ELSEVIER

European Economic Review 47 (2003) 505–520

EUROPEAN ECONOMIC REVIEW

www.elsevier.com/locate/econbase

In-house competition, organizational slack, and the business cycle

Rudolf Kerschbamer[a,b,c,*], Yanni Tournas[d]

[a] Department of Economics, University of Vienna, Hohenstaufengasse 9, A-1010 Vienna, Austria
[b] Department of Economics, University of Innsbruck, Universitätsstrasse 15, A-6020 Innsbruck, Austria
[c] Center for Economic Policy Research, 90-98 Goswell Road, London EC1V 7RR, UK
[d] Department of Economics, Rutgers University, New Brunswick, NJ, USA

Received 1 September 2000; accepted 1 October 2001

Abstract

Multiplant firms pit their facilities against each other for production assignments. The present paper studies the consequences of this practice in a model where production is limited by capacity constraints and asymmetric information allows facilities to accumulate slack. It shows the amount of slack per unit of output to be pro-cyclical. Indeed, as capacity constraints become more acute in economic booms, the power of in-house competition for quota assignments is reduced and slack per unit of output increases, while the opposite is true in downturns. Moreover, in downturns firms may use higher cost facilities even when lower cost plants are not running at capacity since this boosts X-efficiency in low-cost plants. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: D20; D82; F23; L22

Keywords: Competition; Organizational slack; Capacity; Demand uncertainty

Each of our thirteen production facilities is on the run from plant closure. In former times our plant in Gislaved, Sweden, was the tailender. Now the Semperit plant in Traiskirchen, Austria, holds this critical position.

(Dieter von Herz, spokesman for the German tire giant Continental AG) [1]

* Corresponding author. Tel.: +43-1-42-77-374-29; fax: +43-1-42-77-9374.
E-mail address: rudolf.kerschbamer@univie.ac.at (R. Kerschbamer).

[1] The source of the original German citation ("Alle unsere dreizehn Werke rennen, um einer Schließung zuvorzukommen. Früher hatte Gislaved in Schweden die rote Laterne. Jetzt ist Traiskirchen in diesem gefährlichen Bereich".) is an article by Lampl and Sklenar in issue 28/96 of the Austrian weekly *News*.

PII: S0014-2921(02)00185-X

1. Introduction

During the European car industry's most recent recession, the chairman of the board of the German tire manufacturer Continental AG threatened to allocate half of the production quota of its Austrian subsidiary Semperit to the Czech plant Barum. Afraid of losing the production right for 2 million tires per year (the former quota was 4 million) the managing director of Semperit promised cost savings of about 50 million Euro within two years. Only a few months later, the headquarters of the British brake manufacturer Wabco-Westinghouse made use of a similar strategy. It threatened to reduce the output quota of its Austrian production facility a second time, after having allocated part of its quota to a British plant a year before. As in the Continental-Semperit case the management of the Austrian Wabco-Westinghouse plant reacted with a significant downward revision of projected costs. [2]

The corporate practice of playing different facilities against each other is even more common in the United States, where it is known as "whipsawing". U.S. companies active in the automotive industry appear to have been among the first to adopt these methods. [3] Other sectors in which multiplant firms produce fairly homogenous goods and hold considerable overcapacity soon followed. [4]

The pressure that multiplant firms put on individual facilities during downturns has been discussed prominently in the popular press, but has not been investigated in the academic literature. The present paper seeks to fill this gap. It studies the consequences of such pressure for the internal efficiency of multiplant firms over the business cycle. Specifically, it shows two effects: (i) that the amount of slack per unit of output fluctuates pro-cyclically; and (ii) firms may use higher cost facilities even when lower cost plants are not running at capacity (as occurs, in particular, in recessions).

In our analysis, we consider a model that has the following properties: A multiplant firm is faced with random demand for its output. Each of the firm's facilities has a fixed capacity. Because of asymmetric information between the headquarters and

[2] The details of the Continental-Semperit example originate in articles in Austrian printed media, including "Reifenwechsel als Druckmittel" on July 6, 1996, in the daily *Der Standard*, "Das Drama Semperit" in issue 29/1996 of the weekly *Wirtschaftswoche*, and "Semper it—wie lange noch?" in issue 29/1996 of the weekly *Profil*. For our second example see, for instance, the article "Die Bremsen noch unter Kontrolle?" in issue 11/1996 of the Austrian business magazine *Trend*.

[3] Often cited early examples involve Chrysler explicitly playing off its Toledo, Ohio, Jeep plant against a rival plant in Kenosha, Wisconsin, in 1986, and General Motors pitting its two big-car assembly plants against each other in 1991 (cf., e.g., the articles by Bryant and by Hayes in the December 19, 1991, issue of *New York Times*).

[4] A topical example is General Electric (GE): Late in 1999, GE announced that $65 million in cost savings was needed at its refrigerator-freezer manufacturing plant in Bloomington, Indiana. Otherwise, half the production would be moved to a plant in Celaya, Mexico. The Bloomington plant promised to implement cost savings worth $40 million. At about the same time, GE was threatening its "Appliance Park" in Louisville, Kentucky, with a similar extortion scheme. Again, the plant reacted with a substantial downward revision of projected costs. GE agreed to keep the threatened production in Louisville. However, it moved half of Bloomington's output to Mexico saying that the plant's $40 million cost-savings proposal did not meet the given performance goal. (For a summary of events see the feature story "GE Brings Bad Things to Life" in the February 12, 2001, issue of *The Nation*.)

the individual facilities, plant stakeholders (in particular managers and workers) can dissipate corporate resources through slacking, perquisites, empire building, and the like. Thus, asymmetric information causes internal inefficiencies, or slack, in the sense that the firm is producing above the technically efficient production isoquant.

