Capacity Management, Decentralization, and Internal Pricing

Sunil Dutta

Haas School of Business University of California, Berkeley

and

Stefan Reichelstein*

Graduate School of Business Stanford University

October 2008

^{*}We would like to thank seminar participants at the University of Texas at Austin, Harvard University and the Indian School of Business for their comments on an earlier version of this paper, titled "Decentralized Capacity Management" We are also grateful to Alexander Nezlobin for his research assistance.

Abstract: Capacity Management, Decentralization, and Internal Pricing

This paper studies the acquisition and subsequent utilization of production capacity in a decentralized firm comprised of multiple divisions. A central issue in our analysis is whether one or several of the divisions should be structured as investment centers with comprehensive responsibility for acquiring new capacity assets and maintaining existing ones. Ownership of capacity assets may alternatively be centralized, with the provision that the divisions rent capacity from a central unit on a period-by-period basis. The structure of responsibility centers is naturally related to the choice of internal pricing rules for capacity services. We find that an investment center arrangement combined with transfer prices set at full historical cost is efficient if in the short run there is effectively no flexibility in the divisional capacity assignments. In contrast, capacity may be fungible in the sense that even in the short-run the aggregate capacity available can be reassigned in response to fluctuations in the divisional revenues. When capacity is fungible, we identify the advantages of centralized capacity ownership. We also find that under certain conditions a system of negotiated transfer pricing is preferable to cost-based prices because negotiation leads to better coordination of the divisional capacity choices.

1 Introduction

A significant portion of firms' investment expenditures pertain to investments in production capacity. One distinctive characteristic of investments in plant and equipment is that they are long-lived and irreversible. Once the investment expenditure has been incurred, it is usually sunk due to a lack of markets for used assets. The longevity of capacity investments also causes their profitability to be subject to significant uncertainty. Fluctuations in the business environment over time make it generally difficult to predict at the outset whether additional capacity will be fully utilized and, if so, how valuable it will be.¹

The acquisition of new capacity and its subsequent utilization is an even more challenging issue for firms that comprise multiple business units. A prototypical example involves an upstream division which acquires production capacity for its own use and that of one or several downstream divisions which receive manufacturing services from the upstream division. Potential fluctuations in the revenues attainable to the individual divisions make it essential to have a coordination mechanism for balancing the firm-wide demands on capacity. Any such capacity management system must specify "control rights" over existing capacity, responsibility for acquiring new capacity and internal pricing rules to support intrafirm transactions.²

In our model of a two-divisional firm, an upstream division installs and maintains the firm's assets that create production capacity. This arrangement may reflect technical expertise on the part of the upstream division. One natural responsibility center arrangement therefore is to make the upstream division an investment center. Thus, capacity related assets are recorded on the balance sheet of the upstream division, while the downstream division is structured as a profit center that rents capacity from its sister division. As a benchmark result, we identify conditions for such a decentralized structure to result in ef-

¹Capacity choice under uncertainty has been a topic of extensive research in operations management. Traditionally, most of this literature has focused on the problem faced by a single decision-maker seeking to optimize a single investment decision. More recent work has addressed the question of capacity management in multi-agent and multi-period environments; see, for example, Porteus and Whang (1991), Kouvelis and Lariviere (2000) and Van Mieghem (2003). The work by Plambeck and Taylor (2005) on the incentives of contract manufacturers is in several respects closest in spirit to our study.

²The case study by Bastian and Reichelstein (2004) illustrates coordination issues related to capacity utilization at a bearings manufacturer. Martinez-Jerez (2007) describes a new customer profitability measurement system at Charles Schwab. A central issue for the company is how different user groups should be charged for IT related capacity costs.

ficient outcomes provided the downstream division rents capacity in each period at a full cost transfer price which includes depreciation and imputed capital charges for past capacity investments. Specifically, our benchmark result shows that the multi-period game corresponding to this organizational structure has a unique equilibrium with the property that each division's capacity usage is efficient from the firm-wide perspective.

The common reliance on full-cost (transfer) pricing in practice has been a challenge for research in managerial accounting. In most model settings, the use of full-cost prices is predicted to result in double-marginalization.³ The key to the efficiency of full-cost prices in our framework is that the firm makes a sequence of overlapping capacity investments. As shown by Arrow (1964), this dynamic structure makes it possible to identify the marginal cost of one unit capacity for one period of time, despite the inherent jointness that results when each investment creates productive capacity for multiple periods.

Recent work by Rogerson (2008) has shown that the marginal cost of capacity can be captured precisely by a particular set of historical cost charges. Investment expenditures can be allocated over time so that the sum of depreciation charges and imputed interest on the book value of assets is exactly equal to the marginal cost of another unit of capacity in that period. This equivalence requires that investment expenditures be apportioned over time according to what we term the *Relative Practical Capacity* rule. Accordingly, the expenditure for new assets is apportioned in proportion to the capacity available in a given period, relative to the total (discounted) capacity generated over the life of the asset.⁴ The resulting historical cost charges can also be viewed as competitive rental prices: a firm that acquires capacity and then hypothetically rents it out to third parties on a periodic basis would charge these prices if its business is subject to a zero-profit constraint. Thus, the effect of proper intertemporal cost allocations, coupled with full cost transfer prices, is that both divisions are effectively charged the competitive rental prices for capacity in each period.

Our benchmark result on the efficiency of full cost transfer pricing for capacity transfers is obtained in a setting where both divisions' capacity usage is determined at the beginning

³See, for instance, Balakrishnan and Sivaramakrishnan (2002), Goex (2002), Sahay (2003), Wei (2004) and Pfeiffer et al. (2007).

