DISCUSSION PAPER SERIES

No. 2557

IN-HOUSE COMPETITION, ORGANIZATIONAL SLACK AND THE BUSINESS CYCLE

Rudolf Kerschbamer and Yanni Tournas

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

ISSN 0265-8003

IN-HOUSE COMPETITION, ORGANIZATIONAL SLACK AND THE BUSINESS CYCLE

Rudolf Kerschbamer, Universität Wien and CEPR
Yanni Tournas, Rutgers University

Discussion Paper No. 2557
September 2000

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 2557

September 2000

ABSTRACT

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

This Paper analyses the impact of variations of product demand on the amount of internal slack in multi-plant firms in a model in which facilities can produce output at a privately known cost up to a previously-determined capacity level. In such a model, the amount of slack in the firm is shown to be pro-cyclical. Indeed, as capacity constraints become tighter in booms, slack increases in booms, because the power of in-house competition is reduced, while the opposite is true in downturns. Also, in downturns the firm may use high-cost facilities even when low-cost plants are not running at capacity.

JEL Classification: D20, D82, F23 and L22
Keywords: capacity, competition, demand fluctuations and slack

Rudolf Kerschbamer
Institut für Wirtschaftswissenschaften
Universität Wien
Hohenstaufengasse 9
A-1010 Vienna
AUSTRIA
Tel: (43 1) 42 77 374
Fax: (43 1) 42 77 9374
Email: Rudolf.kerschbamer@univie.ac.at

Yanni Tournas
Department of Economics
Rutgers University
360 Dr Martin Luther King Jr. Blvd.
Newark NJ 07102
USA
Tel: (1 201) 648 1320
Fax: (1 201) 648 5819
Email: itournas@andromeda.rutgers.edu

* We thank David Besanko, Jim Dana, Raymond Deneckere and Morton Kamien for helpful suggestions and discussions. Of course, the usual disclaimer applies.

Submitted 27 July 2000

NON-TECHNICAL SUMMARY

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 Reifen AG* to the Czech plant *Barum*. Afraid of losing the production rights for two million tyres per year (the former quota was four million) the managing director of *Semperit Austria* promised cost savings of about 700 million ATS within two years. Only a few months later the headquarters of the US-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.

What can we learn from these examples? A first – fairly trivial – insight is that the amount of capacity available within the boundaries of firms sometimes exceeds the level required to produce the output demanded. Whenever this happens to a multi-plant firm the headquarters has to decide how to allocate the total quantity requirement among the different production facilities. A second – more important – observation is that individual facilities are, in general, keen to maintain or expand their activity levels. The headquarters of multi-plant firms seem to know this and they play their facilities off one another in an attempt to induce them to announce and realize substantial cost savings.

The popular press discusses the pressure multi-plant firms put on individual facilities during downturns in prominent feature articles. This topic has not been investigated in the academic literature, however.

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. To do so, we investigate a model in which the demand for the good produced by a multi-plant firm is stochastic; in which the facilities or plants need capacities in order to produce; and in which asymmetric information between the headquarters on the one hand, and the individual facilities on the other, allows facilities to receive rents under any optimal contract. These rents are then dissipated within the facilities in the form of slacking, perquisites, empire building and other forms of at-the-expense-of-the-firm behaviour. In other words, these rents cause internal inefficiencies, or slack.

We show that in this model the per-unit amount of slack is pro-cyclical. Indeed, as capacity constraints are relaxed in economic downturns, slack decreases in downturns, because idle capacities foster in-house competition among plants for higher production quotas. Exactly the opposite is true for boom periods of the economy where demand exceeds the amount of capacity available within the boundaries of the firm. In those periods, slack increases because tight capacity constraints reduce the power of in-house competition. Thus, during boom periods of the economy, firm profits tend to be high since demand is high and capacity is fully utilized. In contrast, during downturns, the firm is able to improve its profitability by concentrating on the cost side, that is, by reducing organizational slack.

Also, during downturns, production is not necessarily assigned to the cheapest plant. Indeed, a plant may be allowed to produce even if it is known to always have the highest production cost and even if demand is so low that the entire quantity could be produced without employing this facility. An intuitive explanation for this result is that the systematic exclusion of a given plant from the production assignment process impedes in-house competition and that this impediment to competition increases the amount of slack in the remaining plants.

Our first two results are for given capacities in the facilities. Next we determine the optimal capacities. We compare them with the capacities in a benchmark model without asymmetric information, and therewith without X-inefficiency, or slack. We show that there exists a unique critical level of capacity cost for which the optimal capacity choice in our model with X-inefficiency coincides exactly with that of the first best benchmark. If capacity is relatively cheap, i.e., if its price is lower than this critical level, the headquarters over-invests in capacity and *vice versa* for prices that are higher. An explanation for this result is easily provided: Due to the slack at the plant level, the cost of output provision is strictly higher in our model than in the benchmark. This implies that fully-utilized capacity carries more value in the benchmark than in the setting considered here. Exactly the opposite is true for idle capacity, which fosters in-house competition and reduces slack. Now consider capacity prices. If the price of capacity is too high, risking being left with idle capacity is too costly; so the argument for fully utilized capacity applies, leading to under-investment in capacity. By contrast, if capacity is relatively cheap, then ending up with excess capacity is profitable in the benchmark, but even more profitable in our model; thus, over-investment in capacity results.

Although our analysis narrowly focuses on a multi-plant firm context where a central authority has the power to allocate production quotas among multiple facilities, the intuition behind our central result about the amount of slack in the firm over the business cycle is not confined to this framework. We argue that basically the same result can be obtained in a decentralized model in which several capacity-constrained single-plant managerial firms sell a homogeneous product under Bertrand conditions, that is, competing in prices.

Our analysis sheds new light on several interesting empirical observations. For instance, the observation that the extent of cost reductions realized in many industries during recessions can hardly be explained by savings in technological production costs. The explanation suggested by the present analysis is that not only technological production cost but also internal slack is reduced as capacity constraints are relaxed in recessions, because the power of competition is increased. Or, the observation that multi-plant firms do not always shut down their existing high-cost facilities after having installed sufficient capacities in low-cost countries. The explanation suggested by the present Paper is that the firms use their old facilities as a device to reduce internal slack in the new plants.

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 Reifen AG" to the Czech plant "Barum".[1] 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 700 million ATS within two years. Only a few months later the headquarters of the US-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]

What can we learn from these examples? A first – fairly trivial – insight is that the amount of capacity available within the boundaries of firms sometimes exceeds the level required to produce the output demanded. Whenever this happens to a multi-plant firm the headquarters has to decide how to allocate the total quantity requirement among the different production facilities. A second – more important – observation is that individual facilities are, in general, keen to maintain or expand their activity levels. The headquarters

[1]With 13 production facilities in eleven European countries and brand names as "Continental", "Uniroyal", "Semperit", "Gislaved", "Viking" and "Barum" the Continental group holds a share of 8% in the world tire market. This market share makes it the fourth largest tire producer worldwide behind Michelin (20%), Bridgestone (18%) and Goodyear (17%).

[2]The details of the Continental-Semperit example originate in articles in the Austrian printed media, including the article "Reifenwechsel als Druckmittel" on July 6^{th}, 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*".

of multiplant firms seem to know this and they play their facilities off one another in an attempt to induce them to announce and realize substantial cost savings.

The popular press discusses the pressure multiplant firms put on individual facilities during downturns in prominent feature articles. This topic has not been investigated in the academic literature, however. 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. To do so, we investigate a model in which the demand for the good produced by a multi-plant firm is stochastic, in which the facilities or plants need capacities in order to produce, and in which asymmetric information between the headquarters on the one hand and the individual facilities on the other, allows facilities to receive rents under any optimal contract. These rents are then dissipated within the facilities in the form of slacking, perquisites, empire building, and other forms of at-the-expense-of-the-firm behavior. In other words, these rents cause internal inefficiencies, or slack.

Our first main result shows that the amount of slack per unit of output produced behaves pro-cyclicaly in this model. Indeed, as capacity constraints are relaxed in economic downturns, slack decreases in downturns, because idle capacities foster in-house competition among plants for higher production quotas. Exactly the opposite is true for boom periods of the economy where demand exceeds the amount of capacity available within the boundaries of the firm. In those periods slack increases because tight capacity constraints reduce the power of in-house competition. Thus, during boom periods of the economy firm-profits tend to be high since demand is high and capacity is fully utilized. By contrast, during downturns, where total demand falls short of total capacity, the firm is able to improve its profitability by concentrating on the cost side, that is, by reducing organizational slack.

Next 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. An intuitive explanation for this result is that the systematic exclusion of a given plant from the production assignment process impedes in-house competition, and that this impedement to competition increases the amount of slack in the remaining plants.

Our first two results are for given capacities in the facilities. Next we determine the optimal capacities. We compare them with the capacities in a benchmark model without asymmetric information, and therewith without X-inefficiency, or slack. We show that there exists a unique critical level of capacity cost for which the optimal capacity choice in our model with X-inefficiency coincides exactly with that of the first best benchmark. If capacity is relatively cheap, i.e., if its price is lower than this critical level, the headquarters over-invests in capacity and vice versa for prices that are higher. An explanation for this result is easily provided: Due to the slack at the plant level, the cost of output provision is strictly higher in our model than in the benchmark. This implies that fully utilized capacity carries more value in the benchmark than in the setting considered here. Exactly the opposite is true for idle capacity which fosters in-house competition and reduces slack. Now consider capacity prices. If the price of capacity is too high, risking being left with excess capacity is too costly; so the argument for fully utilized capacity applies, leading to under-investment in capacity. By contrast, if capacity is relatively cheap, then ending up with excess capacity is profitable in the benchmark, but even more profitable in our model; thus, over-investment in capacity results.[3]

Next we show that the range of capacity prices for which the headquarters over-invests

[3]Here and throughout the rest of the paper we use the term "over-investment" (or "under-investment") in capacity to describe an ex-ante (i.e., before demand has been realized) situation in which the solution to the headquarters' maximization problem in our model yields strictly more (or less) capacity than would be optimal in the first-best benchmark. By contrast, the term "excess capacity" is used to describe an ex-post situation, in which the amount of capacity available within the boundaries of the firm exceeds the level which is necessary to produce the output demanded.

in capacity is decreasing in uncertainty in demand. Loosely speaking, if demand becomes more variable then over-investment in capacity becomes less likely. This result is driven by the fact that second best capacity has a comparative advantage for intermediate demand realization whereas first best capacity has its strength when demand is extremely high. Since extreme realizations become more likely when volatility of demand increases, the result follows.