Our first result shows that in this model the amount of slack per unit of output fluctuates *pro-cyclically*. Indeed, as capacity constraints become less acute in economic downturns, idle capacities foster in-house competition among plants for higher production quotas. Because of this, slack per unit of output decreases. Exactly the opposite is true for boom periods, where demand exceeds the capacity available within the firm. In those periods, slack per unit of output increases because high demand reduces the headquarters' ability to instigate in-house competition.

Our results also show that during downturns production is not necessarily assigned to the cheapest plant. Indeed, a plant may be allowed to produce even if it has the highest production cost and demand is so low that the entire quantity could be produced without employing the facility. Intuitively, the systematic exclusion of a given plant from the production assignment process impedes in-house competition, and this increases the amount of slack in the remaining plants. Moreover, a new plant might be built even if it is known to have a considerable cost disadvantage in the future. Similarly, an old plant might be kept alive even if it is unprofitable. The explanation suggested by our analysis is that multiplant firms use their less efficient facilities to create a credible threat that production will be allocated to them if the more efficient ones accumulate too much slack.

Both of our model's predictions are supported by empirical evidence that is reviewed in the body of the paper. There, we also discuss alternative explanations for the two empirical regularities that have been offered in the theoretical literature. From a modeling perspective, the present work is most closely related to the second-sourcing literature (see, e.g., Anton and Yao, 1987; Demski et al., 1987; Riordan and Sappington, 1989) and to those papers in the literature on procurement and regulation that show how a carefully designed allocation of production to plants can help to reduce information cost (cf., for instance, Anton and Gertler, 1988; Auriol and Laffont, 1992; Dana and Spier, 1994). However, in contrast to the present paper, in this literature the quantity to be produced is exogenously given and capacity choice is no issue. Although Riordan (1996) is an important exception in this respect, our analysis still differs significantly in that (a) demand is random, and (b) the plants' cost distributions are asymmetric. The first difference drives the result that internal slack fluctuates pro-cyclically; and the second difference is responsible for the result that multiplant firms use high-cost facilities even when low-cost plants are not running at capacity.

Below, we first introduce the model and offer a formal statement of the headquarters' maximization problem. Next, in Section 3, we characterize optimal contracts, capacities, and production assignments. There, we also review empirical evidence that validates the result that multiplant firms use higher cost facilities as a threat against slack in lower cost facilities. Section 4 examines the relationship between the business cycle and operational slack and looks at evidence in support of the result that demand and slack are positively correlated. Section 5 concludes.

2. The model

We take as our model a firm that can sell at most X units of some final good at the price p_x. The price p_x is fixed, the quantity demanded at that price, X, is a random variable that has full support on some interval $[\underline{X}, \overline{X}]$, where $0 \leqslant \underline{X} < \overline{X} < \infty$. The firm has the option to produce the final good in two facilities indexed by A and B. The facilities are run as profit centers and each acts as a single agent. In order to produce, the facilities need capacities. We denote the price per unit of capacity by p_k and the amount of capacity placed at the disposal of facility i $(=A, B)$ by k^i. Each unit of capacity allows a facility to produce up to one unit of output at a constant cost of at least c^i. That is, c^i is the technically feasible minimum cost for plant i. Each c^i a priori belongs to $C^i = \{c^i_L, c^i_H\}$, where $c^i_H - c^i_L = \Delta^i > 0$.[5] The a priori probability that $c^i = c^i_m$ $(m = H, L)$ is denoted by r^i_m. The cost parameters c^A and c^B might be positively but imperfectly correlated. That is, defining $q^i_m \equiv \text{Prob}\{c^j = c^j_L | c^i = c^i_m\}$ for $\{i, j\} = \{A, B\}$ and $m \in \{L, H\}$, it is assumed that

Assumption 1. $1 > q^i_L \geqslant q^i_H > 0 \; \forall i \in \{A, B\}$.

The objective of each facility is to maximize the expected gain from dealing with the headquarters. This gain, or wealth, is given by $t^i - c^i x^i$, where t^i denotes the transfer from the headquarters to facility i, while x^i denotes the quantity produced by this facility. We assume that the wealth is dissipated within the facility through slacking, perquisites, overstaffing, and other forms of at-the-expense-of-the-firm behavior. In other words, this wealth causes X-inefficiency, or slack, in the sense that from the firm's perspective total production cost exceeds technically feasible minimum cost. We assume that the facilities are protected by limited liability so that their wealth is at least 0 ex post.[6]

The headquarters' objective is to maximize expected profit. Profit is given by min $\{x^A + x^B, X\} p_x - t^A - t^B - (k^A + k^B) p_k$.

The time and information structure is as follows: The binary supports of the plant-specific minimum-cost parameters and the support of demand are common knowledge to all parties involved and all share the same prior on $C^A \times C^B$ and on $[\underline{X}, \overline{X}]$. At Stage 1 the headquarters purchases capacity and allocates it among the two facilities. Then she designs the contracts, specifying the production quotas assigned to the facilities and the associated transfers. Later, at Stage 2, demand and minimum costs are drawn from their respective distributions. Demand becomes publicly observable and

[5] The model can easily be extended to allow for more than two types. Although the exposition is messier, the methods and results are essentially the same as in the model here. Similarly, introducing a (horizontally shifting) downward sloping demand curve would complicate the analysis by introducing a pricing decision, without changing any of the results.

