4}{27} \frac{r^3}{k^2}\}$$

When $r = 1$ and $k = 1$, for example, the principal is willing to pay $\frac{2}{3}$. The probability of production in this case is $\frac{4}{9}$. It is interesting to compare the result to a situation where there is only one product with value 1 and uniform costs on $[0, 1]$; in this case, the optimal cut off is found by solving $\max_{\alpha}(1 - \alpha)\alpha$, leading to an optimal value of $\alpha = \frac{1}{2}$. The willingness to pay $\frac{2}{3}$ in the two-dimensional case compared with $\frac{1}{2}$ in the one-dimensional case is explained by the fact that large values for total cost are more likely in the two-dimensional case. Rationing is therefore more expensive because the principal will lose more production for a given cost decrease than in the one-dimensional case. Indeed, in terms of probabilities, he also rations away more production in the two-dimensional case, where the probability of production is $\frac{4}{9}$, compared with $\frac{1}{2}$ in the one-dimensional case.

Consider now a non-proportional production plan

$$\nexists k \geq 0 : y^* = k y^o$$

The intersection between $c_1 + c_2 = 1$ and $c_1 y_1^* + c_2 y_2^* = \alpha$ is $c_1 = \frac{\alpha - y_2^*}{y_1^* - y_2^*}$ and $c_2 = \frac{y_1 - \alpha}{y_1 - y_2}$.

Consider the case where $y_1^* \geq y_2^*$. With no rationing, the agent will claim $c = (1, 0)$, and the principal must pay y_1^*, leading to a profit of $r - y_1^*$. Assume now that $y_1^* \geq \alpha \geq y_2^*$ such that rationing

will actually take place. The probability of production in this case can easily be calculated. It is 2 times the area that remains after the elimination of the rationed unit cost vectors. Multiplying by the profit in case of production $r - \alpha$, we obtain the principal's problem of finding the optimal rationing level

$$\max_{\alpha}(r - \alpha)(1 - (1 - \frac{\alpha}{y_1^*})(\frac{y_1^* - \alpha}{y_1^* - y_2^*}))$$

Solving for the optimal α, we obtain

$$\alpha^* = \frac{1}{3}(r + 2y_1^* - [r^2 - 2ry_1^* + 4(y_1^*)^2 - 3y_1^*y_2^*]^{\frac{1}{2}})$$

Finally, for $y_1^* \geq y_2^* \geq \alpha$, we find that the principal solves

$$\max_{\alpha}(r - \alpha)(\frac{\alpha^2}{y_1^*y_2^*})$$

leading to

$$\alpha^{**} = \frac{2r}{3}$$

Inserting back into the equation, we find that maximum profits are

$$\Pi^{max} = \max\{r - y_1^*, (r - \alpha^*)(1 - (1 - \frac{\alpha^*}{y_1^*})(\frac{y_1^* - \alpha^*}{y_1^* - y_2^*})), \frac{4r^2}{9y_1^*y_2^*}\}$$

where the first value corresponds to no rationing and the last two values correspond to the optimal rationing profits for $\alpha \geq y_2$ and $\alpha \leq y_2$.

Full communication and rationing

Consider first the area A in Fig.. Taking the area of the large triangle and subtracting the small triangle, the full set of possible cost types has area 0.5, and we obtain

$$\text{Prob}(A) = 2 \cdot \{\frac{1}{2}k - \frac{1}{2}(1 - \frac{\alpha}{2} - \frac{\alpha}{2}\frac{1 - k}{k})(k - \frac{\alpha}{2})\}$$

which can be rewritten as

$$\text{Prob}(A) = \alpha\{1 - \frac{\alpha}{4}\frac{1}{k}\}$$

To determine the $\text{Prob}(B)$, we see that the midpoint of the northeast boundary is

$$0.5(\frac{\alpha}{2}, \frac{\alpha}{2}\frac{1 - k}{k}) + 0.5(\frac{\alpha}{2}\frac{1 - k}{k}, \frac{\alpha}{2}) = \frac{1}{2}\frac{\alpha}{2}\frac{1}{k}(1, 1)$$

such that the height of the B triangle is

$$\frac{\alpha}{4}\frac{1}{k}\sqrt{2}$$

The length of the northeast boundary is

$$\frac{\alpha}{2}(1 - \frac{1-k}{k})\sqrt{2}$$

and we therefore have

$$\mathrm{Prob}(B) = \frac{\alpha^2}{4}(\frac{1-k}{k} - 1)\frac{1}{k}$$

because $k \leq 0.5$. We therefore obtain the expected profit as

$$\Pi[c, \alpha, r] = (r - \alpha) \cdot 2 \cdot \alpha \{1 - \frac{\alpha}{4}\frac{1}{k}\} + (r - \frac{\alpha}{2}\frac{1}{k}) \cdot \frac{\alpha^2}{4}(\frac{1-k}{k} - 1)\frac{1}{k}$$