⁴The Relative Practical Capacity rule is conceptually similar to the so-called *Relative Benefit Rule* (Rogerson, 1997), which has played a prominent rule in the literature on performance measurement for investment projects. As the name suggests, though, the relative benefit rule applies to generic investment projects and seeks to match expected future cash inflows with a share of the investment expenditure.

of each period. In contrast to such a setting of *dedicated* capacity, there may be enough flexibility to redeploy the aggregate capacity that is available in the short run, once the divisional managers have received updated information on the revenue opportunities of their respective divisions. When capacity is *fungible*, it is natural to allow the divisional managers to negotiate an adjustment to the initial capacity rights. We refer to the resulting mechanism as adjustable full cost transfer pricing.

By giving divisional managers discretion to negotiate a reallocation of the initial capacity rights, the firm captures trading gains that arise from fluctuations in the divisional revenues. At the same time, we find that the resulting system of adjustable full cost transfer pricing subjects the upstream division to a *dynamic holdup* problem. Since the downstream division only rents capacity in each period, it may have an incentive to drive up its capacity demands opportunistically in one period in anticipation of obtaining the corresponding excess capacity at a low cost through negotiations in future periods. In essence, this dynamic hold-up problem reflects that the downstream division is not accountable for the long-term effect of irreversible capacity demands, yet as an investment center the upstream division cannot divest itself from the corresponding assets and the corresponding fixed cost charges.

To counteract the dynamic hold-up problem described above, the firm may centralize the ownership of capacity assets and regard both divisions as profit centers that with discretion to secure capacity for themselves at the competitive rental price. Since the central office can commit to a policy of full cost transfer pricing, neither division can game the system by securing excessive amounts of capacity. Nonetheless, we identify a *coordination problem* in the divisional capacity requests. While each division correctly internalizes the incremental cost of additional capacity, the corresponding benefits at the divisional level will generally not coincide with the overall benefit to the firm.

Under the rule of adjustable full cost transfer pricing, divisional incentives to secure capacity unilaterally arise from two sources: the autonomous use of capacity and a share of the overall firm-wide revenue that is obtained through negotiated adjustments. We find that the resulting divisional incentives are congruent with the firm-wide objective only in exceptional cases, e.g., the divisional revenue functions can be described by a quadratic function. Yet, in general there will be distortions which can bias the divisional decisions in either direction. In particular, we identify conditions for a system of adjustable full cost transfer pricing (with centralized capacity ownership) to result in over-investment of capacity. This result stands in contrast to earlier incomplete contracting models on transfer pricing.⁵

Our findings provide a rationale for the central office to require that the two divisions coordinate their capacity decisions. One approach to achieving coordination is to require that capacity rights for the downstream division cannot be secured unilaterally but must result from a mutually acceptable negotiation with the other division. Effectively, the upstream division becomes a "gatekeeper" who must approve capacity assignments to the downstream division. The divisions will then have congruent objectives to set the downstream division's capacity assignment at a level which subsequently leads the upstream division to implement a capacity level that is efficient from a firm-wide perspective.⁶ As a consequence, the negotiated transfer price for a unit of capacity will be different from the marginal cost to the firm, yet the resulting capacity investments will nonetheless be efficient.⁷

The conclusion that a negotiated capacity management system will efficiently coordinate the divisional capacity choices is broadly consistent with the findings in Edlin and Reichelstein (1995) and Wielenberg (2000). A complicating issue in these earlier studies on negotiated transfer pricing is that investments are non-verifiable for contracting purposes.⁸ As a consequence, the parties have to contract on a "surrogate" variable such as the quantity or quality of a product to be traded. Clearly, this issue does not arise in our framework where ex-ante agreements pertain to verifiable capacity rights. The main advantage of requiring the downstream division to "clear" its capacity choice with the upstream division is improved coordination. The implied efficiency advantage of negotiated transfer pricing must, of course, be weighed against other factors not accounted for in our analysis. These include

⁵See, for instance, such as Baldenius et al (1999), Anctil and Dutta (1999), Sahay (2003), Wei (2004) and Pfeiffer et al. (2008). In these models the divisions make relationship specific investments that have no value to the investor if the parties do not engage in trade, e.g., the upstream division lowers the unit cost of producing the intermediate product in question. As a consequence, the collective problem is one of mitigating hold-ups and avoiding under-investment.

⁶Without the ability of the downstream division to secure capacity, the upstream division would underinvest, as it would anticipate being held up in the subsequent negotiation.

⁷The notion that firms may want to bias internal prices deliberately is central to the literature on "strategic" transfer prices; see, for example Hughes and Kao (1997), Alles and Datar (1998) and Arya and Mittendorf (2008). In these studies, however, it is a central planner who "distorts" the internal price in order to achieve pre-commitment in the firm's competition with external rivals.

⁸This specification is, of course, crucial to most studies on incomplete contracting, including the pioneering contributions by Williamson (1985) and Grossman and Hart (1988). Tirole (2003) provides a survey of more recent contributions.

the cost of "haggling" and more stringent assumptions of symmetric information across the two divisions.

The remainder of the paper is organized as follows. The model is described in Section 2. Section 3 examines cost-based transfer pricing under the assumption that the upstream division is an investment center that "owns" the capacity assets. Section 4, considers an alternative organizational structure of centralized capacity ownership. We first examine the incentive properties of a cost-based transfer pricing and its possible distortions, and then investigate the efficiency of negotiated transfer pricing which makes the upstream division effectively a gatekeeper for new capacity acquisitions. Extensions of our basic model are provided in Section 5 and we conclude in Section 6.

2 Model Description

Consider a decentralized firm comprised of two divisions and a central office. The two divisions use a collection of common capital assets (capacity) to produce their respective outputs. Because of technical expertise, only the upstream division (Division 1) is in a position to install and maintain the entire productive capacity for both divisions. Our analysis therefore considers initially an organizational structure which views the upstream division as an investment center whose balance sheet reflects the historical cost of past capacity investments. In contrast, the downstream division (Division 2) is evaluated as a profit center.

