In-House Competition, Organizational Slack and the Business Cycle*

Rudolf Kerschbamer[+] and Yanni Tournas[++]

[+]Department of Economics, University of Vienna, and CEPR, London

[++]Department of Economics, Rutgers University

This Version: March 2001

Abstract: Multi-plant firms pit their facilities against each other for production assignments. The present paper studies the consequences of that practice in a model where production is limited by capacity constraints and asymmetric information allows facilities to accumulate slack. The amount of slack per unit of output is shown to be pro-cyclical. Indeed, as capacity constraints become more acute in 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 the firm may use higher-cost facilities even when lower-cost plants are not running at capacity since this boosts X-efficiency in low-cost plants.

JEL Classifications: D20, D82, F23, L22

Keywords: Competition, Organizational Slack, Capacity, Demand Uncertainty

---

*We would like to thank David Besanko, Jim Dana, Raymond Deneckere, Morton Kamien and Gerhard Orosel for helpful discussions, and 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 the Northwestern University, the University of Regensburg, the University of Innsbruck, the University of Bologna and the University of Trento, and to session participants at various conferences. Of course, the usual disclaimer applies.

"Each of our 13 production facilities is on the run from plant closure. In former times our plant in Gislaved, Sweden was the tail ender. Now the Semperit plant in Traiskirchen, Austria holds this critical position."

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

1 Introduction

During the latest recession in the European car industry the chairman of the Board of the German tire giant "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 two million tires per year (the former quota was four million) the managing director of Semperit Austria promised cost savings of about 50 million Euro within two years. Only a few months later the headquarters of the British brake giant "Wabco-Westinghouse" used a similar strategy. It threatened to reduce the output quota of its Austrian production facility a second time after having allocated part of the Austrian 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 off against each other is even more common in the U.S. 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

[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*.

[2]The details of the Continental-Semperit example originate in articles in various issues of Austrian printed media, including the article "Reifenwechsel als Druckmittel" on July 6th, 1996 in the daily *Der Standard*, the article "Das Drama Semperit" in issue 29/1996 of the weekly *Wirtschaftswoche*, and the article "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

sectors in which multi-plant firms produce fairly homogenous goods and hold considerable over-capacity soon followed.[4]

Whereas the popular press discusses in prominent feature articles the pressure that multi-plant firms put on individual facilities during downturns, this topic has not been investigated in the academic literature. The present paper seeks to fill this gap. It studies the consequences of headquarters' pressure for the internal efficiency of multi-plant firms over the business cycle. Specifically, we show two effects: the amount of slack per unit of output fluctuations pro-cyclically; and a firm may use higher-cost facilities even when lower-cost plants are not running at capacity, as it occurs, in particular, in recessions.

In our analysis we consider a model that has the following properties. A multi-plant firm is faced with stochastic demand for its output. Each of the firm's facilities or plants has a fixed capacity. Because of asymmetric information between the headquarters and the individual facilities, plant stakeholders (in particular managers and workers) can dissipate corporate resources in the form of slacking, perquisites, empire building, etc. 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 the amount of slack per unit of output produced fluctuates *pro-cyclically* in this model. Indeed, as capacity constraints become less acute in economic downturns, idle capacities foster in-house competition among plants for higher production

rival plant in Kenosha, Wisconsin in 1986, and General Motors explicitly pitting its two big-car assembly plants in Ypsilanti, Michigan and in Arlington, Texas against each other in 1991 (cf., e.g., the articles by Bryant and by Hayes in the December 19th, 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 the Bloomington plant's output quota 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 12th, 2001 issue of the U.S. magazine *The Nation*.)

quotas. Because of this, slack per unit of output decreases in downturns. Exactly the opposite is true for boom periods of the economy where demand exceeds the amount of total capacity available within the boundaries of the firm. In those periods, slack per unit of output increases because high demand reduces the headquarters' ability to instigate in-house competition. Thus, during boom periods of the economy total profits tend to be high since demand is high and capacity is fully utilized. In contrast, during downturns a multi-plant firm is able to improve its profitability by concentrating on the cost side. In such phases of the business cycle it can reduce organizational slack by making use of in-house competition.

Second, we show that during downturns of the economy production is not necessarily assigned to the cheapest plant. Indeed, a plant may be allowed to produce even if it is known to have always the highest production cost and even if demand is so low that the entire quantity could be produced without employing this 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 *ex ante* 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 multi-plant firms use their technically less efficient facilities as a credible threat that production will be allocated to those facilities, if the more efficient ones accumulate too much slack.

Both predictions of our model are supported by empirical evidence that is reviewed in the next section. There we also discuss other 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, and 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, and Dana and Spier 1994). However, in contrast to the present paper, the quantity to be produced is exogenously given and capacity choice is no issue in this literature. Although Riordan (1996) is an important exception in this respect, our analysis is significantly different in that (a) demand is random, and (b) the plants' cost distributions are asymmetric. The first difference drives our first result that internal slack fluctuates pro-cyclically; and the second difference is responsible for our second result that multi-plant firms use high-cost facilities even when low-cost plants are not running at capacity.