The profit function when $r = 1$ is illustrated in Fig 12 below.
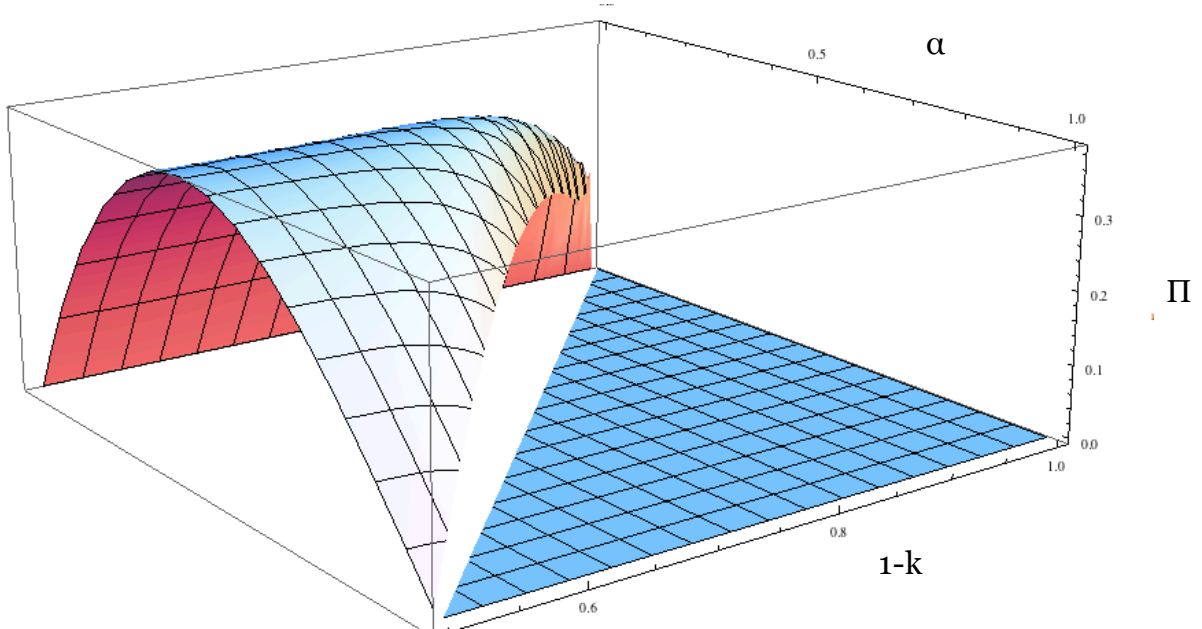

Figure 12   **Profit function when** $r = 1, 0 \leq \alpha \leq 1, 0.5 \leq 1 - k \leq 1$

The optimal solution to the general contract design problem, i.e., the second-best solution with rationing, can therefore be written as

$$\max_{k,\alpha} \ \Pi[k, \alpha, r]$$
$$\text{s.t. } 0 \leq k \leq 0.5$$
$$0 \leq a \leq 2k$$

If we define

$$H = 9 - r - \sqrt{36 - 18r + r^2}$$

the solution can be written as

$$k = 1 - \frac{1}{6}H$$

$$a = \frac{1}{45}(-24 + 36r + 20H - \frac{14}{3}rH - 6H^2 + \frac{2}{9}rH^2 + \frac{7}{9}H^3 - \frac{2}{27}rH^3 - \frac{1}{27}H^4)$$

when $r \leq 1.75$ and

$$k = \frac{1}{4}$$

$$a = \frac{1}{2}$$

when $r \geq 1.75$. The latter corresponds to the optimal non-rationing solution from above.

Appendix B: Extensions

The analysis in the main body of the article can be extended in several ways. In this Appendix, we give a more detailed account of the extensions briefly commented on in Section 7. We discuss situations where the principal has a more elaborate value function, and we discuss the introduction of dynamic features.