Capacity could be measured either in hours or the amount of output produced. New capacity can be acquired at the beginning of each period. It is commonly known that the unit cost of capacity is v. Therefore, the cash expenditure of acquiring b_{t-1} units of capacity at date t-1, the beginning of period t, is given by:

$$C_t = v \cdot b_{t-1}$$

For reasons of notational and expositional parsimony, we assume that assets have a useful life of n = 2 periods. As argued in Section 5 below, all our results would be unchanged for a general useful life of n periods. If b_{t-1} units of capacity are installed at date t - 1, they become fully functional at date t. At the same time, the practical capacity declines to $b_{t-1} \cdot \beta$ at date t + 1. Thus, $\beta \leq 1$ denotes the rate at which the productivity of new capacity declines, possibly due to increased maintenance requirements. The capacity *stock* available for production in period t is therefore given by:

$$k_t = b_{t-1} + \beta \cdot b_{t-2},\tag{1}$$

with $k_0 = 0$. The total capacity available at the beginning of a period can be used by either of the two divisions. While Division 1 has control rights over this capacity, the internal pricing mechanisms we study in this paper allow the downstream division to secure capacity rights in each period, prior to the upstream division deciding on new acquisitions. By k_{2t} we denote the amount of capacity that Division 2 has reserved for itself in period t. By definition, $k_{1t} = k_t - k_{2t}$.

The actual capacity levels made available to the divisions are denoted by q_{it} . They may differ from the initial rights k_{it} to the extent that the two divisions can still trade capacity within a period. If q_{it} units of capacity are ultimately available to Division *i* in period *t*, the corresponding net-revenue is given by $R_i(q_{it}, \theta_{it}, \epsilon_{it})$.⁹ The divisional revenue functions are parameterized by the random vector $(\theta_{it}, \epsilon_{it})$, where the random vector $\theta_t \equiv (\theta_{1t}, \theta_{2t})$ is realized at the beginning of period *t* before the divisions choose their capacity levels for that period, while the random variables $\epsilon_t \equiv (\epsilon_{1t}, \epsilon_{2t})$ represent transitory shocks to the divisional revenues. These shocks materialize after the capacity for period *t* has been decided.

The net-revenue functions $R_i(q_{it}, \theta_{it}, \epsilon_{it})$ are assumed to be increasing and concave in q_{it} for each *i* and each *t*. At the same time, the marginal revenue functions:

$$R_{i}^{'}(q,\theta_{it},\epsilon_{it}) \equiv \frac{\partial R_{i}(q,\theta_{it},\epsilon_{it})}{\partial q}$$

are assumed to be increasing in both θ_{it} and ϵ_{it} .¹⁰ The random variables θ_t 's may be serially correlated. However, the transitory shocks $\{\epsilon_{it}\}$ are assumed to be identically and independently distributed across time; i.e., $Cov(\epsilon_{it}, \epsilon_{i\tau}) = 0$ for each $t \neq \tau$, though in

⁹If one thinks of q_{it} as the amount of output produced for Division *i*, then the net-revenue $R_i(\cdot)$ includes all variable costs of production.

¹⁰The specification that $R'_i(\cdot) > 0$ is always positive reflects that the divisions are assumed to derive positive "salvage value" from their capacity, even beyond the point where they obtain positive contribution margins from their products. We note that this specification is convenient technically, though all of our results still hold if the marginal net-revenues were to drop to zero for q_i sufficiently large.

any given period these shocks may be correlated across divisions; i.e., it is possible to have $Cov(\epsilon_{1t}, \epsilon_{2t}) \neq 0.$

One maintained assumption of our model is that the path of efficient investment levels has the property that the firm *expects* not to have excess capacity. Formally, this condition will be met if the productivity parameters are increasing for sure over time, that is: $\theta_{i,t+1} \ge \theta_{it}$ for all t.¹¹ As a consequence, the *expected* marginal revenues are nondecreasing over time, that is:

$$E_{\epsilon_i}[R'_i(q,\theta_{i,t+1},\epsilon_{i,t+1})] \ge E_{\epsilon_i}[R'_i(q,\theta_{it},\epsilon_{it})],$$
(2)

for all $q \ge 0$ while the *realized* marginal revenues $R'_i(q, \theta_{it}, \epsilon_{it})$ may fluctuate across periods.

At the beginning of period t, both managers observe the realization of the state vector $\theta_t = (\theta_{1t}, \theta_{2t})$.¹² This information is not available to the central office and provides the basic rationale for delegating the investment decisions. Given the realization of the information parameters θ_t , Division 2 can secure capacity rights, k_{2t} , for its own use in the current period. Thereafter Division 1 proceeds with the acquisition of new capacity b_{t-1} .

Capacity is considered fixed in the short run and therefore it is too late to increase capacity for the current period, once the demand shock ϵ_t has been realized. However, in what we term the *fungible capacity* scenario, it is still possible for the two divisions to negotiate an allocation of the currently available capacity $k_t \equiv k_{1t} + k_{2t}$. Let (q_{1t}, q_{2t}) denote the renegotiated capacity levels, with $q_{1t} + q_{2t} = k_t$. In contrast, the scenario of *dedicated capacity* presumes that the initial capacity assignments made at the beginning of each period cannot be changed because of longer lead times. Figure 1 depicts the sequence of events in a given period.



Figure 1: Events in Period t

¹¹For instance, θ may experience consistent growth such that $\theta_{t+1} = \theta_t \cdot (1 + \lambda_t)$ and the support of λ_t is a subset of the non-negative real numbers.

¹²As argued below, some of our results remain valid in their current form if the divisional managers have private information about their own division's revenue.