The present paper relates to several strands of previous work: First and most importantly, the present paper is related to a line of research studying the impact of market competition on internal efficiency of firms. In this literature it has been remarkably hard to generate unambiguous results about competition reducing slack. Hart (1983) and Scharfstein (1988), for instance, show in a hidden information model in which a common shock is transmitted via the market price, that the (informational) effect of an increase in competition by entrepreneurial (profit-maximizing) firms on the internal efficiency of managerial firms crucially depends on the specification of managers' preferences. Hermalin (1992) confirms the ambiguous informational effect of competition in a hidden action model. Horn et al. (1994) study the strategic value of incentive contracts under different market conditions. In their work an increase in the intensity of competition leads to more X-inefficiency. A negative relation between intensity of competition and degree of internal efficiency arises also in Martin's (1993) Cournot principal-agent model, where the principal's marginal benefit of inducing the agent to minimize cost becomes smaller when competition increases. More recently Schmidt (1997) analyses a model in which the basic impact of an increase in competition is that it reduces the profits of firms. In this paper, the overall effect of an increase in competition on the extent of operating slack is again ambiguous: 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.

The main difference between this strand of literature and the present paper is not that this literature analyses the effect of product market competition on internal efficiency while we study that of in-house competition.[4] A key difference consists rather in the way in which an increase in the intensity of competition is modeled. Hart and Scharfstein model increased competition as a higher fraction of entrepreneurial firms. In the two papers that followed (Hermalin 1992, and Horn et al. 1994) the intensity of competition increases through a change in the nature of product market rivalry, from Cournot- to Bertrand-competition. In the paper by Martin (1993) the degree of competition increases as the number of Cournot firms in the market becomes larger. And Schmidt (1997) equates increased competition with lower profits. By contrast, in the present paper the intensity of rivalry among facilities increases when capacity constraints are relaxed in downturns. Since the competitive pressure proceeding from idle capacities has not been analyzed in this line of research, we see our paper primarily as a complement to this literature, that is, to the earlier literature on the effects of competition on the internal efficiency of firms.

The virtues of bad times have also been recognized in the new growth theory. Aghion and Saint-Paul (1998), for instance, 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. 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[5] productivity increases in downturns because the opportunity cost

[4]We argue in Section 4 below that basically the same pro-cyclical relation between industry demand and amount of slack in the firm can be obtained in an decentralized hidden information model where several capacity-constrained single-plant managerial firms supply a homogeneous good under Bertrand conditions, i.e., competing in price.

[5]Besides the (first) Aghion and Saint-Paul model there are, among others, contributions by Davis and

of investing capital and labor resources 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 competition, and because intensified competition reduces slack.

From a modeling perspective the present work is also related to those papers in the procurement and regulation literature that show that a carefully designed allocation of production to plants can help to reduce information cost. For instance, Anton and Gertler (1988) study optimal regulatory policy towards a firm that is a monopolist in an "internal" market and can participate in an "external" market, the latter being outside the regulator's domain of concern. They show that the policy of excluding the regulated firm from the profit opportunities created by the external market if it reports its cost to be high can help to limit agency costs.[6] Related observations have been made by other authors.[7] A major difference to the present paper is, that in this literature the quantity to be produced is exogenously given and capacity choice is no issue. By contrast, demand is random in the present work and the competition-fostering effect of idle capacities is one of our main topics.

The issue of capacity has previously been studied by Riordan (1996). In a procurement model in which the quantity to be produced is exogenously fixed, and in which the orderer first decides about the number of potential suppliers and their capacities and then about the division of production among them, he shows that the capacity of each of the potential suppliers exactly equals demand and that asymmetric information biases the market

Haltiwanger (1990), Gali and Hammour (1992), and Hall (1991).

[6] In a later paper Anton and Gertler (1994) examine regulation in a duopoly model of spatial competition. They show that the regulator can reduce information costs by increasing a relatively more efficient firm's market through a reassignment of consumers at the competitive fringe.

[7] Cf, for instance, Auriol and Laffont (1992), Dana and Spier (1994) and McGuire and Riordan (1995) in the duopoly-versus-monopoly literature, and Laffont and Tirole (1987), McAfee and McMillan (1986) and Riordan and Sappington (1987) in the auctioning literature.

structure in favour of more suppliers. The main difference between Riordan's model and ours is (a) that demand is random in the present paper, and (b) that the plants' cost distributions are asymmetric. Basically, feature (a) drives our first main result (on the relationship between the business cycle and internal slack), and feature (b) the second (that the firm uses high-cost facilities even when low-cost plants are not running at capacity).

Our result about the desirability of maintaining an inefficient plant has parallels in the second-sourcing literature (see, e.g., Anton and Yao 1987, Demski et al. 1987, and Riordan and Sappington 1989) where it has been shown that the occational 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 favour 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).

Before proceeding let us give a short overview over the rest of the paper. The next section (Section 2) introduces the model and offers a formal statement of the headquarters' maximization problem. Section 3 characterizes optimal contracts and capacities and derives the result that high-cost plants are allowed to produce even when low-cost facilities are not running at capacity. The relationship between the business cycle and operational slack is analyzed in Section 4. Section 5 studies the effects of uncertainty in demand on capacities and shows that, in comparison to a benchmark without operational slack, the headquarters either over- or underinvests in capacity. Section 6 concludes. All the proofs are in the Appendix.

2 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. X is a random variable that is uniformly distributed on some interval $[\underline{X}, \overline{X}]$, where $0 \leq \underline{X} < \overline{X} < \infty$.[8] 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.[9] 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 c^i. Each c^i a priori belongs to $C^i = \{c^i_L, c^i_H\}$, where $c^i_H - c^i_L = \Delta^i > 0$.[10] 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: $1 > q^i_L \geq q^i_H > 0 \qquad \forall i \in \{A, B\}$.

The objective of each facility is to maximize the expected gain from dealing with the headquarters. This gain, or surplus, 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 surplus appropriated by a facility is dissipated within this facility in the form of slacking, perquisites, and other forms of at-the-expense-of-the-firm behavior. In other words, this surplus causes slack, or X-inefficiency. We also assume that the facilities are protected by limited liability so that their surplus is at least 0 ex-post.[11]

[8] The uniform distribution facilitates the proofs of Propositions 4 and 5. For the other results we use only the fact that X has full support on some (nondegenerate) interval $[\underline{X}, \overline{X}]$. None of the results relies on the uniform distribution.

[9] Capacity investments are assumed to be nonverifiable and very relation-specific (capacity has no alternative-use value) here. Also the exact characteristics of the good to be delivered ex post are assumed to be unknown ex ante. In such a context in-house ownership of both production facilities is an efficient way of eliminating the potential for hold-up problems ex post, and thus, to induce an efficient amount of investment ex ante (See e.g. Williamson (1975), or Grossman and Hart (1986)).

[10] 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.

[11] There is also a technical reason for introducing ex post individual rationality constraints: From

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 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 unit costs are drawn from their respective distributions. Demand becomes publicly observable and verifiable. Unit cost c^i, however, is privately observed by facility i. After having learned their c^is the facilities simultaneously and confidentially make cost reports to the headquarters.[12] The headquarters collects the reports and sends production recommendations (according to the contract) back to the facilities. Now the facilities decide how much to 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 $X \in [\underline{X}, \overline{X}]$.[13] Here, $x^i(c^i, c^j, X)$ is the output level required of facility i if the cost reports

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 aritificial result.

[12]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 formal framework changes dramatically. In the limit – when the facilities behave as a single entity – we get a single-agent model with two-dimensional uncertainty.

[13]This is not totally correct since the headquarters might wish to offer contracts in which output levels and transfers are stochastic functions of the agents' reports. However, it can be shown that that pure randomization is of no value in the setting considered here.

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 obtains $x^i(c^i, c^j, X)$. To keep the notation symmetric, we adopt the convention that the first cost-report argument in x^i and t^i is the report from plant i while the second is the report from 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 contraints. 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 low cost facility from pretending to have high cost. A trivial solution to this problem is to shut down the facility if it claims to have high cost ($x^i_{Hm} = 0$ for $m \in \{L, H\}$). To evade this solution we impose the following assumption on our problem:[14]

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

[14]In a model with more realistic cost- (increasing marginal cost) and/or demand- (downward sloping demand) conditions our results can be obtained without this assumption.

If a facility declares bankrupt it gets a payoff of 0. 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 (gross revenue minus transfers to the facilities) under incentive compatibility, individual rationality and capacity constraints. Formally, the headquarters' contracting problem at Stage 1 is:

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

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 $X \in [\underline{X}, \overline{X}]$.