[6] There is also a technical reason for introducing ex post individual rationality constraints: From Demski and Sappington (1984) and Crémer and McLean (1985) we know that, with interim individual rationality, any level of correlation in the cost parameters enables the headquarters to extract all the informational rents. This is an artifact of the convenient assumptions of risk neutrality and unlimited punishment. Our ex post constraints enable us to avoid this artificial result.

verifiable. The cost-parameter c^i, however, is privately observed by facility i. After having learned their c^is, the facilities simultaneously make cost reports to the headquarters.[7] These reports become publicly observable. Then the headquarters assigns the production quotas (according to the contract) to the facilities. Now the facilities produce. The quantities produced then become publicly observable and verifiable, and contractual terms are carried out.

What is the optimal contract to be offered by the headquarters at Stage 1? By the revelation principle we can restrict attention, without loss of generality, to contracts of the form $\{x^i(c^i,c^j,X),t^i(c^i,c^j,X)\}$ for $\{i,j\}=\{A,B\}$ where $c^i\in C^i$, $c^j\in C^j$ and $X\in[\underline{X},\overline{X}]$. Here, $x^i(c^i,c^j,X)$ is the output level required of facility i if the cost reports are c^i and c^j and the demand realization is X; $t^i(c^i,c^j,X)$ is the associated transfer, provided facility i produces $x^i(c^i,c^j,X)$. In the sequel, we put the reports into subscripts and omit demand as an argument in these functions (e.g., $x^i_{mn}=x^i(c^i_m,c^j_n,X)$). No confusion should result. With this convention and the definition $u^i_{mn}\equiv t^i_{mn}-c^i_m x^i_{mn}$, where $m,n\in\{L,H\}$, we can equivalently represent each contract by a vector of 8 functions of the form: $(u^i,x^i)=((u^i_{LL},x^i_{LL}),\ldots,(u^i_{HH},x^i_{HH}))$. In what follows we denote a contract combination $\{(u^i,x^i)\}_{|i\in\{A,B\}}$ as (u,x).

We now turn to incentive compatibility, individual rationality, and capacity constraints. Consider a contract (u^i,x^i). Suppose that facility $j\in\{A,B\}$, $j\neq i$, is known to truthfully announce its private information. For type m of facility i to honestly reveal its private information, we must have

$$(IC^i_m)\quad q^i_m u^i_{mL}+(1-q^i_m)u^i_{mH}\geqslant q^i_m[u^i_{nL}+(c^i_n-c^i_m)x^i_{nL}]$$
$$+(1-q^i_m)[u^i_{nH}+(c^i_n-c^i_m)x^i_{nH}],$$

where $\{m,n\}=\{H,L\}$. As is typical in this kind of adverse selection problem the binding IC constraint will be to prevent the facility with low cost from pretending to have high cost. A trivial solution to this problem is simply not to produce in a facility that claims to have high cost, even if demand is so high that the market cannot be served by letting the second facility produce at the capacity limit ($x^i_{Hm}=0$ for all $X\in[\underline{X},\overline{X}]$ and all $m\in\{L,H\}$). This solution is optimal if the price p_x is too low (or, for a given p_x, if the probability that $c^i=c^i_L$ is fairly high). To keep the problem interesting, and to avoid a lot of conditional statements, we introduce the following assumption.

Assumption 2. $p_x>c^i_H+\Delta^i r^i_L q^i_L/r^i_H q^i_H\ \forall i\in\{A,B\}$.

If a facility declares bankruptcy it gets a payoff of zero. Hence, for type m of facility i to respect the contract under all circumstances the inequality

$$(IR^i_{mn})\quad u^i_{mn}\geqslant 0$$

[7] Here and throughout this paper we assume that the facilities behave noncooperatively. If collusion among plants cannot be precluded the set of feasible contracts is further limited. For the design of collusion-proof contracts in correlated environments see Laffont and Martimort (1999) and the references therein.

must hold for all $n \in \{H, L\}$.[8] Obviously, facility i can comply with contractual terms only if the capacity constraint

$$(K^i_{mn}) \quad k^i \geqslant x^i_{mn}$$

is met for all $m, n \in \{H, L\}$. The headquarters wishes to maximize net revenue under incentive compatibility, individual rationality and capacity constraints. Formally, the headquarters' contracting problem at Stage 1 is

$$\begin{aligned}\operatorname*{Max}_{(u,x)} NR \\ &= \sum_{m\in\{H,L\}} r^A_m[q^A_m \min\{x^A_{mL} + x^B_{Lm}, X\} + (1 - q^A_m)\min\{x^A_{mH} + x^B_{Hm}, X\}]\, p_x \\ &\quad - \sum_{i\in\{A,B\}} \sum_{m\in\{H,L\}} r^i_m[q^i_m(c^i_m x^i_{mL} + u^i_{mL}) + (1 - q^i_m)(c^i_m x^i_{mH} + u^i_{mH})]\end{aligned}$$

subject to (IC^i_m), (IR^i_{mn}) and (K^i_{mn}) hold for all $i \in \{A, B\}$, $(m, n) \in \{H, L\}^2$, and $X \in [\underline{X}, \overline{X}]$. The first line in this program is gross revenue (quantity produced times the price), the second line represents total variable cost from the firm's perspective (t^i). Total variable cost consists of two terms, one ($c^i x^i$) standing for the technically feasible minimum cost, and the second (u^i) representing the X-inefficiency, or slack.

Solving the headquarters' contracting problem yields optimal values of u^i_{mn} and x^i_{mn} for all $i \in \{A, B\}$, $X \in [\underline{X}, \overline{X}]$ and $(m, n) \in \{H, L\}^2$. If we substitute the values for a given X in the net revenue function NR and subtract capacity costs, we obtain a reduced form profit function, conditional on X, k^A, and k^B. The headquarters' problem is then to choose k^A and k^B to maximize expected profit over all realizations of X.