The main part of our analysis ignores issues of moral hazard and compensation and instead focuses on the choice of goal congruent performance measures for the divisions. Following the terminology in earlier literature, a performance measure is said to be goal congruent if it induces managers to make decisions that maximize the present value of firmwide cash flows. In our search for goal congruent performance measures, we take it as given that the downstream division is evaluated on the basis of its operating income which consists of its net-revenue, $R_2(\cdot)$, less an internal transfer payment for the capacity service it receives from the other division. In contrast, the upstream division is initially assumed to be an investment center, its financial performance is assumed to be measured by residual income:

$$\pi_{1t} = Inc_{1t} - r \cdot A_{1,t-1}.$$
(3)

Here A_{1t} denotes book value of capacity related assets at the end of period t, and r denotes the firm's cost of capital. The corresponding discount factor is denoted by $\gamma \equiv (1+r)^{-1}$.

The upstream division's measure of income is based on two accruals: the transfer price received from the downstream division and depreciation charges for past capacity investments. The depreciation schedule must satisfy the usual tidiness requirement that the depreciation charges over an asset's useful life add up to the asset's acquisition cost. We let the parameter d represent the depreciation charge at date t per dollar of capacity investment undertaken at date t - 1. The remaining book value $v \cdot b_{t-1} \cdot (1 - d)$ will be depreciated at date t + 1. Thus, the total depreciation charge for Division 1 in period t can be written as:

$$D_t = c \cdot [b_{t-1} \cdot d + b_{t-2} \cdot (1-d)], \tag{4}$$

and the historical cost value of the net assets at the end of period t is given by:

$$A_{1t} = v \cdot [b_t + (1 - d) \cdot b_{t-1}].$$
(5)

To achieve goal congruence, performance measures are required to be *robust* in the sense that the desired incentives hold regardless of the relative bargaining powers of the two managers and even if divisional managers attach weights to future outcomes that differ from those of the firm. Let $u_i = (u_{i1}, ..., u_{iT})$ denote non-negative weights that manager *i* attaches to the sequence of performance measures $\pi_i = (\pi_{1i}, ..., \pi_{iT})$. At the beginning of period 1, manager *i*'s objective function can thus be written as $\sum_{t=1}^{T} u_{it} \cdot E[\pi_{it}]$. One can think of the weights u_i as reflecting a manager's discount factor as well as the bonus coefficients attached to the periodic performance measures. We require goal congruence for all u_i in some open set in $\mathcal{V}_i \subset \mathbb{R}^T_+$. For instance, \mathcal{V}_i could be a neighborhood around $(u \cdot \gamma, u \cdot \gamma^2, ..., u \cdot \gamma^T)$ for some constant bonus coefficient u. Since the transfer pricing policies examined in this paper are based on different behavioral assumptions, including cooperative and non-cooperative notions of equilibrium, we defer the formal definition of strong goal congruence to later sections.

3 Decentralized Capacity Ownership

This section examines an organizational structure in which the upstream division is structured as an investment center. Accordingly, that division is in charge of all capacity investment decisions and its balance sheet reflects the historical cost of assets acquired in previous periods. In contrast, the downstream division rents capacity from the upstream division on a period-by-period basis. We initially focus on pricing rules which allow the downstream division to secure capacity at a price given by the *full cost* of capacity, comprised of historical depreciation charges plus imputed interest. In settings where the upstream division not only provides capacity services but also manufactures an intermediate product for the downstream division, the full cost transfer price would also include applicable variable costs associated with the intermediate product. Such an internal pricing rule appears consistent with the practice of full cost transfer pricing that features prominently in most surveys on transfer pricing.¹³

We first examine a scenario of *dedicated capacity*, in which the random shocks ϵ_t are realized so late in period t that it is impossible for the divisions to redeploy the available capacity stock k_t . Consequently, Division *i*'s initial capacity assignment k_{it} , made at the beginning of period t, is also equal to the capacity ultimately available for its use in that period. Put differently, capacity assignments can only be altered at the beginning of each period, but not within a period.

The firm is a going concern that seeks a path of efficient investment and capacity levels so as to maximize the stream of discounted future cash flows. If a central planner hypothetically

 $^{^{13}}$ See, for instance, Schnell (1995), Merchant (2000), Ernst & Young (2003) and Tang (2002).

had the entire information available to the divisional managers, the investment decisions $b \equiv (b_0, b_1, b_2, ...)$ would be chosen so as to maximize the net present value of the firm's expected future cash flows:

$$\Pi_d(b) = \sum_{t=1}^{\infty} [M_d(b_{t-1} + \beta \cdot b_{t-2}, \theta_t) - v \cdot b_t] \cdot \gamma^t,$$

subject to the non-negativity constraints $b_t \ge 0$. Here, $M_d(b_{t-1} + \beta \cdot b_{t-2}, \theta_t)$ denotes the maximized value of the firm-wide contribution margin:

 $E_{\epsilon} \left[R_1(k_{1t}, \theta_{1t}, \epsilon_{1t}) + R_2(k_{2t}, \theta_{2t}, \epsilon_{2t}) \right],$

subject to the constraint that $k_{1t} + k_{2t} \leq b_{t-1} + \beta \cdot b_{t-2}$.

Lemma 1 When capacity is dedicated, the optimal capacity levels, $(\bar{k}_{1t}^o, \bar{k}_{2t}^o)$, are given by:

$$E_{\epsilon_i} \left[R'_i(\bar{k}^o_{it}, \theta_{it}, \epsilon_{it}) \right] = c, \tag{6}$$

where

$$c = \frac{v}{\gamma + \gamma^2 \cdot \beta}.\tag{7}$$

Proof: All proofs are in the Appendix.

Lemma 1 shows that in the dedicated capacity scenario the firm's optimization problem is separable not only cross-sectionally across the two divisions, but also intertemporally.¹⁴ The non-negativity constraints for new investments, $b_t \ge 0$, will not bind provided the corresponding sequence of capacity levels $k = (k_1, k_2, ...)$ satisfy the monotonicity requirement $k_{t+1} \ge k_t$ for all t. This latter condition will be met whenever the expected marginal revenues satisfy the monotonicity condition in (2).