The paper proceeds as follows. The next section reviews the empirical evidence that relates to our results. Section 3 introduces the model and offers a formal statement of the headquarters' maximization problem. Optimal contracts, capacities, and production assignments are characterized in Section 4. Section 5 examines the relationship between the business cycle and operational slack, and Section 6 concludes.

2 Empirical Evidence

Although there are no direct empirical tests of the two predictions of our analysis, there is nevertheless ample evidence in support of these predictions. This evidence can be found in a diverse array of papers that investigate various questions and use different methods of analysis. First we provide evidence in support of the result that demand and slack are positively correlated over the business cycle, and then we review evidence that validates our second result that multi-plant firms use higher-cost facilities as a threat against slack in lower-cost facilities.

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 1980's. This collapse led to a drastic fall in the demand for iron-ore since the almost exclusive use of this metal is in steel production. The extent to which iron-ore mines where exposed to the demand shock differed considerably across countries. The authors show that mines in those countries insulated from the drop in demand had little or no productivity gains over the 1980's. Mines exposed to the shock, on the other hand, typically had productivity gains ranging from 50 to 100 percent. The authors argue convincingly that the productivity increases where 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.[5]

Further evidence supporting our first result 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 in one decade, but increased by a factor of three in another decade. 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 then 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.[6]

The evidence mentioned thus far concerns particular time periods in specific industries. Of course, one would prefer to have evidence from cross-industry econometric studies illustrating the behavior of slack over the business cycle. To our best knowledge there is no such study, presumably because of notoriously difficult measurement problems.[7] However, there is some indirect evidence: The macroeconomic literature on the behavior of productivity over the business cycle shows that aggregate productivity fluctuates counter-cyclically, when controlled for cyclical capacity utilization (cf., e.g., Basu 1996; for an overview of empirical evidence on the cyclical behavior of marginal costs see Rotemberg

[5] The object of this study fits our modeling framework particularly well: A considerable fraction of firms in the iron-ore market operate more than one mine, the steel collapse was an exogenous event from the viewpoint of iron-ore producers, and the demand reduction had (in the aggregate) to be met by a reduction in iron-ore quantity (and not only by a reduction in price) because the iron-ore input cost amounts to only about 10 percent of the steel's selling price.

[6]Evidence supporting the belief that plants reduce slack when they are in economic distress is also provided by Borenstein and Farrell (1999). The authors show that the increase in the stock-market value of gold mining companies induced by an increase in the price of gold is larger when the gold-price was initially low rather than high. The authors interpret this finding, which is inconsistent with a fairly general theoretical property of value maximization, as evidence confirming the view that increases in wealth will be dissipated in inefficiency.

[7]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.

and Woodford 1999). Of course, this aggregate evidence is not that easy to interpret because many different factors may contribute to productivity deteriorations in booms. This has already been observed by Mitchell (1941, p. 52) who notes that, as activity expands in upturns, "equipment of less than standard efficiency is brought back into use; the price of labor rises while its efficiency falls; the cost of materials, supplies and wares for resale advances faster than selling prices; (...) and all the little wastes incidental to the conduct of business enterprises grow steadily larger." The case studies summarized above indicate however, that slack has a major role to play in explaining counter-cyclical productivity.

The view that slack is an important source for the cost savings in downturns gets some support from evidence concerning one of the examples for the whipsawing tactics of multi-plant firms discussed earlier. In articles on the Continental/Semperit case (see the discussion at the very beginning of the paper) Austrian print media note that, during the fat years of the automotive industry, Semperit employees had enjoyed yearly pay increases that were, on average, 3% higher than those determined in the relevant collective bargaining agreements. At the time when Continental pitted Semperit against Barum for quota assignments, the wages paid by Semperit where, on average, 50%(!) higher than those paid for the same kind of labor in the immediate surroundings. On top of pay, Semperit workers enjoyed very generous secondary benefits as, for instance, a generous plant-specific pension scheme supplementing the Austrian state pension system; a plant-specific sickness insurance scheme providing employees and their relatives a higher level of benefits than the standard national health system at a lower premium; free access to a plant swimming pool, to two plant athletic grounds and to several recreation homes; free work clothes and free cleaning of those clothes; transportation from home to the plant and back at the distance-independent symbolic price of 8.50 ATS (ca. $0.50); arbitrary amounts of free milk; and so on.[8] Of course, one could invoke the efficiency-wage hypothesis to justify all these benefits as part of an optimal *ex ante* contract. But, does this theory really fit assembly-line work where performance is easily observed and verified?

[8]The article "Semper it – wie lange noch?" in issue 29/1996 of the weekly *Profil* discusses these and other privileges. Similar stories appeared in other magazines.

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 provides complementary arguments by focusing on the competitive pressure originating in idle capacities.