General value function

Assume that there are m possible production levels, $y^1, \ldots, y^m$ with corresponding values $V^1, \ldots, V^m$ to the principal. Without loss of generality, we can think of the principal's problem in this setting as a revelation game where the agent reveals his type c, which in turn determines the production plan y^i and the payment level x^i. A contract in this model is therefore a specification of the cost contingent production and payment plans

$$y[\cdot] : \mathbb{R}_0^2 \to \{y^1, \ldots, y^m\}$$

$$x[\cdot] : \mathbb{R}_0^2 \to \mathbb{R}_0$$

Now, it is relatively easy to show that it suffices to look at simple take-it-or-leave-it offers and to leave the implementation to the agent. That is, it suffices for the principal to define a series of offers like

"I will pay you x^i if you produce y^i", $i = 1, \ldots, m$

Faced with these offers, the agent chooses which production plan to implement and collects the corresponding payment. The production implemented is the one maximizing the agent's slack, i.e., that solves the *agent's problem*

$$\arg\max_i \ x^i - cy^i$$

The production plan y^i is implemented if it maximizes the slack $x^i - cy^i$ and leads to an individual rational outcome, $x^i - cy^i \geq 0$. If the best outcome is not individually rational, the agent rejects the contract.

To show that it suffices to look at such contracts, we can think of the general problem of choosing between mutually exclusive "investment projects" like in Antle et al. (1999). It is shown in this paper that an optimal solution has the simple structure of a handicapping scheme. The principal must choose the payment levels, one for each of the projects, to express what he will pay for the various projects. The agent then chooses the investment project to maximize his payment minus his cost of the project.

Let us introduce a shorthand notation $A^i(x, y)$ for the set of types choosing production plan (x^i, y^j) :

$$A^i(x) = \{c \,|\, i \text{ solves the agent's problem}\}$$

Note that we use x as shorthand for the payment menu $(x^i, i = 1 \ldots, m)$. Using this definition, we have that the *principal's problem* is to solve

$$\max_x \sum_i (V^i - x^i) \text{Prob}\{A^i(x)\}$$

In summary, we can easily extend the framework to one with a general value function. Solving the problem may, however, be difficult because the contracts define a series of polyhedron subsets $A^I, i = 1, \ldots, m$ and we need to calculate probabilities on these to determine the principal's expected net-value. With just two possible production plans, $m = 2$, the sets would look like in Fig. 13 below. The borderline between A^1 and A^2 is defined by $x^1 - cy^1 \geq x^2 - cy^2$ or, equivalently, by $x^1 - x^2 \geq c(y^1 - y^2)$.

Figure 13   Using a menu of contracts
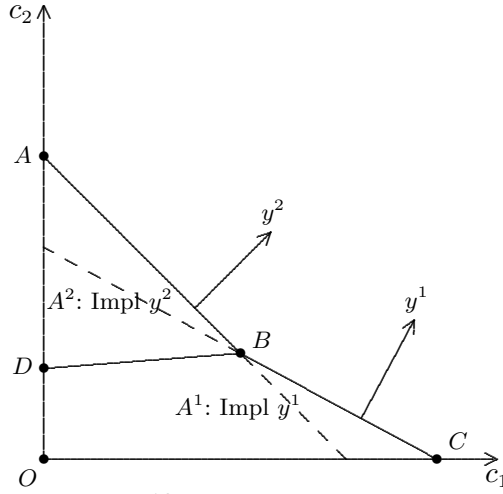
The challenge is therefore to make supplementary assumptions that simplify the calculation of the probabilities of the different outcomes such that the principal's problem can be examined in more detail.

Dynamic model with far-sighted agents

With rational, far-sighted agents, a multi-period model looks exactly the same. Consider the case of two periods. Assume that there are m possible production levels, in Period 1, $y^1, \ldots, y^m$ with corresponding values $V^1, \ldots, V^m$. Similarly, in Period 2, there are n possible production levels, $z, \ldots, z^n$ with corresponding values $W^1, \ldots, W^n$.

Again using Antle et al. (1999), we can think of the mutually exclusive projects as combinations of Period 1 and Period 2 production plans, (y^i, z^j), and it suffices to consider a series of offers like

"I will pay you x^{ij} if you produce (y^i, z^j)", $i = 1, \ldots, m, j = 1, \ldots, n$

The principal's contract design problem is therefore effectively reduced to the choice of $m \times n$ targets $x^{ij}, i = 1, \ldots, m, j = 1, \ldots, n$. They serve two purposes.