Lemma 1 identifies c as the effective long-run marginal cost of capacity. To provide intuition for this characterization, Arrow (1964) and Rogerson (2008) note that the firm can increase its capacity at date t by one unit without affecting its capacity levels in subsequent periods through the following "reshuffling" of future capacity acquisitions: buy one more unit of capacity at date t-1, buy β unit less in period t, buy β^2 more unit in period t+1,

¹⁴The statement in Lemma 1 assumes implicitly that $\bar{k}_{it}^o > 0$. A sufficient condition for this to hold is the following boundary condition: $R'_i(0, \theta_{it}, \epsilon_{it}) > c$ for all $\theta_{it}, \epsilon_{it}$.

and so on. The cost of this variation, evaluated in terms of its present value as of date t-1, is given by:

$$v \cdot \left[1 - \gamma \cdot \beta + \gamma^2 \cdot \beta^2 - \gamma^3 \cdot \beta^3 + \gamma^4 \cdot \beta^4 \dots\right] = v \cdot \frac{1}{1 + \gamma \cdot \beta},$$

and therefore the present value of the variation at date t is:

$$(1+r) \cdot v \cdot \frac{1}{1+\gamma \cdot \beta} \equiv c.$$

Hence, c is the marginal cost of one unit of capacity made available for one period of time.¹⁵ It is useful to note that c is exactly the price that a hypothetical supplier would charge for renting out capacity for one period of time, if the rental business is constrained to make zero economic profit. Accordingly, we will also refer to c as the competitive rental price of capacity.

In the context of a single division, Rogerson (2008) has identified depreciation rules that result in goal congruence with regard to a sequence of overlapping investment projects. Rogerson shows that the depreciation schedule can be set in such a manner that the historical cost charge (the sum of depreciation and imputed interest charges) for one unit of capacity in each period is precisely equal to c, the marginal cost of capacity derived in Lemma 1. Let $z_{t-1,t}$ denote the historical cost charge in period t per dollar of capacity investment undertaken at date t - 1. It consists of the first-period depreciation percentage d and the capital charge r applied to the initial expenditure required for one unit of capacity. Thus:

$$z_{t-1,t} = v \cdot (d+r).$$

Accordingly, $z_{t-2,t}$ denotes the cost charge in period t per dollar of capacity investment undertaken at date t-2, and:

$$z_{t-2,t} = v \cdot [(1-d) + r \cdot (1-d)].$$

The total historical cost charge to Division 1' residual income measure in period t then becomes

¹⁵We recall that for an additional unit of capacity to be available at date t the corresponding investment expenditure, v, is incurred at date t-1. As one would expect, its compounded value $(1+r) \cdot v$, exceeds the date t marginal cost of one unit of capacity, i.e., c. As shown in Section 5 below, the gap between $(1+r) \cdot v$ and c widens as the useful life of investment increases from 2 to T periods.

$$z_t \equiv z_{t-1,t} \cdot b_{t-1} + z_{t-2,t} \cdot b_{t-2}$$

Division 1 will internalize a unit cost of capacity equal to the firm's marginal cost c, provided

$$z_t = c \cdot (b_{t-1} + \beta \cdot b_{t-2}) = c \cdot k_t.$$

Straightforward algebra shows that there is a unique depreciation percentage d that achieves the desired intertemporal cost allocation of investment expenditures. This value of d is given by:

$$d = \frac{1}{\gamma + \gamma^2 \cdot \beta} - r. \tag{8}$$

We note that 0 < d < 1 and:

$$(z_{t-2,t}, z_{t-1,t}) = \left(\frac{\beta \cdot v}{\gamma + \gamma^2 \cdot \beta}, \frac{v}{\gamma + \gamma^2 \cdot \beta}\right) = (\beta \cdot c, c).$$
(9)

Thus the historical cost charge per unit of capacity is indeed c in each period. The above intertemporal cost charges have been referred to as the *Relative Practical Capacity Rule* since the expenditure required to acquire one unit of capacity is apportioned over the next two periods in proportion to the capacity created for that period, relative to the total discounted capacity levels.¹⁶ We note in passing that the depreciation schedule corresponding to the Relative Practical Capacity rule will coincide with straight line depreciation exactly when $\beta = \frac{1+r}{1+2r}$. For instance, if r = .1, the Relative Practical Capacity Rule amounts to straight line depreciation if the practical capacity in the second period declines to 91%.

To summarize, if the firm depreciates investments according to the Relative Practical Capacity rule and the transfer price charged to Division 2 is based on full cost (which

¹⁶The term *Relative Practical Capacity Rule* has been coined in Rajan and Reichelstein (2008), while Rogerson (2008) refers to the Relative Replacement Cost rule to reflect that in his model the cost of new investments falls over time. It should be noted that under the Relative Practical Capacity rule the depreciation charges are not based on the relative magnitude of expected future cash inflows resulting from an investment. The link to expected future cash flows is a crucial ingredient in the *Relative Benefit Allocation Rule* proposed by Rogerson (1997), the *economic* depreciation rule proposed by Hotelling (1925) and the *neutral* depreciation rule advocated by Beaver and Dukes (1974). As demonstrated in Rajan and Reichelstein (2008) these depreciation rules are generally different, though they coincide in certain special cases, most notably if all investments have zero-NPV.

includes the imputed interest charges), then both divisions will be charged the competitive rental price c per unit of capacity in each period. The key difference between the two divisions is that while the downstream division can rent capacity on as needed basis, capacity investments entail a multiperiod commitment for the upstream division. In making its capacity investment decision in the current period, the upstream division has to take into account the resulting historical cost charges that will be deducted from its performance measures in future periods. Given the weights u_t that the divisions attach to their periodic performance measures, we then obtain a multi-stage game in which each division makes one move in each period; i.e., each division chooses its capacity level. We say that a performance measurement system attains strong goal congruence if the resulting game has a subgame perfect equilibrium in which the divisions choose first-best capacity levels in each period.