The virtues of bad times have also been recognized in the new growth theory: Aghion and Saint-Paul (1998) study optimal productivity growth under demand fluctuations in two alternative models, one in which productivity-improving activities are costly in terms of current production, and a second in which the cost of productivity improvements is independent of current production. They show that productivity improvements are counter-cyclical in the first but pro-cyclical in the second model, and that the results for the first model are consistent with empirical evidence whereas those for the second are not. Other new-growth-theoretic contributions deriving counter-cyclical productivity improvements include Davis and Haltiwanger (1990) and Hall (1991). A key difference between this line of research and the present paper lies in the forces driving the productivity improvements in downturns. In the empirically supported new growth models productivity increases in downturns because the opportunity cost of investing capital and labor in productivity improving activities is low when current production is low. By contrast, productivity increases in downturns in the present work because idle capacities intensify the in-house competition among plants for higher production quotas, and because this intensified competition reduces slack.[9]

[9] One of the anonymous referees offers an explanation for counter-cyclical productivity which is similar in spirit to the new-growth theoretic one. 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 in downturns firms are able to settle for lower wages and more efficient work practices. We think

We turn now to our second result that multi-plant firms use higher-cost facilities as a threat against slack in lower-cost facilities. Specifically, multi-plant firms may use high-cost facilities even when low-cost plants are not running at capacity, they may keep unprofitable facilities alive, and they may even build new inefficient plants. This prediction receives support from evidence on one of the examples for the whipsawing tactics of multi-plant firms discussed earlier:[10] When General Motors (GM) explicitly pitted its two big-car assembly plants in Ypsilanti, Michigan, and in Arlington, Texas, against each other in 1991, most analysts expected that the Ypsilanti facility would win the competition, because Ypsilanti had a clear advantage in production costs, and because 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 - which reportedly both concluded that production should be awarded to the Ypsilanti facility.[11] 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 up to the very last moment that they would get the production assignment. When GM announced its decision against the Ypsilanti facility in February 1992 many observers attributed this seemingly irrational 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. The message was that all facilities have to keep slack under control, and that a clear cost advantage doesn't protect a facility from in-house competition.

Additional evidence supporting our second result 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

that this mechanism plays an important role, too, in explaining the empirical regularities.

[10] 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).

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

Sleuwagen (2001) demonstrate that an important determinant of that choice is whether the firm operates a multinational network or not, and that (multi-plant) multinationals keep their unprofitable facilities longer alive than (single-plant) 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 issue June 7, 1999 of *Business Week*. In an article headed "Exploiting Uncertainty" the weekly reports that the U.S. enterprise *Enron Corporation* is about to open three gas-fired power plants in northern Mississippi and western Tennessee that generate electricity at an incremental cost 50% to 70% higher than the industry's best facilities, making them unable to compete most of the time. Again it is argued that the reason for this decision is 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. The present paper provides a different explanation for both findings. Our explanation is similar to the option-theoretic one in observing that each plant gives the opportunity, but not the obligation to produce. The explanations differ 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. Here an additional benefit arises in downturns, as (even technically less efficient) additional facilities help to limit the amount of slack in other, more efficient plants.

Our result about the desirability of maintaining an inefficient plant has also parallels in the second-sourcing literature referred to above, 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 and 1999, or Rothkopf et al. 1997).

3 The Model

We consider a simple model of 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 $0 \leq \underline{X} < \overline{X} < \infty$.[12] 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 of them acts as a single agent. In order to produce the facilities need capacities. Capacities are purchased and installed by the headquarters at the outset. We denote the price per unit of capacity by p_k and the amount of capacity placed at the disposal of facility i 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 (or "potential") cost for plant i. Each c^i *a priori* belongs to $C^i = \{c^i_L, c^i_H\}$, where $c^i_H - c^i_L = \Delta^i > 0$.[13] 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.[14] That is, defining $q^i_m \equiv \text{Prob}\{c^j = c^j_L | c^i = c^i_m\}$ for

[12]Such a demand function arises, for example, when the firm is a monopoly selling a homogeneous product to a random sample X of homogeneous potential buyers each desiring to purchase at most one unit of the product. All buyers have the same quasilinear preferences of the form $p_x q_i + m_i$, where $q_i \in \{0, 1\}$ is consumer i's consumption of the good under study and m_i is his consumption of the numeraire commodity. This is a fairly special specification, of course. (For a reference to a more realistic specification yielding a qualitatively similar demand function see Footnote 5 above.) Introducing instead a standard downward sloping demand curve would complicate the analysis by introducing a pricing decision, without changing any of the results, however. In the modified model movements in demand, now modeled as horizontal shifts in the demand curve, not only lead to quantity- but also to price changes. Since the headquarters (to be introduced below) still tries to minimize the transfers to the facilities for each quantity requirement, the analysis presented below also applies to the modified model.

[13]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 for the simple binary model considered here.

[14]Our current (standard) proof techniques allow for a moderate degree of negative correlation, but they fail if the amount of negative correlation becomes too large. We would expect our main results to hold more generally, for arbitrary degrees of correlation (provided the correlation is not perfect), but did not try to find more general (non-standard) techniques to check for that as we regard the case where the potential costs of plants that produce the same output are (strongly) negatively correlated as rather unrealistic.