First, they define the production selection rules: The productions implemented are the ones maximizing the agent's slack, i.e., that solves the *agent's problem*

$$\arg \max_{i,j} \ x^{ij} - c(y^i + z^j)$$

and satisfy the individual rationality constraint $x^{ij} - c(y^i + z^j) \geq 0$. If there are no such offers, the agent does not contract. Second, they define the payment rule: x^{ij} is the amount of resources transferred when (y^i, z^j) is produced.

Again, we can introduce a shorthand notation $A^{ij}(x)$ for the set of types choosing production plans (y^i, z^j):

$$A^{ij}(x) = \{c \,|\, (i, j) \text{ solves the agent's problem}\}$$

Using this definition, we have that the *principal's problem* is to determine payment levels $x = (x^{ij}, i = 0 \ldots, m, j = 0, \ldots, n)$ to solve

$$\max_x \sum_{i,j} (V^i + W^j - x^{ij}) \text{Prob}\{A^{ij}(x)\}$$

Let us illustrate a few insights from such an extension of the model. We will here, for simplicity, assume that $m = 2$ and $n = 1$ such that there is only one possible production mix in Period 2. In effect, there are therefore two possible contracts in the menu: (y^1, z) and (y^2, z).

An interesting question is how Period 1 behavior is modified by the existence of future production plans. Technically, the problem looks like the single period problem, but an important difference is that *rationing is now targeted to keep agents that are low costs not only in the production of Period 1 outputs but in the production of Period 2 outputs as well*. This is illustrated in Fig. 14 below. If we assume, for example, that product 2 is the primary output in Period 2, the contracts will be targeted more toward types with a low marginal cost of product 2, c_2. Graphically, all of the rationing hyperplanes will be more horizontal.

They will also be located more to the North-East because the cost of rationing increases as it now involves forgoing profit from Period 2.

It is interesting to note also that if z is small, the contracts are designed mainly to enable an optimal adaption of types to Period 1 production. If, in contrast, z is large, the adaption of types to Period 1 production plays a smaller role.

We also see that the choice of Period 1 production reveals information about the agent types. If the agent produces y^1, the principal knows that the agent belongs to the triangle DAB, and if the agent produces y^2, the principal knows that the agent belongs to ODBC. In this sense, Period
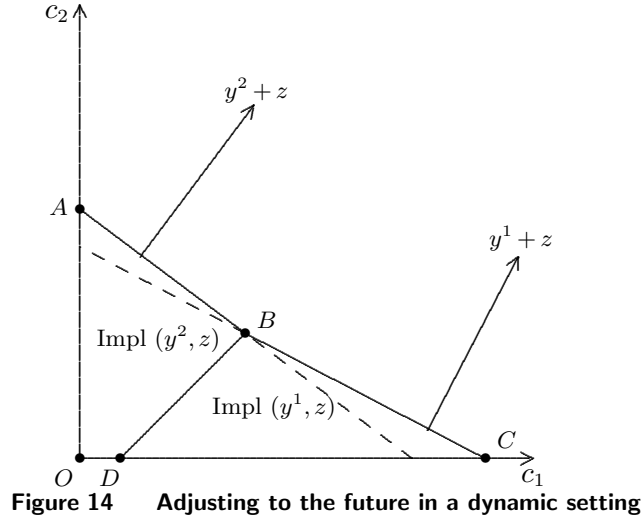
Figure 14   Adjusting to the future in a dynamic setting

1 also serves to *explore* the type space. The contract menu defines an information system. The exploration again depends on the importance of Period 1 and Period 2 production. If the outputs in Period 2 are much larger than those in Period 1, the Period 2 adaption of types to production will dominate. The hyperplanes will be almost similar in the dynamic model and a single Period 2 model. As a consequence, Period 1 mainly serves to explore the type space. If, in contrast, Period 1 production is dominating, the exploration of type space to serve Period 2 will be downplayed, and the hyperplanes will look more like the hyperplanes in a single Period 1 model.

In summary, the principal extracts information about the agent types in Period 1. The exploration depends on the importance of production in Period 2. *If Period 2 is highly valuable and production is large, the Period 1 adaption of types to production plays only a minor role - Period 1 is serving mainly to explore the type space with the aim of supporting Period 2 usage.*

The above extension to a dynamic setting presumes that the agent and the principal are far-sighted and are able to commit ex-ante to multiple-period contracts. If this is not the case, the solution approach does not work.

To illustrate this, let us also look at a simple setting with just one production mix in each of the periods. Let these mixes be y and z, respectively, and let the corresponding values to the principal be V and W as above. Furthermore, let the payment in Period t be denoted be denoted $x^t, t = 1, 2$.