Proposition 1 When capacity is dedicated, full cost transfer pricing achieves strong goal congruence.

As demonstrated in the proof of Proposition 1, the divisional managers face a T-period game which has a unique subgame perfect equilibrium.¹⁷ Irrespective of past decisions, the downstream division will secure a capacity level that is optimal in the short-term, relative to the unit cost c. The upstream division potentially faces the constraint that, in any given period, it may inherit more capacity from past investment decisions than it currently needs. However, provided the divisions' marginal revenues are increasing over time; i.e., condition (2) is met, the requirement of sequential optimality embodied in the concept of subgame perfection implies that, in equilibrium, the upstream division will not find itself in a position with excess capacity.¹⁸

The result in Proposition 1 makes a strong case for full-cost transfer pricing, i.e., a transfer price that comprises variable production costs (effectively set to zero in our model) plus the allocated historical cost of capacity, c. Survey evidence indicates that in practice full cost

¹⁷The game has other, non-perfect Nash equilibria which result in inefficient capacity levels.

¹⁸We note that the result in Proposition 1 does not require the division managers to have symmetric information with regard to θ_{it} . It suffices for each manager to know his own θ_{it} , since the optimal capacity acquisitions are separable across the two divisions. With private information, the formal claim in Proposition 1 needs to be modified since there are no proper subgames. However, the concept of subgame perfect equilibrium could be replaced by another concept requiring sequential rationality, such as Bayesian Perfect equilibrium.

is the most prevalent approach to setting internal prices. In our model, the full cost rule leaves the upstream division with zero economic (residual) profit on internal transactions and, at the same time, provides a goal congruent valuation for the downstream division in its demand for capacity.

As argued by Balakrishnan and Sivaramakrishnan (2002), Goex (2002) and others, it has been difficult for the academic accounting literature to justify the concept of full-cost transfers. Most existing models have focused on one-period settings in which capacity costs were taken as fixed and exogenous. As a consequence, full-cost mechanisms typically run into the problem of double marginalization; that is, the buying entity internalizes a unit cost that exceeds the marginal cost to the firm. Some authors, including Zimmerman (1979), have suggested that fixed cost charges are effective proxies for opportunity costs arising from capacity constraints. In the context of a single capacity investment, however, there is no reason to believe that historical fixed cost charges relate systematically to current opportunity costs.¹⁹

Our rationale for the use of full cost transfer prices hinges crucially on the dynamic of overlapping capacity investments. Since the firm expects to operate at capacity, divisional managers should internalize the incremental cost of capacity; i.e., the unit cost c. The Relative Practical Capacity depreciation rule ensures that the unit cost of both incumbent and new capacity is valued at c in each period. As a consequence, the historical fixed cost charges can be "unitized" without running into a double marginalization problem with regard to the acquisition of new capacity.²⁰

We now relax the specification of dedicated capacity which assumes that capacity usage for each division must be decided at the beginning of the period. A plausible alternative scenario is that the demand shocks ϵ_t are realized relatively early during the period and the

¹⁹Banker and Hughes (1994) examine the relationship between support activity costs and optimal output prices in a classic one period news-vendor setting. Capacity is not a committed resource in their setting, since it is chosen *after* the output price has been decided. Consequently, they find that the marginal cost of capacity is relevant for the subsequent pricing decision. It should be noted that the primary focus of Banker and Hughes (1994) is not on whether full cost are a relevant input in the firm's pricing decision. Instead, they model multiple support activities and show that an activity-based measure of unit cost provides *economically* sufficient information for pricing decisions.

²⁰In contrast, fixed cost charges must be imposed as lump-sum charges in Wei (2000) in order to satisfy two simultaneous objectives: (i) provide divisional managers with an upfront incentive to reduce variable cost in future periods and (ii) ensure efficient internal transfers of goods and services in future periods.

production processes of the two divisions have enough commonalities so that the divisional capacity uses remain *fungible*. While the total capacity, k_t , is determined at the beginning of period t, this resource can be reallocated following the realization of the random shocks ϵ_t . To that end, we assume that the two divisions are free to negotiate an outcome that maximizes the total revenue available, $\sum_{i=1}^{2} R_i(q_{it}, \theta_{it}, \epsilon_{1t})$, subject to the capacity constraint $q_{1t} + q_{2t} \leq k_t$. Provided the optimal quantities $q_i^*(k_t, \theta_t, \epsilon_t)$ are positive, they will satisfy the first-order condition:

$$R'_{1}(q_{1t}^{*}, \theta_{1t}, \epsilon_{1t}) = R'_{2}(k_{t} - q_{1t}^{*}, \theta_{2t}, \epsilon_{2t}).$$
(10)

We also define the shadow price of capacity in period t, given available capacity, k_t , as:

$$S(k_t, \theta_t, \epsilon_t) \equiv R'_i(q_i^*(k_t, \theta_t, \epsilon_t), \theta_{it}, \epsilon_{it}), \qquad (11)$$

provided $q_i^*(k_t, \theta_t, \epsilon_t) > 0$. Thus, the shadow price is the marginal revenue that the divisions could collectively obtain from an additional unit of capacity acquired at the beginning of the period. Clearly, $S(\cdot)$ is increasing in both θ_t and ϵ_t , but decreasing in k_t .