3. Optimal contracts and capacities

3.1. The contracting problem

To facilitate the exposition of the headquarters' contracting problem, we concentrate (with little loss of generality) on a setting where one of the facilities (facility B) is at least as efficient as the other one in an ex ante sense. More precisely, we assume that $c^A_m \geqslant c^B_m$ for $m \in \{L, H\}$ and $r^A_H \geqslant r^B_H$. In this case optimal capacities are characterized by $k^A \leqslant k^B$.[9] We therefore take this into consideration in dealing with the contracting problem.

[8] The term "bankruptcy" should not be taken literally. The idea is rather that the firm cannot force a facility to deliver the good at below technically feasible minimum cost, since plant stakeholders can always quit without penalty.

[9] The formal proof for $k^A \leqslant k^B$ consists of solving the headquarters' contracting problem under the assumption $k^A \geqslant k^B$ and showing that this yields $k^A = k^B$.

Our first result (Lemma 1) characterizes the solution to this problem. In it reference is made to a symmetric and an asymmetric case. In the symmetric case $c_m^A = c_m^B$ and $r_m^A = r_m^B$ for $m \in \{L, H\}$. In the asymmetric case $c_m^A > c_m^B$ for $m \in \{L, H\}$. [10]

Lemma 1. *The solution to the headquarters' contracting problem is characterized by*
(i) $u_{HL}^i = u_{HH}^i = 0$ *and* $q_L^i u_{LL}^i + (1 - q_L^i) u_{LH}^i = [q_L^i x_{HL}^i + (1 - q_L^i) x_{HH}^i] \Delta^i > 0$ *for* $i = A, B$;
(ii) x_{mn}^A *as depicted in Table 1 for the symmetric and in Table 2 for the asymmetric case, and* $x_{mn}^B = \min\{X - x_{nm}^A, k^B\}$.

Proof. The proof uses standard techniques and is available upon request. □

Lemma 1 indicates that if a facility observes the high cost c_H, it is compensated only for the technically feasible minimum cost, while in the favorable environment c_L, it is able to capture corporate resources in the form of slack. The magnitude of resources appropriated by the facility in the favorable environment positively depends upon the output quota assigned to it if it claims to have high cost. This is so because a larger production level assigned to the high-cost plant creates a greater incentive for the low-cost plant to mimic the high-cost one. So, a higher level of slack must be conceded to the low-cost facility to keep it honest. This property of optimal contracts is important for our main results, and we will return to it later.

Let us turn to the allocation of production quotas. To explain this, we introduce a new category of variable cost referred to as the "virtual cost". Virtual cost differs from minimum cost (c^i) in that slack is taken into account. It differs from total variable cost (t^i) in that an ex ante rather than ex post point of view is taken. From an ex ante perspective, the amount of slack in the firm is increased if the output quota assigned to the high-cost facility is increased, while increasing the quantity assigned to the low-cost plant does not give rise to additional slack (see Property (i) in Lemma 1). So the virtual cost of the low-cost plant is just its technically feasible minimum cost, while the virtual cost of the high-cost plant is its minimum cost plus a term that measures the additional amount of slack the low-cost plant accumulates if the quantity produced by the high-cost one is increased by one unit. Denoting the virtual cost by v and adopting the convention that v_{mn}^i stands for the virtual cost in facility i if this facility observes and reports c_m while the second facility reports c_n, we can formally define the virtual cost as follows: $v_{LL}^i = v_{LH}^i = c_L^i$; $v_{HL}^i = c_H^i + (r_L^i q_L^i / r_H^i q_H^i) \Delta^i$; $v_{HH}^i = c_H^i + (r_L^i (1 - q_L^i) / r_H^i (1 - q_H^i)) \Delta^i$.

We are now in the position to explain the allocation of production quotas. In the symmetric case, minimum-, total-, and virtual-cost considerations all lead to the same decision. The resulting allocation is depicted in Table 1 and associated Fig. 1. Fig. 1 defines four different regions in the demand space, denoted by R_1 to R_4. Depending on the phase of the business cycle, that is, on the realization of demand, as well as on the capacities in the facilities, the firm may either have idle capacities (as in an extreme form in region R_1 and in a milder form in R_2 and R_3) or be capacity constrained (R_4).

[10] In the presentation we omit some intermediate cases (such as $c_m^A = c_m^B$ for $m \in \{L, H\}$ and $r_H^A > r_H^B$). The results for these cases correspond to a mixture between the results for the symmetric and those for the asymmetric case.

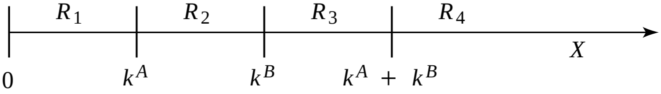

Fig. 1. Capacities and demand.

Table 1
Output allocation in the symmetric case

	x^A_{LL}	x^A_{LH}	x^A_{HL}	x^A_{HH}
R_1	$[0,X]$	X	0	$[0,X]$
R_2	$[0,k^A]$	k^A	0	$[0,k^A]$
R_3	$[X-k^B,k^A]$	k^A	$X-k^B$	$[X-k^B,k^A]$
R_4	k^A	k^A	k^A	k^A

In the presence of idle capacities, the headquarters will always allocate production to the lower cost plant and the higher cost one will carry idle capacity. So, if one plant reports high and the other low cost, all production up to the capacity constraint is allocated to the low-cost plant. And if both plants report either high or low cost, the distribution of production is indeterminate.