$\{i, j\} = \{A, B\}$ and $m \in \{L, H\}$, it is assumed that

Assumption 1: $1 > q_L^i \geq q_H^i > 0 \qquad \forall i \in \{A, B\}$.

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 appropriated by a facility is dissipated within this facility in the form of 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 actual production costs (from the firm's perspective) exceed technically feasible minimum costs.[15] We assume that the facilities are protected by limited liability so that their wealth is at least 0 *ex post*.[16]

The objective of the headquarters 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 potential-cost parameters and the support of demand are common knowledge

[15] The present paper specifies neither who among the plant stakeholders dissipates the resources appropriated by a facility, nor the concrete form resource dissipation takes. Applying instead the standard principal-agent approach where the main problem is to prevent a work- (or effort-) averse manager from shirking would make the model less realistic without changing any of the results (see Kerschbamer and Tournas 1997 for that). We would expect that in practice not only managers but also workers profit from wealth dissipation, for instance, because this makes the life of managers more comfortable (expense preference, preference for seeing pleased faces, etc). Also, reduced effort will in general be only one among many possible forms in which resources are captured. Other, probably more important forms include unprofitable but prestigious investments, inefficient work practices (such as too large crew sizes, excessive manning levels, use of too labor-intensive technologies etc), and generous secondary benefits for managers (free use of prestigious company cars, too generous golden parachutes, etc) and workers (generous plant specific pension schemes, free work clothes, free access for relatives to a plant kindergarten, etc).

[16] 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 evade this artificial result.

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 potential costs are drawn from their respective distributions. Demand becomes publicly observable and verifiable. Potential cost c^i, however, is privately observed by facility i. After having learned their c^is the facilities simultaneously make cost reports to the headquarters.[17] These reports become then publicly observable. Then the headquarters assigns the production quotas (according to the contract) to the facilities. Now the facilities produce. The quantities produced become then 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 $x^i(c^i, c^j, X)$. To keep the notation symmetric, we adopt the convention that the first cost-report argument in x^i and t^i is the report of plant i while the second is the report of j. 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 additional 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.

[17] Here and throughout this paper we assume that it is common knowledge that the facilities behave non-cooperatively. 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.

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

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

where $\{m, n\} = \{H, L\}$. That is, truth telling must be Bayesian Incentive Compatible (IC) for the facility. As is typical in this kind of adverse selection problems the binding IC constraint will be to prevent the facility with low potential cost from pretending to have high cost. A trivial solution to this problem is simply not to produce in a facility if it claims to have high potential cost, even if demand is so high that the market can not 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: $p_x > c^i_H + \Delta^i r^i_L q^i_L / r^i_H q^i_H \qquad \forall i \in \{A, B\}$.

If a facility declares bankruptcy it gets a payoff of 0.[18] Hence, for type m of facility i to respect the contractual terms under all circumstances the inequality

$$(IR^i_{mn}) \qquad\qquad u^i_{mn} \geq 0$$

must hold for all $n \in \{H, L\}$. That is, obeying the contractual terms must be *ex post* Individually Rational (IR) for facility i. Obviously, facility i can comply with contractual terms only if

$$(K^i_{mn}) \qquad\qquad k^i \geq x^i_{mn}$$

holds for all $m, n \in \{H, L\}$. That is, the quantity the facility is required to produce must not exceed its capacity. The headquarters wishes to maximize net revenue under incentive

[18] The term "bankruptcy" should not be taken literally. The basic idea is rather that the firm cannot force a facility to deliver the good at below technically feasible minimum cost as plant stakeholders can always quit without penalty. As noted earlier (see Footnote 16 above) there is also a technical reason for imposing this constraint.

compatibility, individual rationality and capacity constraints. Formally, the headquarters' contracting problem at Stage 1 is:

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

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 $X \in [\underline{X}, \overline{X}]$.[19] The first line in this program is gross revenue (quantity produced times the price), the second line represents actual production cost from the firm's perspective (t^i). Actual production cost consists of two terms, one ($c^i x^i$) standing for the technically feasible minimum cost, and a second term (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.

4 Optimal Contracts and Capacities

The Contracting Problem

We begin the analysis with the headquarters' contracting problem. To facilitate the exposition we concentrate (with little loss of generality) on a setting where one of the facilities (facility B) is at least as efficient as the second one in an *ex ante* sense.[20] More precisely, we assume that $c^A_m \geq c^B_m$ for $m \in \{L, H\}$ and $r^A_H \geq r^B_H$. In this case optimal

[19]Recall that the x^i_{mn} and the u^i_{mn} are functions: The headquarters allocates the production quotas and sets the transfers after having learned the realization of demand.

[20]The missing cases can be analyzed with the same techniques and yield the same qualitative results. We have decided not to present them because they allow no unambiguous ranking of capacity levels. So lots of subcases and sub-subcases have to be considered, burdening the exposition with details that are unimportant for our main findings.

capacities are characterized by $k^A \leq k^B$.[21] We therefore take this into consideration in dealing with the contracting problem.