Imagine first that the principal tries to use the contract designed for a long-sighted agent. This is illustrated by the AB rationing constraint in Fig. 15 below. In the long-sighted setting, all agents in the triangle OAB would accept this contract.

With myopic agents, this is not possible. In Period 1, the principal will need to pay $x^1 = c^B y$ to get all types in OAB to accept the contract. However, this would make all agents in OCB accept the contract in Period 1. Hence, it is not the agents in OAB that end up producing in Period 1. In Period 2, it will be a subset of the agents in OCB that produces. The exact subset of OCB
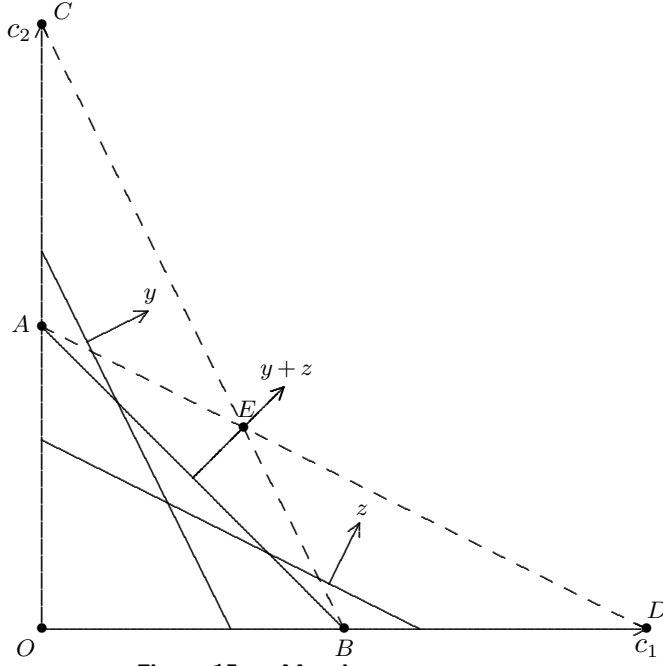
45

Figure 15   Myopic agents

that will produce in Period 2 depends on the payment level. If the principal recalling his original plan decides to pay the highest possible costs of producing z in the OAB triangle, he will pay $c^A z$, where $c^A = (0, A)$, and the types producing in Period 2 will be the types in OAEB. Paying anything less would imply that some of the OAB types will be rationed away in Period 2 and that some of the types in ABE would be asked to produce in Period 2. In summary, therefore, the introduction of myopic agents leads to two types of inefficiencies compared to the case of rational long-sighted agents. Some agents with low $y + z$ costs are at risk of being rationed away, and some agents with high $y + z$ costs are at risk of being included in the production plan.

Another way to illustrate the difficulty of implementing the long-sighted contract is to note that the worst Period 1 costs is $c^B y$, where $c^B = (0, B)$, and the worst Period 2 costs is $c^A z$. To make all agents in OAB accept the production of y and z when they are myopic, the payments will therefore have to be $x^1 = c^B y$ and $x^2 = c^A z$. This corresponds to claiming that all costs are associated with product 1 in Period 1 and with product 2 in Period 2, which of course leads to excessive payments. Indeed, if the principal pays $x^1 = c^B y$ and $x^2 = c^A z$, the types producing in Period 1 will be OCB, and the types producing in Period 2 will be the subset OAEB hereof.

Dynamic model with myopic agents

We have seen above that the general dynamic solution does not work with short-sighted agents. In this sub-section, we derive an optimal solution in one such case.

For simplicity, we first look at a simple setting with just one production mix in each of two periods, Periods 1 and 2. Let these mixes be y and z, respectively, and let the corresponding values

to the principal be V and W. In addition, let the payment in Period t be denoted $x^t, t = 1, 2$. With fixed mixes, we can investigate how the harshness of the rationing depends on the dynamic features and on the product mixes that we implement. We cannot, however, investigate how the dynamic perspective affects the choice of production mix. We shall return to this, however, in some examples below.

We will assume that the agent is myopic. The agent will accept producing in a given period if and only if he earns a non-negative slack. Alternatively, we can think of the agent as facing a liquidity constraint or having limited liability in each period. To limit the possible outcomes, we also follow Tirole (2016) and make one additional assumption, namely, that exit is absorbing. This means that an agent who does not produce in Period 1 also cannot produce in Period 2.