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

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. More precisely, we assume that $c_m^A \geq c_m^B$ for $m \in \{L, H\}$. In this case optimal capacities are characterized by $k^A \leq k^B$, as we will see below. We therefore take this into consideration in dealing with the contracting problem.[15]

Our first result (Lemma 1) characterizes the solution to this problem. In this result reference is made to the variables ξ and δ. These variables are defined by $\xi \equiv r_L^A(1 - q_L^A)(c_H^B - c_L^A) + r_L^B q_L^B \Delta^B$ and $\delta \equiv r_H^A(1 - q_H^A)(c_H^B - c_H^A) + r_L^B(1 - q_L^B)\Delta^B - r_L^A(1 - q_L^A)\Delta^A$. Also, in our first result reference is made to 4 different regions in the demand space. These regions (denoted by R_1 to R_4) are defined in Figure 1.

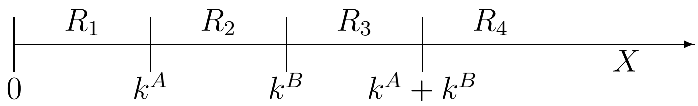

Figure 1: Capacities and Demand

Finally, in Lemma 1 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 $m \in \{L, H\}$.

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

[15] 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 $k^A = k^B$.

(*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* $x^B_{mn} = \min\{X - x^A_{nm}, k^B\}$.

Lemma 1 indicates that a facility gets only a compensation for its production cost if it has observed the high cost-parameter c_H, while it earns a strictly positive surplus, or rent, in the more favourable environment c_L. The magnitude of the rent in the more favourable environment positively depends upon the output quota assigned to the facility if it claims to have high cost. This is easily understood. The low-cost plant gets a rent, since without it, it would always have an incentive of mimicking the high-cost one. Reducing the output quota assigned to the high-cost plant reduces the gain of the low-cost plant for mimicking the high-cost one and therewith its rent. This property of optimal contracts is important for our main results and we will return to it later.

	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

Let us turn to the allocation of production quotas. For the symmetric case this allocation is depicted in Table 1 and associated Figure 1. Figure 1 defines 4 different regions for demand. 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. That is, 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. 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 (depicted in Table 2) behaves similarly, except that virtual rather than real costs are compared. The virtual cost of a given type differs from its real cost in that the informational rents paid to those types that jeopardize the given type are taken into account. Type c_m jeopardizes type c_n if the headquarters has difficulty preventing c_m from mimicking c_n. In the present context the high cost type is jeopardized by the low cost one but not vice versa. So, the virtual cost of the low cost type is just its real cost while the virtual cost of the high cost type is its real cost plus a term that measures the additional rent that must be paid to the low cost type if the quantity produced by the high cost one is increased by one unit. The rest is trivial: (i) If both plants report low costs all production up to capacity is allocated to the more efficient plant B. (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 cost asymmetry so much favours the plant with the high cost-report that it compensates for the increase in the informational rent 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 production 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 $\Delta^B r_L^B q_L^B / r_L^A (1 - q_L^A) > c_L^A - c_H^B$ 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 rent he would get 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 $X \leq k^A$. We record this result as

Proposition 1 *Plant A may be allowed to produce even if it is known to have always higher production costs ($c_L^A > c_H^B$) and even if demand is so low that there remains excess capacity in plant B ($k^B < X$).*

On an intuitive level an explanation for this result is that if the more efficient plant B knows that it is allowed to produce no matter what its cost-report is, inducing truthful revelation is quite difficult, i.e., a high rent is required to accomplish this. 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 size of informational rents.

Note that this result – combined with the observation that capacity-investments are in most cases to some degree sunk – may help to explain the observation that multinational enterprises that build new plants in low-cost countries do not always shut down their old facilities in high-cost countries even if the new facilities are large enough to produce the quantity demanded under most favorable market conditions. The explanation suggested by the preceding analysis is, that the headquarters use their old facilities as a credible threat that output will be allocated back to those facilities if the reported production costs in the new plants are too high.

The Capacity Choice Problem

The next step is to determine the optimal capacity levels for the facilities. Optimal capacities are found by setting the expected shadow value of a marginal unit of capacity equal to the capacity price. 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.*

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 cost realization 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 unit 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 $r_L^A(1-q_L^A)(c_H^B-c_L^A)$.[16] 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 rents received by the facilities. As we have seen earlier the magnitude of the rent earned by the low cost facility 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 mimicke the high cost one and therewith its rent. Since we are talking about a situation in which facility A reports the low and facility B the high unit cost, the quantity of interest is x_{HL}^B, and reducing this quantity by one unit leads to a reduction in B's rent by $q_L^B\Delta^B$ as can be seen from condition (i) of Lemma 1. Since the event that facility B gets a rent has probability r_L^B, the impact of the slack-reduction benefit of an additional unit of capacity in plant A is

[16] An effect that is closely related to our *best realization benefit* has earlier been identified by Auriol and Laffont (1992) in a paper on the optimal structure of an industry. The authors show that one of the merits of a duopoly (in comparison to a monopoly) is the increased sample-size that gives a higher probability of drawing a low marginal cost for the industry. They refer to this as the *sampling effect.* A difference to our best realization benefit is that the sampling effect requires an increase in the number of draws (in our context, an increase in the number of plants) while our best realization benefit arises with a given number of draws because a good draw (= low cost realization) can better be exploited.

given by $r_L^B q_L^B \Delta^B$.

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

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 production and information costs.

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 informational rent 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' cost-parameters 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 price of capacity. 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 $\delta > 0$).

Also note that Corollary 1 implies that optimal capacities are characterized by $k^A = k^B$ in the symmetric, and by $k^A < k^B$ in the asymmetric case. To see this for the symmetric case suppose to the contrary that (without loss of generality since the plants are otherweise identical) $k^A < k^B$. Then a transfer of one unit of capacity from fracility B to facility A results in an increase in the headquarters' expected payoff by $\xi > 0$ if X falls in the interval $(k^A, k^B]$ and in no change in this payoff if X falls outside this range.[17] The argument for the asymmetric case is similar: Suppose $k^B \leq k^A$. Then a reallocation of one unit of capacity from plant A to plant B results even in the worst case (where $\xi > 0$ and $\delta > 0$) in an increase in the headquarters' expected payoff if $X > k^B$ and in no change in this payoff otherwise.[18] Thus, the optimality of $k^A \neq k^B$ is contradicted in the symmetric, and the optimality of $k^B \leq k^A$ is contradicted in the asymmetric case.

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

[17]Notice that in the symmetric case $\xi = r_L^A(1 - q_L^A)(c_H^B - c_L^A) + r_L^B q_L^B \Delta^B = r_L^B(1 - q_L^B)(c_H^B - c_L^B) + r_L^B q_L^B \Delta^B = r_L^B(c_H^A - c_L^B)$.

[18]To see this notice that for $k^B \leq k^A$ Table 4 looks exactly as the actual Table 4 except that Region 2 is defined by $(k^B, k^A]$ (and Region 3 by $(k^A, k^A + k^B)$) and that the shadow value of a marginal unit of capacity in this region is 0 for plant A and $r_L^B(c_H^A - c_L^B) + \max\{0, -\xi\} + \max\{0, -\delta\}$ for plant B. Also notice that in the worst case ($\xi > 0, \delta > 0$) the shadow value of capacity in Region 3 is $r_L^B(c_H^A - c_L^B)$ for plant B and that $r_L^B(c_H^A - c_L^B) > r_L^B(c_H^B - c_L^B) > r_L^B(c_H^B - c_L^B) - r_H^B(c_H^A - c_H^B) = \xi + \delta$, where the latter is the shadow value of capacity for plant A.

4 Organizational Slack and the Business Cycle

The goal of this section is to analyse the effect of variations in product demand on the amount of internal slack, measured by the size of the rents dissipated within the facilities. As mentioned earlier, earlier work from other authors had difficulty to generate unambiguous results in this dimensions. By contrast, we get a sharp unambiguous result:

Proposition 2 *Denote the (expected) amount of slack per unit of output produced by $\phi(x)$. That is,*

$$\phi(x) \equiv \frac{1}{x} \sum_{i \in \{A,B\}} r_L^i [q_L^i u_{LL}^i + (1 - q_L^i) u_{LH}^i)],$$

where $x = \min\{X, 2k\}$. Then $\phi(x)$ is increasing in x for all x and all k. Furthermore, $\phi(x)$ is strictly increasing in x for all $x \in (k, 2k)$.

Proposition 2 tells us that the per-unit slack is growing in the level of demand, i.e. is pro-cyclical, for given capacities in the facilities. In other words, for given capacity levels, 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. Here notice that this result doesn't depend on any correlation in the cost-parameters of the facilities. The driving force is rather the facilities' greediness for higher production quotas.

Although Proposition 2 has been derived in a multi-plant firm context where a central authority, the headquarters, has the power to allocate production among multiple facilities the intuition behind this result extends well beyond this framework. Basically the same pro-cyclical relationship between industry demand and amount of slack could be obtained, for instance, in a decentralized model where several capacity-constrained single-plant firms, each run by a manager with private information on unit cost, supply a homogeneous good under Bertrand conditions. In such a model a low-cost manager has little incentive to misrepresent his private information during downturns because in downturns high-cost firms are unable to stand the market test. Things are different in boom periods of the economy where demand exceeds the amount of capacity available within the industry. In such periods X-inefficiency increases because tight capacity constraints reduce the power of Bertrand competition.[19]

Proposition 2 sheds new light on the finding that in many industries substantial cost savings are realized during downturns. Empirical evidence supportive of this finding is provided by the literature on productivity growth and technical progress (cf., for instance, Gali and Hammour 1992, or Malley and Muscatelli 1996). The finding is also supported by casual observations. Schmidt (1997), for instance, reports that during the latest recession in the car industry almost all suppliers reduced their cost by 20 to 30% within two years. He argues that there has been no accompanying change in technology allowing for *savings in technological production-cost* in that order. So, the question arises where these savings stem from. The answer suggested by the present analysis is that *internal efficiency has improved* during the last recession since idle capacities intensify competition, and since intensified competition reduces slack.[20]

[19]Since Proposition 2 is for given capacities in the facilities it doesn't matter that capacity decisions are decentralized in a decentralized model. Rather important, however, is the production assignment process. In the symmetric setting the headquarters' production assignment in the present model (as depicted in Table 1) is exactly the same as it would be under Bertrand competition, i.e., when firms compete in prices.