Our first result (Lemma 1) characterizes the solution to this problem. In this result reference is made to a symmetric and an asymmetric case. In the symmetric case $c^A_m = c^B_m$ and $r^A_m = r^B_m$ for $m \in \{L, H\}$. In the asymmetric case $c^A_m > c^B_m$ for $m \in \{L, H\}$.[22]

Lemma 1 *The solution to the headquarters' contracting problem is characterized by*

(*i*) $u^i_{HL} = u^i_{HH} = 0$ *and* $q^i_L u^i_{LL} + (1 - q^i_L) u^i_{LH} = [q^i_L x^i_{HL} + (1 - q^i_L) x^i_{HH}]\Delta^i > 0$ *for* $i = A, B$;

(*ii*) x^A_{mn} *as depicted in Table 1 for the symmetric and in Table 2 for the asymmetric case, and* $x^B_{mn} = \min\{X - x^A_{nm}, k^B\}$.

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

Lemma 1 indicates that a facility gets only a compensation for the technically feasible minimum cost if it observes and reports the high cost-parameter c_H, while it is able to capture corporate resources in the form of slack in the more favorable environment c_L. The magnitude of resources appropriated by the facility in the favorable environment positively depends upon the output quota assigned to the facility if it claims to have high potential cost. This is easily understood. The low-cost plant is allowed to grow fat, since otherwise, it would always have an incentive of mimicking the high-cost one. Reducing the output quota assigned to the high-cost plant reduces the incentive for the low-cost plant of mimicking the high-cost one and therewith the amount of fat the low cost facility must be allowed to accumulate. 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 allocation we introduce

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

[22] In the presentation we omit some intermediate cases (such as $c^A_m = c^B_m$ for $m \in \{L, H\}$ and $r^A_H > r^B_H$; or $c^A_m = c^B_m$, $c^A_n = c^B_n$ for $m, n = \{L, H\}$ and $r^A_H \geq r^B_H$). The results for these cases correspond to a mixture between the results for the symmetric and those for the asymmetric case.

a new category of variable cost referred to as the "virtual cost". Virtual cost differs from potential cost (c^i) in that slack is taken into account. It differs from actual 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 potential cost while the virtual cost of the high cost plant is its potential cost plus a term that measures the additional amount of slack the low cost plant must be allowed to accumulate 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^i_{mn} 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: $v^i_{LL} = v^i_{LH} = c^i_L$; $v^i_{HL} = c^i_H + (r^i_L q^i_L / r^i_H q^i_H)\Delta^i$; $v^i_{HH} = c^i_H + (r^i_L(1-q^i_L)/r^i_H(1-q^i_H))\Delta^i$.

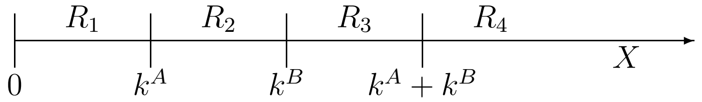

Figure 1: Capacities and Demand

	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

Table 1: Output Allocation in the Symmetric Case

We are now in the position to explain the allocation of production quotas. Consider first the symmetric case. In the symmetric case potential-, actual- and virtual-cost considerations all lead to the same decision. The resulting allocation is depicted in Table 1 and associated Figure 1. Figure 1 defines 4 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, and 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 regions R_2 and R_3), or be capacity constrained (R_4). In the presence of idle capacities the headquarters will always allocate production to the least-cost plant and the higher-cost one will carry idle capacity. So, if one plant reports high and the other reports low costs all production up to the capacity constraint is allocated to the low cost plant. And if both plants report either high or low costs the distribution of production is indeterminate.

	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

Table 2: Output Allocation in the Asymmetric Case

The asymmetric case behaves similarly, except that potential-, actual- and virtual-cost considerations do not longer lead to the same decision. Since contracts are designed *ex ante* and since slack reduces the firm's profit, virtual costs will drive the decision. The resulting output allocation is depicted in Table 2. In this Table reference is made to the variables ξ and δ. These variables are defined as $\xi \equiv (v^B_{HL} - v^A_{LH})r^B_H q^B_H$ and $\delta \equiv (v^B_{HH} - v^A_{HH})r^B_H(1 - q^B_H)$. In words, ξ is the difference in the virtual costs of the two plants in the situation where $(c^A, c^B) = (c^A_L, c^B_H)$ times the probability of this event occurring, and δ represents that term for the situation where $(c^A, c^B) = (c^A_H, c^B_H)$. 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 as this plant has also 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 costs and the other reports low costs all production up to the capacity constraint is allocated to the low cost plant unless the underlying asymmetry in potential 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 potential costs 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 $X \leq k^A$. Similar arguments for the case where $\delta > 0$ and $(c^A, c^B) = (c_H^A, c_H^B)$ lead to the following result:

Proposition 1 *Iff either $\xi > 0$ and $(c^A, c^B) = (c_L^A, c_H^B)$, or $\delta > 0$ and $(c^A, c^B) = (c_H^A, c_H^B)$ then plant A is allowed to produce even if it is known to have always higher potential costs ($c_L^A > c_H^B$) and even if demand is so low that there remains excess capacity in plant B ($k^B < X$).*

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 is, it is able to accumulate a high level of fat. By contrast, if production is awarded to the worse plant A if B claims to have high cost, competition among the facilities for the right to produce limits the amount of slack.[23]

Proposition 1 sheds new light on the evidence (discussed in the introduction) regarding

[23] 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). We do not regard this as a problem in the present context: The relationship between the headquarters of a multi-plant firm and the individual facilities is a repeated relationship. It is well known that a player in a repeated relationship (here the headquarters) may be willing to incur a short run cost (the cost of implementing an *ex post* inefficient output allocation) in order to get a higher payoff (the payoff in the commitment solution characterized in Lemma 1) in the long run.