The net present value of the firm's expected future cash flows is given by:

$$\Pi_f(b) = \sum_{t=1}^{\infty} E_{\epsilon} \left[M_f(b_{t-1} + \beta \cdot b_{t-2}, \theta_t, \epsilon_t) - v \cdot b_t \right] \cdot \gamma^t,$$

where the maximized contribution margin now takes the form:

$$M_f(k_t, \theta_t, \epsilon_t) = R_1(q_1^*(k_t, \theta_t, \epsilon_t), \theta_{1t}, \epsilon_{1t}) + R_2(k_t - q_1^*(k_t, \theta_t, \epsilon_t), \theta_{2t}, \epsilon_{2t}).$$

Using the Envelope Theorem, we obtain the following analogue of Lemma 1.

Lemma 2 When capacity is fungible, the optimal capacity levels, k_t^o , are given by:

$$E_{\epsilon}\left[S(k_t^o, \theta_t, \epsilon_t)\right] = c. \tag{12}$$

We note that with dedicated capacity the optimal \bar{k}_{it}^o for each division depends only on θ_{it} . With fungible capacity, in contrast, the optimal aggregate k_t^o depends on both θ_{1t} and θ_{2t} . The proof of Lemma 2 shows that, for any given capacity level k, the expected shadow prices are increasing over time. As a consequence, the first-best capacity levels given by (12) are also increasing over time, which in turn implies that the non-negativity constraints $b_t \geq 0$ again do not bind.

Since the relevant information embodied in the shocks ϵ_t is assumed to be known only to the divisional managers, and they are assumed to have symmetric information about the attainable net-revenues, the two divisions can split the "trading surplus" of $M_f(k_t, \theta_t, \epsilon_t) - \sum_{i=1}^2 R_i(k_{it}, \theta_{it}, \epsilon_{it})$ between them. Let $\delta \in [0, 1]$ denote the fraction of the total surplus that accrues to Division 1. Thus, the parameter δ measures the relative bargaining power of Division 1, with the case of $\delta = \frac{1}{2}$ corresponding to the familiar Nash bargaining outcome. The negotiated adjustment in the transfer payment, ΔTP is then given by:

$$R_1(q_1^*(k_t,\theta_t,\epsilon_t),\theta_{1t},\epsilon_{1t}) + \Delta TP = R_1(k_{1t},\theta_{1t},\epsilon_{1t}) + \delta \cdot \left[M_f(k_t,\theta_t,\epsilon_t) - \sum_{i=1}^2 R_i(k_{it},\theta_{it},\epsilon_{it}) \right].$$

At the same time, Division 2 obtains:

$$R_{2}(k_{t}-q_{1}^{*}(k_{t},\theta_{t},\epsilon_{t}),\theta_{2t},\epsilon_{2t})-\Delta TP = R_{2}(k_{2t},\theta_{2t},\epsilon_{2t})+(1-\delta)\cdot\left[M_{f}(k_{t},\theta_{t},\epsilon_{t})-\sum_{i=1}^{2}R_{i}(k_{it},\theta_{it},\epsilon_{it})\right]$$

These payoffs ignore the transfer payment $c \cdot k_{2t}$ that Division 2 makes at the beginning of the period, as these payoffs are viewed as sunk at the negotiation stage. The total transfer payment made by Division 2 in return for the ex-post efficient quantity $q_2^*(k_t, \theta_t, \epsilon_t)$ is then given $c \cdot k_{2t} + \Delta TP$. Clearly, $\Delta TP > 0$ if and only if $q_2^*(k_t, \theta_t, \epsilon_t) > k_{2t}$. We refer to the resulting "hybrid" transfer pricing mechanism as *adjustable full cost transfer pricing*.

At first glance, the possibility of reallocating the initial capacity rights appears to be an effective mechanism for capturing the trading gains that arise from random fluctuations in the divisional revenues. However, the following result shows that the prospect of such negotiations compromises the divisions' long-term incentives.

Proposition 2 When capacity is fungible, adjustable full cost transfer pricing fails to achieve strong goal congruence.

The proof of Proposition 2 shows that, for some performance measure weights u_t , there is no (subgame perfect) equilibrium which results in efficient capacity investments. In particular, the proof identifies a *dynamic holdup problem* that results when the downstream division drives up its capacity demand opportunistically in an early period in order to acquire some of the resulting excess capacity in later periods through negotiation. Doing so is generally a cheaper for the downstream division than securing capacity upfront at the transfer price c. Such a strategy will be particularly profitable for the downstream division if the performance measure weights u_{2t} are such that the downstream division assigns more weight to the later periods.

It should be noted that the dynamic holdup problem can emerge only if the downstream division anticipates negotiation over actual capacity usage in subsequent periods. In the dedicated capacity scenario examined above, the downstream division could not possibly gain by driving up capacity strategically because it cannot appropriate any excess capacity through negotiation. The essence of the dynamic holdup problem is that the downstream division has the power to force long-term asset commitments without being accountable in the long-term. That power becomes detrimental if the downstream division anticipates future negotiations over actual capacity usage.

4 Centralized Capacity Ownership

One alternative to the divisional structure examined in the previous section is to centralize capacity ownership at the corporate level. In the context of our model, both divisions would then effectively become profit centers that can secure capacity rights from a central office on a period-by-period basis. The central office owns the assets and in each period acquires sufficient capacity so as to fulfill the divisional requests made at the beginning of that period. The downstream division will no longer be able to "hold-up" the upstream division, since the latter is no longer the residual claimant of capacity rights.