[20]Schmidt (1997) explains the puzzle by the "threat of liquidation effect": The decrease in profits during recessions makes it more likely that firms that do not manage to reduce costs are driven out of the market. This gives a direct incentive to managers to work harder. We view this and our own explanation as complements not as substitutes.

5 Capacity and Uncertainty of Demand

Having analysed how variations in product demand affect the amount of organizational slack for given capacities in the facilities, we now turn to the choice of capacity for each plant. Our first result here compares the optimal capacity in our model with X-inefficiency – which we denote by $\hat{k}$ – with that in a benchmark in which the cost parameters c^A and c^B are observable and verifiable. We denote the optimal capacity level for this benchmark by k^*.

Proposition 3 *There exists a unique critical price of capacity $\tilde{p}_k$ for which $\hat{k} = k^*$. If $p_k > \tilde{p}_k$, then $\hat{k} < k^*$. If $p_k < \tilde{p}_k$, then $\hat{k} > k^*$.*

Proposition 3 tells us that there exists a unique critical price of capacity for which the optimal capacity level in our model with X-inefficiency coincides exactly with that of the first best benchmark. If capacity is relatively cheap, i.e., if its price is lower than this critical level, the headquarters over-invests in capacity (relative to the first best benchmark), while the opposite is true for prices that are higher than the critical level. This result is easily explained: Since idle capacities limit slack at the plant level, the shadow value of an additional unit of idle capacity is strictly higher in our model than in the benchmark without X-inefficiency. Exactly the opposite is true for fully utilized capacity: Due to the absence of slack, the cost of output provision is strictly lower in the benchmark, the shadow value of fully used capacity therefore higher. Now consider capacity cost. If the cost of capacity is high, risking to be left with excess capacity is too expensive. So the argument for fully utilized capacity applies, leading to under-investment in capacity. On the other hand, if capacity is relatively cheap, running the risk of a situation in which capacity is not matched by demand is profitable in the benchmark but even more profitable in our model. Thus, over-investment in capacity results.

The low capacity-cost part of Proposition 3 closely resembles an earlier result by Riordan (1996). In a procurement model in which the *quantity to be produced is exogenously* *fixed*, and in which the orderer first decides about the number of potential suppliers and their capacities and then about the division of production among them, he shows that the capacity of each of the potential suppliers exactly equals demand and that asymmetric information biases the market structure in favour of more suppliers. Together these two facts imply an unambiguous over-investment result. That the current model can also generate under-investment is not surprising given the difference in the modelling assumptions: Riordan derives his results under the supposition that it is always optimal to provide sufficient capacity to cover demand. Since the *total quantity requirement* of the headquarters *is driven by fluctuations in market demand* in the present model, and since the headquarters must decide about capacities before knowing which state of demand is going to realize, such an assumption would be difficult to motivate in our work.

Note that our central result on the relationship between demand and the amount of slack in the firm doesn't depend on capacity costs and therewith also not on whether over- oder under-investment in capacity prevails. What is needed to reduce organizational slack is rather excess capacity. While *over-investment in capacity* refers to an ex ante situation in which second best capacities exceed first best levels, *excess capacity* describes an ex post situation in which capacity exceeds demand.[21] Since demand is random in the present model, excess capacity arises also in the under-investment range of capacity prices.

Up to now we have assumed that demand is uniformly distributed over a given interval $[\underline{X}, \overline{X}]$. A next question of interest is, how our results change if we vary the "extent of uncertainty in demand", i.e., if we modify the bounds of this interval. To keep things simple we consider mean preserving spreads of the form $\underline{X}^{new} = \underline{X} - \mu$ and $\overline{X}^{new} = \overline{X} + \mu$, where $0 \leq \mu \leq \underline{X}$. Obviously, Lemma 1, Propositions 1 and 2, and Corollary 1 remain unaffected by such a change. Proposition 3 holds in a slightly modified form in which $\hat{k}$, k^* and $\tilde{p}_k$ are replaced by $\hat{k}(\mu)$, $k^*(\mu)$ and $\tilde{p}_k(\mu)$. An obvious question to ask is, how

[21] Ex ante here refers to a situation before demand has been realized.

optimal first and second best capacities change if the volatility of demand increases. The answer seems to be straightforward: There is an option value on each unit of capacity for being able to choose the lower cost firm (in virtual cost terms) ex post. That option value should increase with higher volatility. Thus, an increase in the volatility of demand should increase both first and second best capacities. Furthermore, since the option value of capacity is higher in our model with X-inefficiency than in the first best benchmark the range of capacity prices for which over-investment in capacity prevails should increase in uncertainty of demand, i.e., $\tilde{p}_k(\mu)$ should be an increasing function. Things are more complicated, however, as the following results show:

Proposition 4 *The impact of an increase in the volatility of demand on optimal capacities (i.e., the derivative of $\hat{k}(\mu)$ and $k^*(\mu)$ with respect to μ) is ambiguous in sign.*

Proposition 5 *The critical price of capacity $\tilde{p}_k(\mu)$ is decreasing in μ. Furthermore, for each pair $(\underline{X}, \overline{X})$ there exists some $\epsilon \in [0, \min\{0, .75\underline{X} - 0.25\overline{X}\}]$ such that $\tilde{p}_k(\mu)$ is strictly decreasing in μ for all $\mu \in [\epsilon, \underline{X}]$.* [22]

What went wrong with our option value story? Consider first Proposition 4, i.e., the impact of an increase in volatility of demand on optimal capacities. To explain this result, let us return to Table 4 showing the shadow value of capacity for different demand realizations. In the symmetric case considered here, the shadow value of capacity is zero for $X \leq k$, $\xi = r_L(1 - q_L)(c_H - c_L) + r_L q_L \Delta$ for $X \in (k, 2k)$, and $p_k - r_L(c_L + \Delta) - r_H c_H$ for $X \geq 2k$. An optimal level of k is found by setting the expected shadow value of capacity equal to the capacity price, p_k. Suppose for the sake of the argument that the price of capacity is fairly low; more precisely, suppose $p_k \in (0, \xi)$. Further suppose that $\mu = 0$. Then the optimal second best level of capacity for each facility, $\hat{k}$, lies in the interval $(\underline{X}, \overline{X})$, where the shadow value is zero with probability $(\hat{k} - \underline{X})/(\overline{X} - \underline{X})$

[22] If $\overline{X} \geq 3\underline{X}$ then $\tilde{p}_k(\mu)$ is strictly decreasing for all $\mu \in [0, \underline{X}]$.

and ξ with probability $(\overline{X} - \hat{k})/(\overline{X} - \underline{X})$.[23] For $\mu > 0$ these probabilities change to $(\hat{k} - \underline{X} + \mu)/(\overline{X} - \underline{X} + 2\mu)$ and $(\overline{X} - \hat{k} + \mu)/(\overline{X} - \underline{X} + 2\mu)$. So, for the range of capacity prices under consideration, the impact of μ on $\hat{k}$ crucially depends upon which of these two probabilities is increasing in μ. First suppose that p_k is close to ξ implying that $\hat{k}$ is close to $\underline{X}$. Then the first of these two probabilities is increasing, and $\hat{k}$ therefore decreasing in μ. Exactly the opposite is true for p_k close to 0 (implying $\hat{k}$ close to $\overline{X}$). The story for capacity prices exceeding ξ is similar. Again, the impact of an increase in the volatility of demand on $\hat{k}$ crucially depends upon which of the three shadow values of capacity in Table 3 gets more probability weight as μ increases. And this depends upon the initial level of $\hat{k}$ and therewith on the prevailing price of capacity. Since the shadow value of capacity for the first best benchmark is obtained by setting slack, represented by the term Δ in the expressions in Table 3, equal to zero, the same is true for k^*.

Now consider Proposition 5 saying that the range of capacity prices for which we get over-investment in capacity is decreasing in uncertainty in demand. Loosely speaking, if "demand becomes more variable" (μ increases) then "over-investment in capacity becomes less likely". To see the intuition behind this result, recall that the shadow value of capacity in our model with X-inefficiency exceeds that in the first best benchmark if $X \in (k, 2k)$ while the opposite is true for $X > 2k$. For $X \leq k$, on the other hand, capacity carries the same shadow value – namely zero – in both settings. So, for $\hat{k}$ to be equal to k^*, both, demand realizations in $(k, 2k)$ and demand realizations exceeding $2k$, must have strictly positive probability weight. For this to be the case, capacities must lie in the interval $[.5(\underline{X} - \mu), .5(\overline{X} + \mu)]$. Now, fix capacities in this interval and increase μ. Then the probability that X exceeds $2k$ cannot decrease while the probability that X falls in the intermediate range $(k, 2k)$ cannot increase. Thus, the balance shifts to the benchmark's favour. To restore the balance, optimal capacities must increase and $\tilde{p}_k$ therefore decrease in μ. To summarize: Proposition 5 is driven by the fact that first best capacity has

[23] Here we assume that $2\underline{X} > \overline{X}$. The argument for the opposite case is similar.

a comparative advantage for extreme demand realizations ($X \geq 2k$) while second best capacity has its strength when realized demand is at an intermediate level ($X \in (k, 2k)$). Since extreme realizations become more likely when volatility of demand increases, the result follows.