GM's decision to award production to the Arlington plant despite that plant's clear cost disadvantage relative to the Ypsilanti facility. The explanation suggested by the present analysis is that GM used the *ex post* inefficient output allocation as a means to send a message to all its plants. The message was that all facilities have to keep slack under control, and that a clear cost advantage doesn't guarantee a production assignment.

The present analysis also indicates that Ypsilanti's stakeholders, in particular workers and managers, might have been wrong in blaming GM for breach of faith (see Buchholz 1999 for quotes and other details): They had interpreted GM's announcement that "the plant managers' reports (would be) the key deciding factor" as meaning that the plant with the lowest reported cost would get the production assignment. The present analysis suggests another interpretation: The plant which reports low cost *relative to the own cost distribution* wins the competition.

The Capacity Choice Problem

The next step is to determine the optimal capacity levels for the facilities.[24] Optimal capacities are found by setting the expected shadow value of a marginal unit of capacity equal to the capacity price.[25] The shadow value of additional units of capacity under different demand realizations is as depicted in Table 4.

Corollary 1 *The shadow value of a marginal unit of capacity is as depicted in Table 4.*

Proof Follows immediately from Lemma 1. □

[24] Closely related to this decision is the choice of the number of facilities. In the present paper we have assumed that the firm can operate at most two plants (with the interpretation that plant i is operated if and only if $k^i > 0$). An obvious extension would be to introduce a fixed cost of creating a new facility (without such a cost there is a natural tendency in our model to operate infinitely many infinitely small plants), and to solve for both the optimal number of facilities and the optimal capacity level for each facility. This extension does not yield any interesting additional insights, however.

[25] Note that we assume that the headquarters knows the supports of the potential-cost distributions of the two plants (but not the actual cost realizations, of course) even before the plants are built (capacity investments are made). That this modeling detail is not that unrealistic, and that a new plant might be built even if it is *ex ante* known to have a considerable cost disadvantage is confirmed by the evidence (discussed in the introduction) regarding *Enron Corp.*'s decision to build low efficiency new plants.

Consider first the symmetric case. For the symmetric case Table 4 simplifies to Table 3. This table is easily explained. In extreme downturns, that is, in Region 1, the headquarters can produce the whole output in whichever facility she wants. Since capacity places no restriction in this case, adding an additional unit of this resource to one of the facilities creates no value.

	Plant A
	best realization benefit — slack reduction benefit
R_1	0
R_2&R_3	$\overbrace{r_L^A(1-q_L^A)(c_H^B-c_L^A)}+\overbrace{r_L^B q_L^B \Delta^B}$
R_4	$p_x - r_L^A(c_L^A+\Delta^A) - r_H^A c_H^A$

Table 3: Shadow Value of Capacity in the Symmetric Setting

If demand exceeds the capacity in plant A then the shadow value of an additional unit of this resource crucially depends on whether demand is higher (in R_4) or lower (in R_2 and R_3) than total capacity. Let us assume first demand is lower. To see where the shadow value of an additional unit of capacity in this case comes from let us return to Table 1. If demand falls in Region 2 and both facilities report the same potential cost then the headquarters is indifferent between carrying out production in facility A and producing in plant B. An extra unit of capacity in A has therefore no value. The same holds if A has drawn the high and B the low potential cost since the headquarters prefers to have the whole output produced in B in this case. An extra unit of capacity in A has, however, value if A has drawn the low and B the high cost-parameter. In this case an additional unit of capacity in facility A allows the headquarters to produce an additional unit of output in A instead of producing it in B. This generates two kinds of benefits: First, a *best-realization benefit*: production of an additional unit can be carried out at the low cost c_L^A rather than the high cost c_H^B. Since the event that facility A has drawn the low and facility B the high unit cost has probability $r_L^A(1-q_L^A)$, the impact of the best-realization benefit is given by $r_L^A(1-q_L^A)(c_H^B-c_L^A)$. Idle capacities have a second, more interesting advantage which we call the *slack-reduction benefit*. To understand this second benefit it is important to remember the determinants of the slack accumulated by the facilities. As we have seen earlier the magnitude of corporate resources the low cost facility is able to capture in the form of slack positively depends upon the production quota assigned to this facility if it reports its cost to be high. Where does the slack reduction benefit of an extra unit of capacity in facility A now come from? This benefit arises because the additional unit of capacity in plant A allows the headquarters to reduce the production quota assigned to plant B, if B claims to have high cost. This reduces the incentive of the low-cost realization of plant B to mimic the high cost one and therewith the slack. Since we are talking about a situation in which facility A reports the low and facility B the high potential cost, the quantity of interest is x^B_{HL}, and reducing this quantity by one unit leads to a reduction in B's fat by $q^B_L \Delta^B$ as can be seen from condition (i) of Lemma 1. Since the event that facility B is able to grow fat has probability r^B_L, the impact of the slack-reduction benefit of an additional unit of capacity in plant A is given by $r^B_L q^B_L \Delta^B$.