The structure of an optimal solution to the two period problem is illustrated in Figure 16 below. The types A^1 that accept production in Period 1

$$A^1 = \{c \mid cy \le x^1\}$$

are the types below the rationing hyperplane in the left panel, and the types A^2 that also accept in Period 2

$$A^2 = \{c \mid cy \le x^1, cz \le x^2\}$$

are the ones below both hyperplanes in the right panel. In Period 1, we ration using the requirement $cy^1 \le x^1$, and in Period 2, we ration further by adding the constraint $cz \le x^2$.
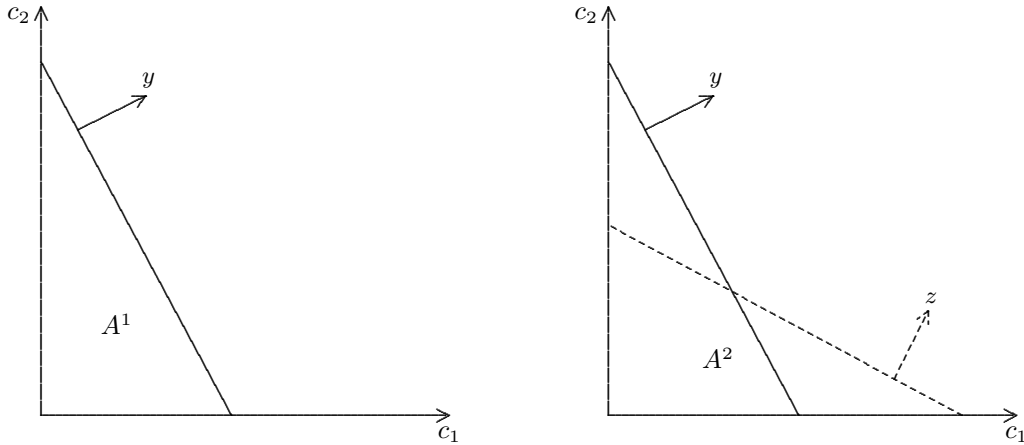

Figure 16   Optimal rationing under changing production mix

It is interesting to note that with myopic agents and absorbing exit, the Period 2 problem is exactly like the problem that we have studied above. The starting point in Period 2 is the

47

information that $x^o = x^1$ suffices to produce $y^o = y$. Hence, in a dynamic setting, we can think of the initial information (x^o, y^o) as being endogenously generated by the outcome in Period 1.

The principal's contract design problem with production plans y and z reduces to the choice of payment levels x^1 and x^2, one for each of the periods, to solve

$$\max_{x^1, x^2} (V - x^1) G_{Z^1}(x^1) + G_{Z^1}(x^1)(W - x^2) G_{Z^2 | Z^1}(x^2 \mid x^1)$$

where $Z^1 = cy$, $Z^2 = cz$, $G_{Z^1}(x^1) = \text{Prob}\{Z^1 \leq x^1\}$ is the cumulative distribution function for Z^1 and $G_{Z^2 | Z^1}(x^2 \mid x^1) = \text{Prob}\{Z^2 \leq x^2 \mid Z^1 \leq x^1\}$ is the conditional cumulative distribution function for Z^2 given the event $Z^1 \leq x^1$. The first term in the objective function, $(V - x^1) G_{Z^1}(x^1)$, is the net value to the principal if the agent accepts producing y^1 for a compensation of x^1 times the chance that the agent accepts this. The second term, $G_{Z^1}(x^1)(W - x^2) G_{Z^2 | Z^1}(x^2 \mid x^1)$ is the probability that the agent does not exit in Period 1 times the net expected surplus from Period 2 where the agent is offered x^2 to produce z.

The principal's problem is, as previously, a screening problem, and because we have two-dimensional types, we can think of the screening both in terms of the direction in type space and the harshness of the screening. The direction is illustrated by the y vector in the left panel of Figure 16. We can also think of it in terms of the agents unit cost ratio, $\frac{c_2}{c_1}$, i.e., the slope of a line from $(0,0)$ to c. The harshness is determined by the payment levels $x^t, t = 1, 2$. The less we pay, the harsher the rationing is. Changing the harshness corresponds to a parallel shift in the rationing hyperplanes or "budget lines" in Figure 16.