6 Concluding Remarks

This paper has studied the effect of variations of product demand on the amount of organizational slack in multi-plant firms in a model in which plants can produce output at a privately known cost up to a previously determined capacity level. The private information on production costs allow facilities to receive rents which are dissipated within the facilities in form of slacking, perquisites, and other forms of at-the-expense-of-the-firm behavior. We have shown that in such a model, the amount of operating slack in the firm is pro-cyclical. Indeed, as capacity constraints become tighter in booms, slack increases in booms, because tight capacity constraints hamper competition among plants for higher production quotas. The converse argument applies for downturns. In such periods slack decreases because the intensity of rivalry for quota assignments increases when capacity constraints are relaxed. Thus, during boom periods of the economy firm-profits tend to be high since demand is high and capacity is fully utilized. By contrast, during downturns the firm is able to improve its profitability by concentrating on the cost side, that is, by reducing organizational slack.

We have also shown, that during downturns of the economy the firm may use high-cost facilities even when low-cost plants are not running at capacity. The reason for the *ex post* inefficient production assignment to a high-cost plant is, loosely speaking, the desire to improve internal efficiency in low-cost plants. Regarding capacity we have seen that, in comparison to a benchmark model without X-inefficiency, the firm has an incentive to overinvest in capacity when capacity cost is low, while the opposite is true for high capacity prices. This result is due to the fact that fully utilized capacity has a higher shadow value in the benchmark where average production cost is lower, while idle capacity is more advantageous in our model where it reduces slack at the plant level.

Although our analysis has narrowly focused on a multi-plant firm context where a central authority has the power to allocate production quotas among multiple facilities we don't think that the intuition behind our central result about the amount of slack in the firm over the business cycle is confined to this framework. Indeed, we have argued that the same result can be obtained in a decentralized model in which several capacity-constrained single-plant managerial firms sell a homogeneous product under Bertrand conditions, that is, competing in prices.

Our analysis sheds new light on several interesting empirical observations. For instance, the observation that the extent of cost reductions realized in many industries during recessions can hardly be explained by savings in technological production costs. The explanation suggested by the present analysis is that not only technological production cost but also internal slack is reduced as capacity constraints are relaxed in recessions, because the power of competition is increased.[24] Or, the observation that multi-plant firms do not always shut down their existing high-cost facilities after having installed sufficient capacities in low-cost countries. The explanation suggested by the present paper is that the firms use their old facilities as a device to reduce internal slack in the new plants.

[24] An alternative explanation for this observation is provided by the new growth literature (cf., for instance, Davis and Haltiwanger 1990, Hall 1992, Gali and Hammour 1992, and Aghion and Saint-Paul 1998). In this literature productivity increases in recessions because the opportunity cost of productivity-improving activities goes down in recessions by more than their return. We view this and our own explanation as complements, not as substitutes.

APPENDIX

Proof of Lemma 1

The proof begins by analysing a relaxed program in which the downward incentive constraints IC_H^A and IC_H^B are not included. Later we show that the solution to this relaxed program admits a continuum of transfer prices which satisfy these constraints. First notice that the upward incentive constraints IC_L^A and IC_L^B cannot be relaxed by *increasing* production. Also, there is no other gain from producing output that cannot be sold. Our original relaxed program is therefore equivalent to

$$\text{Max}_{(u,x)}\text{NR}' =$$

$$= \sum_{i \in \{A,B\}} \sum_{m \in \{H,L\}} r_m^i [q_m^i (x_{mL}^i (p_x - c_m^i) - u_{mL}^i) + (1 - q_m^i)(x_{mH}^i (p_x - c_m^i) - u_{mH})]$$

subject to

$$(D_{mn}) \qquad X \geq x_{mn}^A + x_{nm}^B,$$

(IC_H^i), (IR_{mn}^i) and (K_{mn}^i) hold for all $i \in \{A, B\}$, $(m, n) \in \{H, L\}^2$ and $X \in [\underline{X}, \overline{X}]$. Letting $\beta_m^i, \alpha_{mn}^i, \gamma_{mn}^i$ and λ_{mn} be the Lagrange multipliers associated with (IC_m^i), (IR_{mn}^i), (K_{mn}^i) and (D_{mn}) the first order conditions for a solution to this problem include (we focus on one plant and delete the i superscript):

$$(1) \qquad -r_L q_L + \beta_L q_L + \alpha_{LL} = 0$$

$$(2) \qquad -r_L(1 - q_L) + \beta_L(1 - q_L) + \alpha_{LH} = 0$$

$$(3) \qquad -r_H q_H - \beta_L q_L + \alpha_{HL} = 0$$

$$(4) \qquad -r_H(1 - q_H) - \beta_L(1 - q_L) + \alpha_{HH} = 0$$

Standard arguments reveal that β_L, α_{HL} and α_{HH} are strictly positive while $\alpha_{LL} = \alpha_{LH} = 0$. From adding up (1) and (3) we get $\alpha_{HL} = r_L q_L + r_H q_H$. Also, from (2) and (4), $\alpha_{HH} = r_L(1 - q_L) + r_H(1 - q_H)$. Thus, $\beta_L = r_L$ by (1).

In addition, condition (5i) – (8i) have to be satisfied for $i = A, B$. In these conditions $\lambda_+ = \lambda_{LH}$ and $\lambda_- = \lambda_{HL}$ if $i = A$ and $\lambda_+ = \lambda_{HL}$ and $\lambda_- = \lambda_{LH}$ if $i = B$.

(5i) $\quad r^i_L q^i_L(p_x - c^i_L) - \gamma^i_{LL} - \lambda_{LL} \leq 0, \quad x^i_{LL} \geq 0$ and c.s.;

(6i) $\quad r^i_L(1 - q^i_L)(p_x - c^i_L) - \gamma^i_{LH} - \lambda_+ \leq 0, \quad x^i_{LH} \geq 0$ and c.s.;

(7i) $\quad r^i_H q^i_H(p_x - c^i_H) - r^i_L q^i_L \Delta^i - \gamma^i_{HL} - \lambda_- \leq 0, \quad x^i_{HL} \geq 0$ and c.s.;

(8i) $\quad r^i_H(1 - q^i_H)(p_x - c^i_H) - r^i_L(1 - q^i_L)\Delta^i - \gamma^i_{HH} - \lambda_{HH} \leq 0, \quad x^i_{HH} \geq 0$ and c.s.

Four different cases are to be considered with respect to demand realization:

Case 1: $k^A + k^B \leq X$**.** In this case conditions (K^i_{LL}) - (K^i_{HH}) are more demanding than conditions (D_{LL}) - (D_{HH}) so that $\lambda_{LL} = \lambda_{LH} = \lambda_{HL} = \lambda_{HH} = 0$. Under Assumption 2 the first inequality in (5i) - (8i) holds as an equation (note that, by Assumption 1, $q^i_L/q^i_H \geq 1 \geq (1 - q^i_L)/(1 - q^i_H)$) and $x^i_{HL} = x^i_{HH} = x^i_{LL} = x^i_{LH} = k^i$, for $i = A, B$.

Case 2: $k^B < X < k^A + k^B$**.** In this case Assumption 2 guarantees that $\lambda_{mn} > 0$ for all $(m, n) \in \{L, H\}^2$: If $\lambda_{LL} = 0$, then $\gamma^A_{LL} = r^A_L q^A_L(p_x - c^A_L)$ [see Case 1] and $x^A_{LL} = k^A$. Since the same argument holds for plant B condition (D_{LL}) is violated, contrary to the hypothesis. Similarly for λ_{LH}, λ_{HL} and λ_{HH}. Next observe, that $\lambda_{LL} > 0$ (together with D_{LL} and K^B_{LL}) implies that $x^A_{LL} > 0$. By a similar argument, $x^i_{mn} > 0$ for all $i \in \{A, B\}$ and $m, n \in \{L, H\}$. This, in turn, implies that the first inequality in (5i) - (8i) holds as an equation for $i \in \{A, B\}$.[25] Next notice that $\gamma^A_{HL} = 0$. To see this suppose to the contrary that $\gamma^A_{HL} > 0$. Then $x^A_{HL} = k^A$. Furthermore, since condition (7A) must hold with equality, condition (6B) can be satisfied only if $\gamma^B_{LH} > 0$ (by the definition of the conditional probabilities $r^A_H q^A_H = r^B_L(1 - q^B_L)$; by assumption $c^A_H \geq c^B_H > c^B_L$) so that $x^B_{LH} = k^B$. But $x^A_{HL} = k^A$ and $x^B_{LH} = k^B$ together violate (D_{LH}). The fact that $\gamma^A_{HL} = 0$, in turn, implies that $\lambda_{HL} = r^A_H q^A_H(p_x - c^A_H) - r^A_L q^A_L \Delta^A$ and $\gamma^B_{LH} = r^B_L(1 - q^B_L)(c^A_H - c^B_L) + r^A_L q^A_L \Delta^A$ so that $x^B_{LH} = k^B$ and $x^A_{HL} = X - k^B$. Next notice that $\gamma^A_{LL} = 0$. To see this suppose to

[25] In the sequel we omit the phrase "the first inequality in" and refer to the first part of condition (Y_i) simply as condition (Y_i).