	Plant A	Plant B
R_1	0	0
R_2	$\max\{0,\xi\} + \max\{0,\delta\}$	0
R_3	$\max\{0,\xi\} + \max\{0,\delta\}$	$r^B_L(c^A_H - c^B_L) + \max\{0,-\xi\} + \max\{0,-\delta\}$
R_4	$p_x - r^A_L(c^A_L + \Delta^A) - r^A_H c^A_H$	$p_x - r^B_L(c^B_L + \Delta^B) - r^B_H c^B_H$

Table 4: Shadow Value of Capacity

The rest of Table 3 is easily explained: In boom periods of the economy, where demand exceeds the amount of capacity available within the boundaries of the firm (Region 4), an additional unit of capacity in any of the facilities allows the headquarters to produce and sell an additional unit of output, a unit that would not have been produced (and sold) otherwise. Thus, the benefit of this unit is simply the market price of output minus actual production 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 production cost for each realization of $c = (c^A, c^B)$ and in which there is excess capacity for sure: If $c_L^A > c_H^B$ but $\xi > 0$ (or $\delta > 0$) then the shadow value of capacity in plant A is strictly positive even if $X \leq k^A + k^B$. The reason for this is again the above mentioned 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 $c_L^A > c_H^B$ and $c = (c_L^A, c_H^B)$ 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, $r_L^B q_L^B \Delta^B$, however, remains positive. So, if in absolute terms, the slack-reduction effect exceeds the best-realization effect ($\xi > 0$) then facility A is allowed to produce, and extra units of capacity in plant A have positive value, despite the high production cost. The argument for the case $c = (c_H^A, c_H^B)$ and $\delta > 0$ is similar.

Note, that in a first-best benchmark in which the facilities' potential costs are observable and verifiable the shadow value of capacity in plant A would be zero if $c_L^A > c_H^B$ and $X \leq k^A + k^B$. 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. In contrast to this, in our second-best world there exists a range of capacity prices for which k^A is strictly positive even if $c_L^A > c_H^B$ (but $\xi > 0$ or $\delta > 0$). This might help to explain the econometric result by Pennings and Sleuwagen (2001) that (multi-plant) multinational enterprises keep their unprofitable facilities longer alive than (single-plant) national corporations, and the evidence (discussed in *Business Week*) indicating that multi-plant firms build new facilities that are technically so inefficient that they are unable to compete most of the time. The explanation suggested by the preceding analysis is that multi-plant firms build, or keep alive, technically less efficient facilities as a credible threat that output will be allocated to those facilities if the more efficient ones try to accumulate too much slack.

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

5 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 2 *Denote the (expected) amount of slack per unit of output produced by $\overline{u}(x)$. That is,*

$$\overline{u}(x) \equiv \frac{1}{x} \sum_{i \in \{A,B\}} r_L^i [q_L^i x_{HL}^i + (1 - q_L^i) x_{HH}^i)] \Delta^i,$$

where x_{HL}^i and x_{HH}^i are as shown in Table 1 for all $X \in [\underline{X}, \overline{X}]$, and where $x = \min\{X, 2k\}$. Then $\overline{u}(x)$ is increasing in x for all x and all k. Furthermore, $\overline{u}(x)$ is strictly increasing in x for all $x \in (k, 2k)$.

Proof 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 $\overline{u}(x)$ and taking into account that $k^A = k^B = k$, say, yields the increasing function

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

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

Proposition 2 tells us that the (expected) per-unit slack is growing in production, x, for given capacities in the facilities. Since production equals demand up to the capacity limit this result 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 actual cost (which equals expected virtual cost) is growing in production, too. We record this result as

Implication 1 *Denote the (expected) actual cost per unit of output produced by $\overline{t}(x)$. That is,*

$$\overline{t}(x) = \frac{1}{x} \sum_{i \in \{A,B\}} \left[\sum_{m \in \{H,L\}} r_m^i [q_m^i c_m^i x_{mL}^i + (1 - q_m^i) c_m^i x_{mH}^i] + r_L^i [q_L^i x_{HL}^i + (1 - q_L^i) x_{HH}^i] \Delta^i \right],$$

where x^i_{mn} *is as shown in Table 1 for all* $(m,n) \in \{H, L\}$ *and all* $X \in [\underline{X}, \overline{X}]$*, and where* $x = \min\{X, 2k\}$*. Then* $\bar{t}(x)$ *is increasing in* x *for all* x *and all* k*. Furthermore,* $\bar{t}(x)$ *is strictly increasing in* x *for all* $x \in (k, 2k)$*.*

Proof Similar to that of Proposition 2 and therefore omitted. □

Implication 1 helps to explain the immense productivity improvements in downturns found in the case studies discussed earlier. The explanation suggested by the present analysis is that internal efficiency has improved in the recession since idle capacities intensify competition, and since intensified competition reduces slack.