The asymmetric case behaves similarly, except that minimum-, total-, and virtual-cost considerations no longer lead to the same decision. Since contracts are designed ex ante and slack reduces the firm's profit, virtual costs will drive the decision. The resulting output allocation is depicted in Table 2. In it, reference is made to the variables ξ and δ. The term ξ applies to the situation where facility A has drawn the low and facility B the high cost, δ to the situation where both plants have high cost. The variables stand for the amount by which the virtual cost of facility A exceeds that of B in the respective situation, multiplied by the a priori probability of the situation occurring. Formally we have $\xi \equiv (v^B_{HL} - v^A_{LH})r^B_H q^B_H$ and $\delta \equiv (v^B_{HH} - v^A_{HH})r^B_H(1-q^B_H)$. So, $\xi >\ (=,<)\ 0$ if and only if $v^B_{HL} > (=,<)\ v^A_{LH}$, and similarly for δ. Taking this into account, Table 2 is easily understood: (i) If both plants report low costs, all production up to capacity is allocated to the technically more efficient plant B since this plant also has the lower virtual cost. (ii) If both plants report high costs, all production is given to the plant with the lowest virtual cost (again up to its capacity constraint). (iii) If one plant reports high and the other low costs, all production up to the capacity constraint is

Table 2
Output allocation in the asymmetric case

	x^A_{LL}	x^A_{LH}			x^A_{HL}	x^A_{HH}		
		$\xi < 0$	$\xi > 0$	$\xi = 0$		$\delta < 0$	$\delta > 0$	$\delta = 0$
R_1	0	0	X	$[0,X]$	0	0	X	$[0,X]$
R_2	0	0	k^A	$[0,k^A]$	0	0	k^A	$[0,k^A]$
R_3	$X-k^B$	$X-k^B$	k^A	$[X-k^B,k^A]$	$X-k^B$	$X-k^B$	k^A	$[X-k^B,k^A]$
R_4	k^A	k^A	k^A	k^A	k^A	k^A	k^A	k^A

allocated to the low-cost plant unless the underlying asymmetry in minimum costs so much favors the plant with the high cost report that it compensates for the increase in slack induced by an increase in the high-cost quantity.

Here note that plant A may be allowed to produce even if it is known to have the highest minimum cost in each environment (i.e., even if $c_L^A > c_H^B$) and even if demand is so low that the entire quantity could be produced without employing this facility. To see this possibility, suppose that $\xi > 0$ and $(c^A, c^B) = (c_L^A, c_H^B)$. Then producing in plant B is (in ex ante terms) more expensive than producing in A since the cost difference to B's favor is smaller than the additional slack it would accumulate if x_{HL}^B is increased by one unit. Thus, plant A is assigned to produce $\min\{X, k^A\}$ while B gets only the rest, which is zero if $X \leqslant k^A$.[11] We record this observation as

Proposition 1. *Suppose that $\xi > 0$ and $(c^A, c^B) \neq (c_L^A, c_L^B)$. Then the plant that reports low cost relative to the own cost-distribution wins the competition ($x_{LH}^i = \min\{X, k^i\}$) even if its cost report is higher than that of the rival in absolute terms ($i = A$ and $c_L^A > c_H^B$).*

Proof. Evident from Table 2. □

On an intuitive level an explanation for this result is that if the technically more efficient plant B knows that it is allowed to produce no matter what its cost report, it is able to accumulate a high level of fat. By contrast, if production is awarded to the less efficient plant A if B claims to have high cost, competition among the facilities limits the amount of slack. Proposition 1 has parallels in the second-sourcing literature referred to in the introduction, where it has been shown that the occasional replacement of a low-cost supplier (or, a more efficient incumbent) by a high-cost supplier (a less efficient entrant) might help to limit the informational rent of the former. Broadly similar effects are also at work in asymmetric auctions, where it is well known that it may pay the seller to favor a low value bidder in order to encourage aggressive bidding by others (see, for instance, Maskin and Riley (1985, 2000) or Rothkopf et al. (1997)).

Evidence supporting the prediction in Proposition 1 comes from one of the examples of whipsawing tactics discussed earlier:[12] In 1991, when General Motors (GM) pitted its big-car assembly plants in Ypsilanti, Michigan, and Arlington, Texas, against each other, most analysts expected that Ypsilanti would win the competition because it had a clear advantage in production costs and transportation costs between it and supplier plants in the Midwest were lower than those for the Arlington plant. This view was supported by reports of two studies, one by GM and one by an independent firm,

[11] While the output allocation in the symmetric case is ex post efficient and therewith renegotiation-proof, the output allocation in the asymmetric case is not. For the design of renegotiation-proof contracts see the articles published in the symposium "Incomplete Contracts and Renegotiations" in Vol. 34, 1990, of this journal, and the references therein.

[12] A good source for the evidence discussed in this paragraph (with quotations of workers and managers of the two facilities and many other interesting details) is a political science case study by Buchholz (1999).