the contrary that $\gamma^A_{LL} > 0$. Then $x^A_{LL} = k^A$. Furthermore, since condition $(5A)$ must hold as an equation, condition $(5B)$ can be met only if $\gamma^B_{LL} > 0$ (notice that $r^A_L q^A_L = r^B_L q^B_L$ and $c^A_L \geq c^B_L$) so that $x^B_{LL} = k^B$. But $x^A_{LL} = k^A$ and $x^B_{LL} = k^B$ together violate (D_{LL}). The fact that $\gamma^A_{LL} = 0$, in turn, implies that $\lambda_{LL} = r^A_L q^A_L (p_x - c^A_L)$ and $\gamma^B_{LL} = r^B_L q^B_L (c^A_L - c^B_L)$ so that $x^B_{LL} = k^B$ and $x^A_{LL} = X - k^B$ if $c^A_L > c^B_L$, and $x^B_{LL} \in [X - k^A, k^B]$ and $x^A_{LL} = X - x^B_{LL}$ otherwise (i.e., if $c^A_L = c^B_L$). For the rest we have to distinguish between the symmetric and the asymmetric case. In the *symmetric case* the above arguments can be reversed to show that $\gamma^B_{HL} = 0$ and $\gamma^A_{LH} = r^A_L(1 - q^A_L)(c^B_H - c^A_L) + r^B_L q^B_L \Delta^B$ implying $x^A_{LH} = k^A$ and $x^B_{HL} = X - k^A$. Furthermore, $\gamma^A_{HH} = \gamma^B_{HH} = 0$ (since $x^A_{HH} = k^A$ and $x^B_{HH} = k^B$ together violate (D_{HH})), and $\lambda_{HH} = r^i_H(1 - q^i_H)(p_x - c^i_H) - r^i_L(1 - q^i_L)\Delta^i$ so that $x^A_{HH} \in [X - k^B, k^A]$ and $x^B_{HH} = X - x^A_{HH}$. The *asymmetric case* is more complicated. First consider conditions $(8A)$ and $(8B)$. Three different cases have to be considered. If $\delta \equiv r^A_H(1 - q^A_H)(c^B_H - c^A_H) + r^B_L(1 - q^B_L)\Delta^B - r^A_L(1 - q^A_L)\Delta^A < 0$, then $\gamma^A_{HH} = 0$, $\lambda_{HH} = r^A_H(1 - q^A_H)(p_x - c^A_H) - r^A_L(1 - q^A_L)\Delta^A$ and $\gamma^B_{HH} = |\delta|$ by arguments similar to those presented above. Thus, $x^B_{HH} = k^B$ and $x^A_{HH} = X - k^B$. Symmetric arguments reveal that $\gamma^B_{HH} = 0$, $\lambda_{HH} = r^B_H(1 - q^B_H)(p_x - c^B_H) - r^B_L(1 - q^B_L)\Delta^B$ and $\gamma^A_{HH} = \delta$ so that $x^A_{HH} = k^A$ and $x^B_{HH} = X - k^A$ whenever $\delta > 0$. Finally assume that $\delta = 0$. Then $\gamma^A_{HH} = \gamma^B_{HH} = 0$ and $\lambda_{HH} = r^A_H(1 - q^A_H)(p_x - c^A_H) - r^A_L(1 - q^A_L)\Delta^A = r^B_H(1 - q^B_H)(p_x - c^B_H) - r^B_L(1 - q^B_L)\Delta^B$ so that $x^A_{HH} \in [X - k^B, k^A]$ and $x^B_{HH} = X - x^A_{HH}$, since $(8A)$ and $(8B)$ holding as equalities implies $\gamma^A_{HH} = \gamma^B_{HH}$ in this case, and since $\gamma^A_{HH} > 0$ and $\gamma^B_{HH} > 0$ together violate (D_{HH}). Applying the same logic to conditions $(6A)$ and $(7B)$ reveals that $\gamma^A_{LH} = 0$, $\lambda_{LH} = r^A_L(1 - q^A_L)(p_x - c^A_L)$, $\gamma^B_{HL} = |\xi|$, $x^B_{HL} = k^B$ and $x^A_{LH} = X - k^B$ whenever $\xi \equiv r^A_L(1 - q^A_L)(c^B_H - c^A_L) + r^B_L q^B_L \Delta^B < 0$, that $\gamma^B_{HL} = 0$, $\lambda_{LH} = r^B_H q^B_H (p_x - c^B_H) - r^B_L q^B_L \Delta^B$, $\gamma^A_{LH} = \xi$, $x^A_{LH} = k^A$ and $x^B_{HL} = X - k^A$ if $\xi > 0$, and that $\gamma^B_{HL} = \gamma^A_{LH} = 0$, $\lambda_{LH} = r^A_L(1 - q^A_L)(p_x - c^A_L) = r^B_H q^B_H (p_x - c^B_H) - r^B_L q^B_L \Delta^B$, $x^A_{LH} \in [X - k^B, k^A]$ and $x^B_{HL} = X - x^A_{LH}$ if $\xi = 0$.

Case 3: $k^A < X \leq k^B$. In this case we get again $\lambda_{mn} > 0$ for all $(m,n) \in \{L,H\}^2$ (by the arguments in the proof of Case 2). In addition, $\gamma^B_{mn} = 0$ for all $(m,n) \in \{L,H\}^2$ (to see this, suppose that $\gamma^B_{mn} > 0$; then $x^B_{mn} = k^B$, which violates the respective (D) constraint) and $x^B_{mn} > 0$ for all $(m,n) \in \{L,H\}^2$ (this follows from $\lambda_{mn} > 0$ for all $(l,m) \in \{L,H\}^2$ and from $k^A < X$). Also, $\gamma^A_{LL} = 0$ (if $\gamma^A_{LL} > 0$ then $x^A_{LL} = k^A > 0$; this implies that condition $(5A)$ holds as an equation; if condition $(5A)$ holds as an equation, condition $(5B)$ can be satisfied only if $\gamma^B_{LL} > 0$; but this contradicts $\gamma^B_{LL} = 0$) and $\gamma^A_{HL} = 0$ (along similar lines of argument) so that $\lambda_{LL} = r^B_L q^B_L(p_x - c^B_L) \geq r^A_L q^A_L(p_x - c^A_L)$ and $\lambda_{HL} = r^B_L(1-q^B_L)(p_x - c^B_L) > r^A_H q^A_H(p_x - c^A_H) - r^A_L q^A_L \Delta^A$. This, in turn, implies that $x^A_{HL} = 0$ and $x^B_{LH} = X$. Furthermore, $x^B_{LL} = X$ and $x^A_{LL} = 0$ if $c^A_L > c^B_L$, and $x^B_{LL} \in [X - k^A, X]$ and $x^A_{LL} = X - x^B_{LL}$ otherwise (i.e., if $c^A_L = c^B_L$). For the rest we again have to distinguish between the symmetric and the asymmetric case. In the *symmetric case* $\gamma^A_{LH} > 0$ (to see this, suppose that $\gamma^A_{LH} = 0$; then follow the line of arguments presented in the proof of $\gamma^A_{LH} > 0$ in Case 2 above) so that $x^A_{LH} = k^A$, $x^B_{HL} = X - k^A$, $\lambda_{LH} = r^B_H q^B_H(p_x - c^B_H) - r^B_L q^B_L \Delta^B$ and $\gamma^A_{LH} = r^A_L(1-q^A_L)(c^B_H - c^A_L) + r^B_L q^B_L \Delta^B$. Furthermore, $\gamma^A_{HH} = 0$, $\lambda_{HH} = r^i_H(1-q^i_H)(p_x - c^i_H) - r^i_L(1-q^i_L)\Delta^i$, $x^A_{HH} \in [0, k^A]$ and $x^B_{HH} = X - x^A_{HH}$. In the *asymmetric case* we have again to distinguish lots of subcases. First consider conditions $(6A)$ and $(7B)$. From $\gamma^B_{HL} = 0$, $x^B_{HL} > 0$ and $x^A_{LH} \geq 0$ we get $\lambda_{LH} = r^B_H q^B_H(p_x - c^B_H) - r^B_L q^B_L \Delta^B \geq r^A_L(1-q^A_L)(p_x - c^A_L) - \gamma^A_{LH}$. If $\xi < 0$ then this relation implies that $\gamma^A_{LH} = 0$, $x^A_{LH} = 0$ and $x^B_{HL} = X$. Similary, if $\xi > 0$ then $\gamma^A_{LH} = \xi$, $x^A_{LH} = k^A$ and $x^B_{HL} = X - k^A$. And if $\xi = 0$ then $\gamma^A_{LH} = 0$, the above relation holds as an equality and we get $x^A_{LH} \in [0, k^A]$ and $x^B_{HL} = X - x^A_{LH}$. Next consider conditions $(8A)$ and $(8B)$. From $\gamma^B_{HH} = 0$, $x^B_{HH} > 0$ and $x^A_{HH} \geq 0$ we get $\lambda_{HH} = r^B_H(1-q^B_H)(p_x - c^B_H) - r^B_L(1-q^B_L)\Delta^B \geq r^A_H(1-q^A_H)(p_x - c^A_H) - r^A_L(1-q^A_L)\Delta^A - \gamma^A_{HH}$. If $\delta < 0$ then this relation implies $\gamma^A_{HH} = 0$, $x^A_{HH} = 0$ and $x^B_{HH} = X$. If $\delta > 0$ then $\gamma^A_{HH} = \delta$, $x^A_{HH} = k^A$ and $x^B_{HH} = X - k^A$. And if $\delta = 0$, then this relation holds as an equation and $x^A_{HH} \in [0, k^A]$ and $x^B_{HH} = X - x^A_{HH}$.