In our dynamic model, two aspects of demand may change. First, the value of production in Periods 1 and 2, V and W, may differ. Second, the required production (i.e., the levels and production mixes), y and z, may differ. If $W > V$ but production is constant we can say that Period 2 is more *lucrative* than Period 1. If, however, production increases over time, $z_1 > y_1, z_2 > y_2$ while the values stay constant $V = W$, we can say that Period 2 is more *demanding* than Period 1. A less lucrative and/or a more demanding period is also less profitable because the difference between values generated and the associated total costs is declining.

Let us now consider two special instances.

Constant production, $y = z$

Consider the case of a constant production $y = z$, where one period is more lucrative than the other, $V \neq W$. The principal's problem in this case can be formulated as follows:

$$\max_{x^1, x^2} (V - x^1) G_Z(x^1) + G_Z(x^1)(W - x^2) G_{Z|Z}(x^2 \mid x^1)$$

48

where G_Z is the cumulative cost distribution for total costs $Z = cy = cz$ and $G_{Z|Z}(\cdot \mid x^1)$ is the cumulative distribution of costs Z given that costs are below x^1.

We can safely restrict the payment in Period 2 to $x^2 \leq x^1$ because if $x^2 \geq x^1$, we have $G_{Z|Z}(x^2 \mid x^1) = 1$ and the objective becomes one of maximizing $[(V - x^1) + (W - x^2)]G_Z(x^1)$, which is declining in x^2. Furthermore, when $x^2 \leq x^1$, we have

$$G_Z(x^1)G_{Z|Z}(x^2 \mid x^1) = \text{Prob}\{Z \leq x^2, Z \leq x^1\} = \text{Prob}\{Z \leq x^2\} = G_Z(x^2)$$

and the principal's problem reduces to

$$\max_{x^2 \leq x^1} (V - x^1)G_Z(x^1) + (W - x^2)G_Z(x^2)$$

When demand becomes less lucrative over time, $W < V$, unrestricted optimization over x^1 and x^2 will satisfy the constraint $x^2 \leq x^1$, cf the usual one-period, first-order condition, and the principal's problem reduces to

$$\max_{x^1, x^2} (V - x^1)G_Z(x^1) + (W - x^2)G_Z(x^2)$$

which decomposes into two single-period problems. The optimal solution satisfies

$$x^1 = V - \frac{G_Z(x^1)}{g_Z(x^1)} \text{ and } x^2 = W - \frac{G_Z(x^2)}{g_Z(x^2)}$$

corresponding to two myopic solutions.

When demand becomes more lucrative over time, $W > V$, the principal wants a high chance of production in Period 2, i.e., $x^2 = x^1 = x$, and the contracting problem reduces to

$$\max_x (V - x)G_Z(x) + (W - x)G_Z(x)$$

The first order condition for an optimal solution in this case satisfies

$$x = \frac{V + W}{2} - \frac{G_Z(x)}{g_Z(x)}$$

In summary, therefore, in the dynamic solution with fixed production, $y = z$, the principal in Period 1 takes into account the types of agents he would like to have in Period 2. *If demand becomes more lucrative over time, he rations less harshly in Period 1. If demand becomes less lucrative over time, the rationing in Period 1 is not affected.* There is no reason to ration in Period 1 to prepare for a less lucrative Period 2; whatever rationing is called for can just as well be implemented directly in Period 2.

Non-constant production, $y \neq z$

We now consider the general case with non-constant production levels. The principal's optimization problem in this case is

$$\max_{x^1, x^2}(V - x^1)G_{Z^1}(x^1) + G_{Z^1}(x^1)(W - x^2)G_{Z^2|Z^1}(x^2 \mid x^1)$$

where as above, $Z^1 = cy$ and $Z^2 = cz$.

The first-order condition for an optimal x^1 is

$$V g_{Z^1}(x^1) - x^1 g_{Z^1}(x^1) - G_{Z^1}(x^1) +$$

$$g_{Z^1}(x^1)(W - x^2)G_{Z^2|Z^1}(z^2 \mid z^1) + G_{Z^1}(x^1)(W - x^2)\frac{\partial G_{Z^2|Z^1}(x^2 \mid x^1)}{\partial x^1} = 0$$

i.e.

$$x^1 = V - \frac{G_{Z^1}(x^1)}{g_{Z^1}(x^1)} + (W - x^2)G_{Z^2|Z^1}(x^2 \mid x^1) + \frac{G_{Z^1}(x^1)}{g_{Z^1}(x^1)}(W - x^2)\frac{\partial G_{Z^2|Z^1}(x^2 \mid x^1)}{\partial x^1}$$

The first two terms are the usual terms in our first-order characterization of x^1. The last two terms, in contrast, originate from the interaction with Period 2. We see that these terms are positive. The presence of Period 2 therefore makes the principal weakly increase the payment to the agent in the first period. Hence, what we can say at the general level is that *the dynamic perspective tends to reduce rationing in the first period.*

We see also that the reduction in Period 1 rationing is larger, i.e., we ration less, the higher is

$$\frac{\partial G_{Z^2|Z^1}(x^2 \mid x^1)}{\partial x^1}$$

That is, we ration less in Period 1 if a small increase in Period 1 payment leads to large increase in the probability of having agents left that have low Period 2 total costs.