Table 3
Shadow value of capacity in the symmetric setting

	Plant A
R_1	0
R_2 and R_3	$\overbrace{r_L^A(1-q_L^A)(c_H^B-c_L^A)}^{\text{Best realization benefit}}+\overbrace{r_L^B q_L^B \Delta^B}^{\text{Slack reduction benefit}}$
R_4	$p_x - r_L^A(c_L^A+\Delta^A) - r_H^A c_H^A$

which reportedly both concluded that production should be awarded to the Ypsilanti facility.[13] Aware of its clear cost disadvantage, the Arlington plant reacted promptly with adjustments in work practices, schedules, and so on. Nothing similar happened at the Ypsilanti plant, where managers and workers were sure that they would get the production assignment. When GM announced its decision against Ypsilanti in February 1992, many observers attributed the decision to politics. The present paper suggests an economic explanation: GM used the strategy of implementing an ex post inefficient output allocation to send a message to all its plants that they must keep slack under control, and that a clear cost advantage does not protect a facility from in-house competition. Here note that Ypsilanti's stakeholders, in particular workers and managers, might have been wrong in blaming GM for breach of faith: They had interpreted GM's announcement that "the plant managers' reports [would be] the key deciding factor" as meaning that the plant reporting the lowest absolute cost would get the production assignment. The present analysis suggests that the plant reporting low cost *relative to the own cost-distribution* wins the competition.

3.2. The capacity choice problem

The next step is to determine the capacities of the facilities. Optimal capacities are found by setting the expected shadow value of a marginal unit of capacity equal to the capacity price. As is easily verified, the shadow value of additional units of capacity under different demand realizations is as depicted in Table 4.

Consider first the symmetric case displayed in the simplified Table 3. In extreme downturns (in R_1) capacity places no restriction. So, adding an additional unit of this resource to one of the facilities creates no value.

If demand exceeds the capacity in plant A then the shadow value of an additional unit of capacity crucially depends on whether demand is higher (in R_4) or lower (in R_2 and R_3) than total capacity. Let us first assume demand is lower. In this case, an additional unit of capacity in facility i $(=A, B)$ allows the headquarters to produce an additional unit of output in i instead of producing it in $j \neq i$. This generates two kinds

[13] When the confidential GM study became public two years later it turned out that the reports were correct: The study concluded that by producing in Ypsilanti, GM could save \$74 million annually.

Table 4
Shadow value of capacity

	Plant A	Plant B
R_1	0	0
R_2	$\max\{0,\xi\}+\max\{0,\delta\}$	0
R_3	$\max\{0,\xi\}+\max\{0,\delta\}$	$r_L^B(c_H^A-c_L^B)+\max\{0,-\xi\}+\max\{0,-\delta\}$
R_4	$p_x-r_L^A(c_L^A+\Delta^A)-r_H^Ac_H^A$	$p_x-r_L^B(c_L^B+\Delta^B)-r_H^Bc_H^B$

of benefits: First, a *best-realization benefit*. This benefit arises in the situation where facility i reports the low and j the high cost because the additional unit of capacity in i allows production of an additional unit of output at the low cost c_L^i rather than the high cost c_H^j. Since the probability of the relevant event is $r_L^i(1-q_L^i)$, the impact of the best-realization benefit is given by $r_L^i(1-q_L^i)(c_H^j-c_L^i)$. Idle capacities have a second, more interesting advantage, which we call the *slack-reduction benefit*. This benefit arises in the situation where both facilities have drawn the low cost because the additional unit of capacity in i reduces the production quota assigned to j, if j claims to have high cost. This reduces the incentive of the low-cost realization of facility j to mimic the high-cost one, and therewith the slack. Since we are talking about a situation in which both facilities have the low cost, the quantity of interest is x_{HL}^j, and reducing this quantity by one unit leads to a reduction in j's fat by $q_L^j\Delta^j$ as can be seen from condition (i) of Lemma 1. Since the event that facility j is able to grow fat has probability r_L^j, the impact of the slack-reduction benefit of an additional unit of capacity in plant i is given by $r_L^jq_L^j\Delta^j$.

The rest of Table 3 is easily explained: In boom periods (Region 4) an additional unit of capacity in any of the facilities allows the headquarters to produce and sell an additional unit of output. Thus, the benefit of this unit is simply the market price of output minus total variable cost.

Allowing now for asymmetries, Table 4 shows that an extra unit of capacity in plant A can have positive value even in a situation in which this plant is known to have the highest cost in each environment ($c_L^A>c_H^B$) and in which there is excess capacity for sure ($\overline{X}\leqslant k^A+k^B$). The reason for this is again the slack in facility B, which is reduced by an increase in k^A. In terms of best-realization and slack-reduction benefits, the situation is as follows: If $c_L^A>c_H^B$ then the best-realization benefit, $r_L^A(1-q_L^A)(c_H^B-c_L^A)$, is unambiguously strictly negative. The slack-reduction benefit, $r_L^Bq_L^B\Delta^B$, however, remains positive. So, if in absolute terms, the slack-reduction effect exceeds the best-realization effect ($\xi>0$), then extra units of capacity in plant A have positive value despite the high production cost. Similar arguments for the case $c=(c_H^A,c_H^B)$ and $\delta>0$ lead to the result recorded in Proposition 2. Here note that in a first-best benchmark in which the facilities' technically feasible minimum costs are observable and verifiable, the shadow value of capacity in plant A would be zero if $c_L^A>c_H^B$ and $\overline{X}\leqslant k^A+k^B$. Thus, setting k^A equal to zero would be optimal in this benchmark if $c_L^A>c_H^B$, irrespective of the capacity-price p_k.

Proposition 2. *Suppose that* $\max\{\xi,\delta\}>0$. *Then there exists a range of capacity prices for which plant A is operated* ($k^A>0$) *even if it is ex ante known to have a sure cost disadvantage in the future* ($c_L^A>c_H^B$).