Case 4: $X \leq k^A$. In this case conditions (D_{LL}) - (D_{HH}) are more demanding than conditions (K^i_{LL}) to (K^i_{HH}) so that $\gamma^i_{LL} = \gamma^i_{LH} = \gamma^i_{HL} = \gamma^i_{HH} = 0$ for $i = A, B$. Now remember that $c^A_m \geq c^B_m$ for $m \in \{L, H\}$. Therefore, condition $(6B)$ implies that condition $(7A)$ holds with strict inequality so that $\lambda_{HL} = r^B_L(1 - q^B_L)(p_x - c^B_L)$, $x^A_{HL} = 0$ and $x^B_{LH} = X$. For the rest we distinguish again between the symmetric and the asymmetric case. In the *symmetric case* condition $(6A)$ implies that condition $(7B)$ holds as a strict inequality so that $\lambda_{LH} = r^A_L(1 - q^A_L)(p_x - c^A_L)$, $x^B_{HL} = 0$ and $x^A_{LH} = X$. Under Assumption 2 the other conditions hold as equations and we get $\lambda_{LL} = r^i_L q^i_L(p_x - c^i_L)$, $\lambda_{HH} = r^i_H(1 - q^i_H)(p_x - c^i_H) - r^i_L(1 - q^i_L)\Delta^i$, $x^A_{mm} \in [0, X]$ and $x^B_{mm} = X - x^A_{mm}$ for $m \in \{L, H\}$. In the *asymmetric case* $(c^A_L > c^B_L)$ condition $(5B)$ implies that condition $(5A)$ holds as a strict inequality so that $\lambda_{LL} = r^B_L q^B_L(p_x - c^B_L)$, $x^A_{LL} = 0$ and $x^B_{LL} = X$. Carrying out the same type of analysis with conditions $(8A)$ and $(8B)$ reveals that $\lambda_{HH} = r^B_H(1 - q^B_H)(p_x - c^B_H) - r^B_L(1 - q^B_L)\Delta^B > r^A_H(1 - q^A_H)(p_x - c^A_H) - r^A_L(1 - q^A_L)\Delta^A$, $x^A_{HH} = 0$ and $x^B_{HH} = X$ if $\delta < 0$, $\lambda_{HH} = r^A_H(1 - q^A_H)(p_x - c^A_H) - r^A_L(1 - q^A_L)\Delta^A > r^B_H(1 - q^B_H)(p_x - c^B_H) - r^B_L(1 - q^B_L)\Delta^B$, $x^A_{HH} = X$ and $x^B_{HH} = 0$ if $\delta > 0$, and $\lambda_{HH} = r^i_H(1 - q^i_H)(p_x - c^i_H) - r^i_L(1 - q^i_L)\Delta^i$, $x^A_{HH} \in [0, X]$ and $x^B_{HH} = X - x^A_{HH}$ if $\delta = 0$. Similarly, from conditions $(6A)$ and $(7B)$, $\lambda_{LH} = r^B_H q^B_H(p_x - c^B_H) - r^B_L q^B_L \Delta^B > r^A_L(1 - q^A_L)(p_x - c^A_L)$, $x^A_{LH} = 0$ and $x^B_{HL} = X$ if $\xi < 0$, $\lambda_{LH} = r^A_L(1 - q^A_L)(p_x - c^A_L) > r^B_H q^B_H(p_x - c^B_H) - r^B_L q^B_L \Delta^B$, $x^A_{LH} = X$ and $x^B_{HL} = 0$ if $\xi > 0$, and $\lambda_{LH} = r^A_L(1 - q^A_L)(p_x - c^A_L) = r^B_H q^B_H(p_x - c^B_H) - r^B_L q^B_L \Delta^B$, $x^A_{LH} \in [0, X]$ and $x^B_{HL} = X - x^A_{LH}$ if $\xi = 0$.

It remains to be verified that at a solution to the relaxed program the missing "downward" incentive constraints IC^A_H and IC^B_H are satisfied. To see this, first remember that $\alpha^i_{HL} > 0$ and $\alpha^i_{HH} > 0$ so that $u_{HL} = 0$ and $u_{HH} = 0$. Substitititing into (IC_H) and (IC_L) and taking into account that $\beta^i_L > 0$ yields

$$(9) \qquad q_L u_{LL} + (1 - q_L)u_{LH} = [q_L x_{HL} + (1 - q_L)x_{HH}]\Delta > 0;$$

$$(10) \qquad q_H u_{LL} + (1 - q_H)u_{LH} \leq [q_H x_{LL} + (1 - q_H)x_{LH}]\Delta.$$

To be admissible as a solution to our problem, u_{LL} and u_{LH} have to satisfy (9), (10), (IR_{LH}) and (IR_{LL}). It is straightforward to verify that there exists a continuum of values for u_{LL} and u_{LH} which satisfy these conditions. In the symmetric solution to the symmetric case, where $x^A_{mm} = x^B_{mm}$ for $m \in \{L, H\}$, the desired outcome can even be implemented in dominant strategies. This is accomplished by setting $u_{LL} = x_{HL}\Delta$ and $u_{LH} = x_{HH}\Delta$. Conditions (9), (IR_{LL}) and (IR_{LH}) are obviously satisfied; and $x_{LL} > x_{HL}$ and $x_{LH} > x_{HH}$ implies that $q_L x_{HL} + (1 - q_L)x_{HH} < q_H x_{LL} + (1 - q_H)x_{LH}$ so that (10) holds as a strict inequality.

Proof of Corollary 1

The shadow value of additional units of capacity is calculated by summing up the Langrangian multipliers γ^i_{mn} over all $(m, n) \in \{L, H\}^2$ for the different cases considered in the proof of Lemma 1.

Proof of Proposition 2

Let $x\phi(x)$ denote the total expected rent paid from the headquarters to the agents if the production level is $x = x^A + x^B$. Then,

$$x\phi(x) = \sum_{i\in\{A,B\}} r^i_L[q^i_L u^i_{LL} + (1 - q^i_L)u^i_{LH}] = \sum_{i\in\{A,B\}} r^i_L[q^i_L x^i_{HL} + (1 - q^i_L)x^i_{HH}]\Delta^i.$$

By symmetry, $r^A_L = r^B_L = r_L$, $q^A_L = q^B_L = q_L$ and $\Delta^A = \Delta^B = \Delta$. Inserting the optimal values for x^i_{HL} and x^i_{HH} $(i \in \{A, B\})$ from Lemma 1 into $\phi(x)$ and taking into account that $k^A = k^B = k$, say, yields the increasing function

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

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

Proof of Proposition 3

Substituting the optimal values for u^i_{mn} and x^i_{mn} for all $i \in \{A, B\}$ and all $(m, n) \in \{L, H\}^2$ in the objective function NR and taking into account that $k^A = k^B = k$ yields the following reduced form profit function:

$$\Pi(X,k) = \begin{cases} X[p_x - (1-r_H(1-q_H))c_L - r_H(1-q_H)c_H - r_L(1-q_L)\Delta] & \text{for } X \in [\underline{X}, k] \\ p_x X - r_L(q_L X + (1-q_L)2k)c_L - r_H(q_H 2(X-k) + & \\ \qquad (1-q_H)X)c_H - r_L(q_L 2(X-k) + (1-q_L)X)\Delta & \text{for } X \in (k, 2k) \\ 2k[p_x - r_L(c_L+\Delta) - r_H c_H] & \text{for } X \in [2k, \overline{X}] \end{cases}$$

Taking the derivative with respect to k gives

$$\Pi'(X,k) = \begin{cases} 0 & \text{for } X \in [\underline{X}, k] \\ 2[r_L(1-q_L)(c_H - c_L) + r_L q_L \Delta] & \text{for } X \in (k, 2k) \\ 2[p_x - r_L(c_L+\Delta) - r_H c_H] & \text{for } X \in [2k, \overline{X}] \end{cases}$$

For future reference define

$$\begin{aligned} \alpha &= p_x - r_L(c_L+\Delta) - r_H c_H; \\ \xi &= r_L(1-q_L)(c_H - c_L) + r_l q_L \Delta. \end{aligned} \tag{11}$$

The next step is to take the expectation over $[\underline{X}, \overline{X}]$. We denote the resulting function by $\Pi'(k)$. Two different cases have to be considered:

Case 1: $2\underline{X} \geq \overline{X}$. In this case $\Pi'(k)$ is given by

$$\Pi'(k) = \begin{cases} 2\alpha & \text{for} \quad 2k \leq \underline{X} \\ 2[\alpha(\overline{X} - 2k) + \xi(2k - \underline{X})]/(\overline{X} - \underline{X}) & \text{for} \quad 2k \in [\underline{X}, \overline{X}] \\ 2\xi & \text{for} \quad 2k \in [\overline{X}, 2\underline{X}] \\ 2\xi(\overline{X} - k)/(\overline{X} - \underline{X}) & \text{for} \quad 2k \in [2\underline{X}, 2\overline{X}] \\ 0 & \text{for} \quad 2k \geq 2\overline{X} \end{cases}$$

Case 2: $2\underline{X} < \overline{X}$. In this case $\Pi'(k)$ is given by

$$\Pi'(k) = \begin{cases} 2\alpha & \text{for} \quad 2k \leq \underline{X} \\ 2[\alpha(\overline{X} - 2k) + \xi(2k - \underline{X})]/(\overline{X} - \underline{X}) & \text{for} \quad 2k \in [\underline{X}, 2\underline{X}] \\ 2[\alpha(\overline{X} - 2k) + \xi k]/(\overline{X} - \underline{X}) & \text{for} \quad 2k \in [2\underline{X}, \overline{X}] \\ 2\xi(\overline{X} - k)/(\overline{X} - \underline{X}) & \text{for} \quad 2k \in [\overline{X}, 2\overline{X}] \\ 0 & \text{for} \quad 2k \geq 2\overline{X} \end{cases}$$

The corresponding function for the first best benchmark (we mark this function with an asterisk) is obtained by setting the term Δ in α and ξ (as defined in (11)) equal to zero. An optimal value of k (denoted by $\hat{k}$ or k^*, respectively) is found by setting $\Pi'(k)$ (or $\Pi^{*\prime}(k)$, respectively) equal to $2p_k$[26]. To verify the proposition, first notice that in both cases the value of the function $\Pi^{*\prime}(k)$ is strictly higher than the value of $\Pi'(k)$ for all $k \in [0, .5\underline{X}]$ and that the value of $\Pi^{*\prime}(k)$ is strictly lower than that of $\Pi'(k)$ for all $k \in [.5\overline{X}, \overline{X})$. Existence of a critical price of capacity $\tilde{p}_k$ then follows from continuity of the $\Pi'(\cdot)$'s and uniqueness from the fact that the difference $\Pi^{*\prime}(k) - \Pi'(k)$ is strictly decreasing for $k \in (.5\underline{X}, .5\overline{X})$.