It is not always the case that we will strictly increase the payments in Period 1, however. We have already seen this, cf. our analysis of the case with less lucrative markets, $V > W$ and a constant production level $y = z$. In this case, there was no reason to ration less in the first period. We could simply ration away the too costly types in Period 2.

In the case of two-dimensional types and varying production levels, however, the situation is more complicated because it is a matter of not only how much we ration but also in which direction of the type space we ration, and this depends on the production levels. Rationing in Period 1 is reduced in this case because the agent types that may not be useful in Period 1 may be useful in Period 2. We illustrate this in Fig. 16. Here the share of required product 2 increases from Period 1 to Period 2. In Period 1, we are therefore mainly interested in eliminating types with relatively large c_1 costs. However, these are precisely the types that are particularly useful in Period 2, where

the demand for product 2 is higher such that we would like to remove types with relatively large c_2 costs.

The first-order condition for x^2 looks simpler

$$-G_{Z^2|Z^1}(x^2 \mid x^1) + (W - x^2)g_{Z^2|Z^1}(x^2 \mid x^1) = 0$$

leading to

$$x^2 = W - \frac{G_{Z^2|Z^1}(x^2 \mid x^1)}{g_{Z^2|Z^1}(x^2 \mid x^1)}$$

The amount of rationing in Period 2 depends, as always, on the (inverse) hazard rate of the total costs. The hazard rate, however, depends on what we have learned from Period 1. It hereby depends both on how harshly we have rationed in Period 1 and on which types have been eliminated. These are the dependencies we have studied at length in the previous sections and that give rise to the mix stickiness. From the point of view of Period 2 alone, Period 1 has served as an *exploration period*, and Period 2 is the *exploitation period*.

Mix interaction

We have emphasized in the dynamic analysis above how the harshness of Period 1 rationing depends on 1) the value of Period 2 production and b) the production mix in Period 2. The more valuable Period 2 is and the more the Period 2 mix deviates from the Period 1 mix, the less we ration in Period 1.

If there are several possible production plans in the two periods, the principal can choose not only to adjust the harshness but also the direction of the rationing in Period 1.

When we have fully rational, long-sighted agents that are not subject to limited liability constraints in each period, the interaction is particularly clear. The principal will already in Period 1 ration against types that are not cost efficient in the production of the combined output of Period 1 and Period 2. If Period 2 production is large and highly valuable, for example, the Period 1 adaption of types to production plays only a minor role - Period 1 is serving mainly to explore the type space with the aim of supporting Period 2 usage.

Myopic agents with limited liability severely reduce the possibility of adjusting the Period 1 mix to prepare for Period 2 production. When the agent is myopic, the future mainly impacts how harshly we ration in Period 1. Some interaction of the mixes is still possible, however, as we shall now illustrate in our example.

Numerical example (ctd) To illustrate how the choice of mix in the two periods may interact, we can look at the results in Table 1. Assume that the principal in Period 1 is indifferent between any of the production plans

51

$$Y = \{(2,0), (1.8, 0.2), (1.6, 0.4), \ldots, (1,1)\}$$

Now, imagine that in Period 2, the principal needs $z = (1.8, 0.2)$ and that the Period 2 value is $W = r = 1$. In this case, the optimal mix for the principal is $y = (1,1)$ in Period 1, since then the expected profits from Period 2 is 0.21. If, however, the value in Period 2 is $W = r = 2$, it would be better for the principal to choose $y = (1.8, 0.2)$ already in Period 1 since this would (after a rescaling) correspond to the $z = (1,1)$ solution in Table 1.

We therefore see that it would be optimal for the principal to adjust the Period 1 mix to the Period 2 mix when Period 2 is very valuable. If, however, Period 2 is not too valuable, such that severe rationing is called for in Period 2 anyway, it would be better to deviate from the Period 2 mix in Period 1 since this makes rationing in Period 2 more effective.