6 Concluding Remarks

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

An interesting and important question left open in the present work is the following: 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, from reductions in labor force, and from changes in work practices. Here, a natural question to ask is, to what extent these adjustments can be summarized under the heading 'slack-reduction'? When the cuts regard compensation parts that go 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* hold up of plant stakeholders 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* hold up of employees is unlikely to be the main source for the immense productivity improvements found in the case studies discussed earlier.

References

[1] Aghion, P. and G. Saint-Paul, "Virtues of Bad Times: Interaction between Productivity Growth and Economic Fluctuations", *Macroeconomic Dynamics* **2** (1998), 322-344.

[2] Anton, J. and P. Gertler, "External Markets and Regulation", *Journal of Public Economics* **37** (1988), 243-260.

[3] Anton, J. and D. Yao, "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement", *Rand Journal of Economics* **18** (1987), 57-76.

[4] Auriol, E. and J. J. Laffont, "Regulation by Duopoly", *Journal of Economics and Management Strategy* **1** (1992), 507-533.

[5] Baily, M. and H. Gersbach, "Efficiency in Manufacturing and the Need for Global Competition", *Brooking Papers on Economic Activity* Microeconomics (1995), 307-347.

[6] Basu, S., "Procyclical Productivity: Increasing Returns or Cyclical Utilization?", *Quarterly Journal of Economics* **111** (1996), 719-751.

[7] Borenstein, S. and J. Farrell, "Do Investors Forecast Fat Firms? Evidence from the Gold Mining Industry", *NBER Working Paper W7075, Cambridge, MA* (1999).

[8] Buchholz, D., "General Motors: A Case Study", *Mimeo, Department of Political Science, Duke University, Durham, North Carolina* (1999).

[9] Crémer, J. and R. McLean, "Optimal Selling Strategies under Uncertainty for a Discriminant Monopolist when Demands are Interdependent", *Econometrica* **53** (1985), 345-361.

[10] Dana, J. and K. Spier, "Designing a Private Industry: Government Auctions and Endogenous Market Structure", *Journal of Public Economics* **53** (1994), 127-147.

[11] Davis, S. and J. Haliwanger, "Job Creation, Job Destruction, and Job Reallocation over the Business Cycle", *NBER Macroeconomics Annual, Cambridge MA* (1990).

[12] Demski, J. and D. Sappington, "Optimal Incentives with Multiple Agents", *Journal of Economic Theory* **33** (1984), 152-171.

[13] Demski, J., D. Sappington and P. Spiller, "Managing Supplier Switching", *Rand Journal of Economics* **18** (1987), 77-97.

[14] Hall, R., "Recessions as Re-organizations", *NBER Macroeconomic Annual, Cambridge MA* (1991).

[15] Kerschbamer, R. and Y. Tournas, "Excess Capacity as an Incentive Device", *Discussion Paper 1625, CEPR, London, UK* (1997).

[16] Laffont, J. J. and D. Martimort, "Mechanism Design with Collusion and Correlation", *Mimeo, Universite des Sciences Sociales, Toulouse, France* (1999).

[17] Maskin, E., and J. Riley, "Auction Theory with Private Values", *American Economic Review* **75** (1985), 150-155.

[18] Maskin, E., and J. Riley, "Asymmetric Auctions", *Review of Economic Studies* (1999), forthcoming.

[19] Mitchell, W., "Business Cycles and their Causes", *University of California Press, Berkeley* (1941).

[20] Pennings, E. and L. Sleuwagen, "Exit, Downscaling or International Relocation of Production", *Mimeo, Universitat Pompeu Fabra, Barcelona, Spain* (2000).

[21] Prescott, E., "Needed: A Theory of Total Factor Productivity", *International Economic Review* **39** (1998), 525-551.

[22] Riordan, M., "Contracting with Qualified Suppliers", *International Economic Review* **37** (1996), 115-128.

[23] Riordan, M. and D. Sappington, "Second Sourcing", *Rand Journal of Economics* **20** (1989), 41-58.

[24] Rotemberg, J. and M. Woodfood, "The Cyclical Behavior of Prices and Costs", in J. Taylor and M. Woodfood (eds.), *Handbook of Macroeconomics*, Vol. 1B, North-Holland, Amsterdam (1999).

[25] Rothkopf, M., R. Harstad and Y. Fu, "Is Subsidizing Inefficient Bidders Actually Costly", *Mimeo, Department of Economics, Rutgers University* (1997).

[26] Sanchez, J. and J. Schmitz, "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, Minnesota* (2000).

[27] Schmidt, K., "Managerial Incentives and Product Market Competition", *Review of Economic Studies* **64** (1997), 191-213.