Proof of Proposition 4

Adapting the derivate $\Pi'(k)$ derived in the proof of Proposition 3 to allow for $\mu > 0$ (we denote the adapted derivate by $\Pi'(k, \mu)$)[27], setting this derivate equal to $2p_k$, and solving for $\hat{k}(p_k, \mu)$ yields

[26] Since $\Pi^{*\prime}(0) > \Pi'(0) > 0$ and $\Pi^{*\prime}(2\overline{X}) = \Pi'(2\overline{X}) = 0$ and since the $\Pi'(\cdot)$'s are continuous, an interior solution exists for all $p_k \in (0, \alpha)$. Since the $\Pi'(\cdot)$'s are only decreasing but not strictly decreasing the solution might not be unique.

[27] The formulas in the proof of Proposition 3 can easily be adjusted by replacing $\underline{X}$ in the terms and bounds by $\underline{X} - \mu$, and $\overline{X}$ by $\overline{X} + \mu$. For example, $\overline{X} - \underline{X}$ changes to $\overline{X} - \underline{X} + 2\mu$ and $\overline{X} - (2 + q_L)\underline{X}$ to $\overline{X} - (2 + q_L)\underline{X} + (3 + q_L)\mu$.

Case 1: $2(\underline{X} - \mu) \geq \overline{X} + \mu$.

$$\hat{k}(p_k, \mu) = \begin{cases} \in [0, .5(\underline{X} - \mu)] & \text{if} \quad p_k = \alpha \\ \frac{\overline{X}(\alpha - p_k) + \underline{X}(p_k - \xi) + \mu(\alpha + \xi - 2p_k)}{2(\alpha - \xi)} & \text{if} \quad p_k \in (\xi, \alpha) \\ \in [.5(\overline{X} + \mu), \underline{X} - \mu] & \text{if} \quad p_k = \xi \\ \frac{\overline{X}(\xi - p_k) + \underline{X} p_k + \mu(\xi - 2p_k)}{\xi} & \text{if} \quad p_k \in (0, \xi) \end{cases}$$

Case 2: $2(\underline{X} - \mu) < \overline{X} + \mu$.

$$\hat{k}(p_k, \mu) = \begin{cases} \in [0, .5(\underline{X} - \mu)] & \text{if} \quad p_k = \alpha \\ \frac{\overline{X}(\alpha - p_k) + \underline{X}(p_k - \xi) + \mu(\alpha + \xi - 2p_k)}{2(\alpha - \xi)} & \text{if} \quad p_k \in (\frac{\alpha \overline{X} - \underline{X}(2\alpha - \xi) + \mu(3\alpha - \xi)}{\overline{X} - \underline{X} + 2\mu}, \alpha) \\ \frac{\overline{X}(\alpha - p_k) + \underline{X} p_k + \mu(\alpha - 2p_k)}{2\alpha - \xi} & \text{if} \quad p_k \in [\frac{\xi(\overline{X} + \mu)}{2(\overline{X} - \underline{X} + 2\mu)}, \frac{\alpha \overline{X} - \underline{X}(2\alpha - \xi) + \mu(3\alpha - \xi)}{\overline{X} - \underline{X} + 2\mu}] \\ \frac{\overline{X}(\xi - p_k) + \underline{X} p_k + \mu(\xi - 2p_k)}{\xi} & \text{if} \quad p_k \in (0, \frac{\xi(\overline{X} + \mu)}{2(\overline{X} - \underline{X} + 2\mu))}) \end{cases}$$

The corresponding function for the first best benchmark is again obtained by setting the term Δ in the definition of α and ξ equal to zero. To verify the proposition, consider Case 1. Here, the derivative of $\hat{k}(p_k, \mu)$ with respect to μ is strictly positive for $p_k \in (0, .5\xi)$, strictly negative for $p_k \in (.5\xi, \xi)$, again positive for $p_k \in (\xi, .5(\alpha + \xi))$ and again negative for $p_k \in (.5(\alpha + \xi), \alpha)$. Case 2 behaves similarly.[28]

Proof of Proposition 5

Let $\tilde{k}$ be defined by the equation $\Pi'(\tilde{k}) = \Pi^{*\prime}(\tilde{k})$. Using the formulas derived in the proof of Proposition 3 it is easily verified that for $\mu = 0$, $\tilde{k}$ is uniqueley determined by $r_L \Delta(\overline{X} - 2\tilde{k}) = r_L q_L \Delta(2\tilde{k} - \underline{X})$ if $\overline{X} < \underline{X}(2 + q_L)$ and by $r_L \Delta(\overline{X} - 2\tilde{k}) = r_L q_L \Delta \tilde{k}$ otherwise. For $\mu > 0$ the corresponding equations are $r_L \Delta(\overline{X} - 2\tilde{k} + \mu) = r_L q_L \Delta(2\tilde{k} - \underline{X} + \mu)$ if $\overline{X} < \underline{X}(2 + q_L) - \mu(3 + q_L)$ and $r_L \Delta(\overline{X} - 2\tilde{k} + \mu) = r_L q_L \Delta \tilde{k}$ otherwise. Solving these

[28]In Case 2, four subcases have to be distinguished depending on whether the terms $(\overline{X} + \mu) - 3(\underline{X} - \mu)$ and $(\alpha - \xi)(\overline{X} + \mu) - \alpha(\underline{X} - \mu)$ are positive or negative. In each of these subcases the derivative of $\hat{k}(p_k, \mu)$ with respect μ changes sign at least once.

equations for $\tilde{k}$ yields

$$\tilde{k} = \begin{cases} \frac{\overline{X}+\underline{X}q_L+\mu(1-q_L)}{2(1+q_L)} & \text{if } \overline{X}+\mu \leq (\underline{X}-\mu)(2+q_L) \\ \frac{\overline{X}+\mu}{2+q_L} & \text{otherwise.} \end{cases}$$

Substituting these expressions in the respective marginal benefit functions yields

$$\tilde{p}_k(\mu) = \begin{cases} \frac{\xi+\alpha q_L}{1+q_L} & \text{if } \overline{X}+\mu \leq (\underline{X}-\mu)(2+q_L) \\ \frac{\xi+\alpha q_L}{2+q_L} \cdot \frac{\overline{X}+\mu}{\overline{X}-\underline{X}+2\mu} & \text{otherwise,} \end{cases}$$

where α and ξ are as defined in (11). As is easily verified, $\tilde{p}_k(\mu)$ is decreasing in μ for all $\mu \in (0, \underline{X})$. It is strictly decreasing in μ if $\overline{X} > \underline{X}(2+q_L) - \mu(3+q_L)$. A simple qualification under which this condition holds as a strict inequality for arbitrary $(\underline{X}, \overline{X})$-pairs is $\mu > .75\underline{X} - .25\overline{X}$. If, on the other hand, $\overline{X} > \underline{X}(2+q_L)$, then this condition holds as a strict inequality for any $\mu \geq 0$.

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 P. Gertler, "Regulation, Local Monopolies and Spatial Competition" *Mimeo, Fuqua School of Business, Duke University* (1994).

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

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

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

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

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

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

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

[11] Gali, J. and M. Hammour, "Long Run Effects of Business Cycles", *Working Paper FB-9226, Columbia University, New York* (1992).

[12] Grossman, S. and O. Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", *Journal of Political Economy* **94** (1986), 691-719.

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

[14] Hart, O., "The Market Mechanism as an Incentive Scheme", *Bell Journal of Economics* **14** (1983), 366-382.

[15] Hermalin, B., "The Effects of Competition on Executive Behavior", *Rand Journal of Economics* **23** (1992), 350-365.

[16] Horn, H., H. Lang and S. Lundgren, "Competition, Long Run Contracts and Internal Inefficiencies in Firms", *European Economic Review* **38** (1994), 213-233.

[17] Laffont, J. J. and J. Tirole, "Auctioning Incentive Contracts", *Journal of Political Economy* **95** (1987), 921-937.

[18] Malley, J. and V. Muscatelli, "Business Cycles and Productivity Growth: Are Temporary Downturns Productive or Wasteful?", *Discussion Paper 9604, University of Glasgow* (1996).

[19] McAfee, R. P. and J. Mcmillan, "Bidding for Contracts: A Principal Agent Analysis", *Rand Journal of Economics* **17** (1986), 326-338.

[20] McGuire, T. and M. Riordan, "Incomplete Information and Optimal Market Structure: Public Purchases from Private Providers", *Journal of Public Economics* **56** (1995), 125-141.

[21] Martin, S., "Endogenous Firm Efficiency in a Cournot Principal-Agent Model", *Journal of Economic Theory* **59** (1993), 445-450.

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

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

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

[25] Riordan, M. and D. Sappington, "Awarding Monopoly Franchises", *American Economic Review*, **77** (1987), 375-387.

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

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

[28] Scharfstein, D., "Product Market Competition and Managerial Slack", *Rand Journal of Economics* **14** (1988), 147-155.

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

[30] Williamson, O., *"Markets and Hierarchies: Analysis and Antitrust Implications"*, New York, Free Press, 1975.