Proof. With the aid of Table 4 it is easily verified that $\lim_{p_k\to 0}(k^A,k^B)=(\overline{X},\overline{X})$ whenever $\max\{\xi,\delta\}>0$ (and that $k^A=0$ for any $p_k>0$ whenever $\max\{\xi,\delta\}\leqslant 0$). □

Evidence supporting the prediction in Proposition 2 comes from a recent econometric study analyzing the factors influencing a firm's choice between exit, downscaling, and relocation in reaction to a decline in performance. Based on a sample of Belgian firms, Pennings and Sleuwagen (2000) demonstrate that an important determinant in that choice is whether the firm operates a multinational network or not, and that multiplant multinationals keep their unprofitable facilities alive longer than singleplant national firms. The authors provide an option-theoretic explanation for this finding. They argue that by keeping their unprofitable facilities alive, multinationals preserve the opportunity to produce under more favorable market conditions, and that such an option has lower value for national enterprises. Other evidence consistent with our second result is discussed in the June 7, 1999, issue of *Business Week*, in an article headed "Exploiting Uncertainty". The article reports that U.S.-based Enron Corporation was about to open three gas-fired power plants in northern Mississippi and western Tennessee that would generate electricity at an incremental cost 50–70 percent higher than the industry's best facilities, making them unable to compete most of the time. It argued that the reason for this decision was that each plant gives a real-option to produce under favorable market conditions, and that, by building less efficient plants, a firm can save a lot on construction. We would provide a different explanation for both findings. Ours is similar to the option-theoretic explanation in observing that each plant gives the opportunity, but not the obligation to produce, but differs in regard to the source of value for that option. In the real-option approach, inefficient plants have value because they can be used when demand is high. This source of value is also present in our model, but an additional benefit arises in downturns, where additional facilities (even the technically less efficient) help to limit the amount of slack in other (more efficient) plants. This difference suggests that it is possible to empirically discriminate between the two explanations by checking whether the inefficient plants are operated only in good times (as predicted by the option-theoretic approach) or also in downturns.

To simplify the exposition, we concentrate in the following section on the symmetric case and denote the capacity level for this case by k ($=k^A=k^B$).

4. Organizational slack and the business cycle

The goal of this section is to analyze the effect of variations in product demand on the amount of internal slack. As noted earlier we get a sharp unambiguous result in this dimension:

Proposition 3. *The expected per-unit slack is increasing in production at given capacities in the facilities.*

Proof. Denote the (expected) amount of slack per unit of output produced by $\bar{u}(x)$. That is, $\bar{u}(x) \equiv \frac{1}{x}\sum_{i\in\{A,B\}} r_L^i[q_L^i x_{HL}^i + (1-q_L^i)x_{HH}^i)]\Delta^i$, where x_{HL}^i and x_{HH}^i are as shown in Table 1 for all $X \in [\underline{X}, \overline{X}]$, and where $x = \min\{X, 2k\}$. By symmetry, $r_L^A = r_L^B = r_L$, $q_L^A = q_L^B = q_L$, and $\Delta^A = \Delta^B = \Delta$. Inserting the optimal values for x_{HL}^i and x_{HH}^i ($i \in \{A, B\}$) from Lemma 1 into $\bar{u}(x)$ yields the increasing function

$$\overline{u}(x) = \begin{cases} r_L(1-q_L)\Delta & \text{for} \quad X \leqslant k \\ r_L(1-q_L)\Delta + r_L q_L \Delta \dfrac{2(X-k)}{X} & \text{for} \quad k < X < 2k \\ r_L(1-q_L)\Delta + r_L q_L \Delta & \text{for} \quad 2k \leqslant X \end{cases}$$

which is strictly increasing for $X \in (k, 2k)$. □

Since production equals demand up to the capacity limit, Proposition 3 can be interpreted as showing that X-inefficiency losses are less severe during downturns of the economy than in states of high demand. This is simply a consequence of the slack-reduction benefit just discussed: If demand is low then there exist idle capacities within the boundaries of the firm. Idle capacities intensify in-house competition among plants for higher production quotas. This intensified competition, in turn, reduces X-inefficiency. Since idle capacities carry not only a slack-reduction but also a best-realization benefit, and since there is no offsetting variable cost, expected total variable cost (which equals expected virtual cost) is increasing in production, too. We record this result as

Proposition 4. *Expected total variable cost per unit of output is increasing in production at given capacities in the facilities.*

Proof. Similar to that of Proposition 3 and therefore omitted. □

Convincing empirical evidence supporting the result that demand and slack are positively correlated comes from a recent case study by Sanchez and Schmitz (2000). The vantage point of this study is the world steel market collapse in the early 1980s, which led to a drastic fall in the demand for iron ore. The authors show that iron ore mines in countries insulated from the drop in demand had little or no productivity gains during the 1980s, while mines exposed to the shock typically had productivity gains ranging from 50 to 100 percent. The authors argue convincingly that the productivity increases were driven by continuing mines, using existing (up to capacity constraints) increasing-returns-to-scale technologies, improving their performance by reducing slack in order to escape the imminent production reduction.

Further evidence supporting our results on the behaviour of slack and variable cost over the business cycle is provided by a case study on the time pattern of productivity in the subsurface coal-mining industry in the United States. Prescott (1998) shows that productivity declined by a factor of two during one decade, but increased by a factor of three during another. He argues that the crucial factor in explaining these productivity movements is the price of coal substitutes. When the price of substitutes is high, plant stakeholders have massive incentives to resist cost cuts because coal production will be high anyway. The incentive to resist disappears when the price of substitutes is low, making the correlation between the price of substitutes and productivity highly negative. [14]

An alternative explanation as to why slack decreases in downturns is suggested by Schmidt (1997) in a theoretical paper on the impact of product market competition on managerial effort. In Schmidt's framework, an increase in the intensity of competition, modeled as a decrease in firm profits, generally has an ambiguous effect on effort: On the one hand, lower profits induce the management to work harder to avoid liquidation; on the other hand, lower profits reduce the owner's incentive to motivate the management appropriately. Schmidt argues that the former effect might dominate in downturns leading to the observed reduction in slack. Our explanation for pro-cyclical slack is not in conflict with this analysis but rather provides complementary arguments by focusing on the competitive pressure originating in idle capacities. [15]

5. Concluding remarks

Our analysis has shown that the pressure for internal efficiency exerted by multiplant firms has two consequences. The amount of slack per unit of output fluctuates pro-cyclically, and multiplant firms use higher cost facilities as a threat against slack in lower cost facilities.

The present paper leaves open an interesting and important question: In the examples discussed in the introduction, part of the cost savings induced by the headquarters' whipsawing tactics seems to come from cuts in wages, salaries, and fringe benefits as well as from reductions in labor force and changes in work practices. Here, a natural question to ask is, to what extent can these adjustments be summarized under the heading "slack-reduction"? When the cuts regard compensation-parts beyond those in an optimal ex ante contract, the use of this term seems justified. However, part of the changes might represent an inefficient ex post holdup of plant stakeholders

[14] The evidence mentioned thus far concerns particular time periods in specific industries. To the best of our knowledge there is no cross-industry econometric study illustrating the behavior of slack over the business cycle, presumably because of notoriously difficult measurement problems. An exception is Baily and Gersbach (1995) who find slack to be pro-cyclical. However, because of the long run focus of their study they did not investigate this observation.

[15] One of the anonymous referees offers an alternative explanation for counter-cyclical productivity. He/She argues that employees have decreased bargaining power in downturns as their outside options are not as good as during booms. Given this, any standard bargaining model in which firms negotiate with employees over wages and work practices would predict that during downturns firms are able to settle for lower wages and more efficient work practices. We think that this mechanism plays an important role, too, in explaining the empirical regularities.

disqualifying our interpretation. Thus, an important but difficult empirical task would be to disentangle the different sources for the observed cost savings. Such an investigation is complicated by the fact that efficient and inefficient variations in work practices, labor remuneration, etc. have to be carefully distinguished. Although this task still remains to be done, we are convinced that ex post holdup of employees is unlikely to be the main source for the immense productivity improvements found in the case studies discussed earlier.

Acknowledgements

We would like to thank David Besanko, Jim Dana, Raymond Deneckere, Morton Kamien, and Gerhard Orosel for helpful discussions as well as the editor, Xavier Vives, and three anonymous referees for their valuable suggestions. Special thanks are due to Jan Boone, Hans Gersbach, and Enrico Pennings for their help in providing a solid empirical basis for the paper. Thanks for useful comments are also due to seminar participants at Northwestern University, University of Regensburg, University of Innsbruck, University of Bologna, and University of Trento as well as to session participants at various conferences. Of course, the usual disclaimer applies.

References

Anton, J., Gertler, P., 1988. External markets and regulation. Journal of Public Economics 37, 243–260.

Anton, J., Yao, D., 1987. Second sourcing and the experience curve: price competition in defense procurement. Rand Journal of Economics 18, 57–76.

Auriol, E., Laffont, J.J., 1992. Regulation by duopoly. Journal of Economics and Management Strategy 1, 507–533.

Baily, M., Gersbach, H., 1995. Efficiency in manufacturing and the need for global competition. Brookings papers on Economic Activity, Microeconomics, pp. 307–347.

Buchholz, D., 1999. General Motors: A case study. Mimeo., Department of Political Science, Duke University, Durham, North Carolina.

Crémer, J., McLean, R., 1985. Optimal selling strategies under uncertainty for a discriminant monopolist when demands are interdependent. Econometrica 53, 345–361.

Dana, J., Spier, K., 1994. Designing a private industry: Government auctions and endogenous market structure. Journal of Public Economics 53, 127–147.

Demski, J., Sappington, D., 1984. Optimal incentives with multiple agents. Journal of Economic Theory 33, 152–171.

Demski, J., Sappington, D., Spiller, P., 1987. Managing supplier switching. Rand Journal of Economics 18, 77–97.

Laffont, J.J., Martimort, D., 1999. Mechanism design with collusion and correlation. Mimeo., Université des Sciences Sociales, Toulouse, France.

Maskin, E., Riley, J., 1985. Auction theory with private values. American Economic Review 75, 150–155.

Maskin, E., Riley, J., 2000. Asymmetric auctions. Review of Economic Studies 67, 413–438.

Pennings, E., Sleuwagen, L., 2000. Exit, downscaling or international relocation of production. Mimeo., Universitat Pompeu Fabra, Barcelona, Spain.

Prescott, E., 1998. Needed: A theory of total factor productivity. International Economic Review 39, 525–551.

Riordan, M., 1996. Contracting with qualified suppliers. International Economic Review 37, 115–128.
Riordan, M., Sappington, D., 1989. Second sourcing. Rand Journal of Economics 20, 41–58.
Rothkopf, M., Harstad, R., Fu, Y., 1997. Is subsidizing inefficient bidders actually costly? Mimeo., Department of Economics, Rutgers University, New Brunswick, NJ.
Sanchez, J., Schmitz, J., 2000. Threats to industry survival and labor productivity: World iron-ore markets in the 1980's. Research Department Staff Report 263, Federal Reserve Bank of Minneapolis, Minneapolis, MN.
Schmidt, K., 1997. Managerial incentives and product market competition. Review of Economic Studies 64, 191–213.