4.1 Cost-Based Transfer Pricing

We begin with cost-based transfer prices which allow the divisions to rent capacity from the central office at the full cost transfer price c. Since this price is the historical cost of capacity under the relative practical capacity depreciation rule, the central unit will show a residual income of zero in each period, provided the divisions do not "game" the system by forcing the central office to acquire excess capacity,

To investigate the divisional incentives, we first retain the sequential decision structure in which the downstream division communicates its capacity requirements to the upstream division, which then secures enough capacity rights from the central office to meet the needs of both divisions.²¹ Suppose the initial capacity choices leave Division i with k_{it} units of capacity in period t. After the two managers have observed the realization of the period tdemand shock ϵ_t , they will again divide the total capacity k_t so as to maximize the sum of revenues for the two divisions. The effective net-revenue to Division i then becomes:

$$R_{1}^{*}(k_{1t}, k_{2t}|\theta_{t}, \epsilon_{t}) = (1 - \delta) \cdot R_{1}(k_{1t}, \theta_{1t}, \epsilon_{1t}) + \delta \cdot [M_{f}(k_{t}, \theta_{t}, \epsilon_{t}) - R_{2}(k_{2t}, \theta_{2t}, \epsilon_{2t})].$$

and

$$R_{2}^{*}(k_{1t}, k_{2t}|\theta_{t}, \epsilon_{t}) = \delta \cdot R_{2}(k_{2t}, \theta_{2t}, \epsilon_{2t}) + (1 - \delta) \cdot \left[M_{f}(k_{t}, \theta_{t}, \epsilon_{t}) - R_{1}(k_{1t}, \theta_{1t}, \epsilon_{1t})\right].$$

Taking division 2's capacity request k_{2t} as given, manager 1 will choose k_{1t} to maximize:

$$E_{\epsilon} \left[R_1^*(k_{1t}, k_{2t} | \theta_t, \epsilon_t) \right] - c \cdot k_{1t}.$$

Anticipating Division 1's response $k_{1t}^*(k_{2t})$, Division 2 has the first-mover advantage and chooses k_{2t} to maximize:

$$E_{\epsilon_t} [R_2^*(k_{1t}^*(k_{2t}), k_{2t}|\theta_t, \epsilon_t)] - c \cdot k_{2t}.$$

It is useful to observe that in the extreme case where Division 1 extracts the entire negotiation surplus ($\delta = 1$), Division 1's objective simplifies to $E_{\epsilon}[M_f(k_t, \theta_t, \epsilon_t)] - c \cdot (k_t - k_{2t})$. As a consequence, Division 1 would fully internalize the firm's objective and choose the efficient capacity level k_t^o . Similarly, in the other corner case of $\delta = 0$, Division 2 would internalize the firm's objective and choose its demand k_{2t} such that Division 1 responds with the efficient capacity level k_t^o . For any $\delta \in (0, 1)$, Division 1's (expected) marginal revenue of acquiring another unit of capacity is:

$$E_{\epsilon}\left[\frac{\partial}{\partial k_{1t}}R_{1}^{*}(k_{1t},k_{2t}|\theta_{t},\epsilon_{t})\right] = E_{\epsilon}\left[(1-\delta)\cdot R_{1}^{'}(k_{1t},\theta_{1t},\epsilon_{1t}) + \delta\cdot S(k_{1t}+k_{2t},\theta_{t},\epsilon_{t})\right].$$
 (13)

It turns out that $k_{1t} = 0$ cannot be part of a Nash equilibrium (see the proof of Lemma 3). Consequently, the constraint $k_{1t} \ge 0$ will not bind in equilibrium and Division 1's optimal response k_{1t}^* satisfies the following first-order condition:

 $^{^{21}}$ We contrast alternative sequencing arrangements, i.e., sequential versus simultaneous capacity requests, at the end of this section.

$$E_{\epsilon_{t}}\left[(1-\delta)\cdot R_{1}'(k_{1t}^{*},\theta_{1t},\epsilon_{1t}) + \delta\cdot S(k_{1t}^{*}+k_{2t},\theta_{t},\epsilon_{t})\right] = c.$$
 (14)

Since $R'_1(\cdot)$ and $S(\cdot)$ are decreasing functions of k_{1t} , Division 1's objective function is globally concave and therefore there is a unique best response k^*_{1t} , given as the solution to equation (14).

Lemma 3 For $\delta \in (0,1)$, Division 1's response function, $k_{1t}^*(k_{2t})$, satisfies: $-1 < \frac{\partial k_{1t}^*}{\partial k_{2t}} < 0$.

Lemma 3 implies that the total capacity stock, $k_t = k_{1t}^*(k_{2t}) + k_{2t}$, chosen by the upstream division in period t satisfies $0 < \frac{\partial k_t}{\partial k_{2t}} < 1$. For each additional unit of capacity secured upfront by the downstream division, Division 1 will increase the total capacity by less than a unit. Division 2's optimal choice k_{2t}^* satisfies the first-order condition:

$$E_{\epsilon_t}\left[\delta \cdot R_2'(k_{2t}^*, \theta_{2t}, \epsilon_{2t}) + (1 - \delta) \cdot S(k_t^*, \theta_t, \epsilon_t) + \frac{\partial k_{1t}^*}{\partial k_{2t}} \left\{S(k_t^*, \theta_t, \epsilon_t) - c\right\}\right] = c, \quad (15)$$

where $k_t^* \equiv k_{2t}^* + k_{2t}^*(k_{1t}^*)$.

It is instructive to interpret the marginal revenues that each division obtains from securing capacity for itself at the beginning of period t. The second term on the left-hand side of both (14) and (15) represents the firm's aggregate and *optimized* marginal revenue, given by the (expected) shadow price of capacity. Since the divisions individually only receive a share of the aggregate return (given by δ and $1 - \delta$, respectively), this part of the investment return entails a "classical" holdup problem.²² Yet, the divisions also derive *autonomous* value from the capacity available to them, even if the overall capacity were not to be reallocated expost. The corresponding marginal revenues are given by the first terms on the left-hand side of equations (14) and (15), respectively. The overall incentives to acquire capacity therefore stem both from the unilateral "stand-alone" use of capacity as well as the prospect of trading capacity with the other division.²³ For the downstream division, the marginal

 $^{^{22}}$ Earlier papers on transfer pricing that have examined this hold-up effect include Edlin and Reichelstein (1995), Baldenius et al. (1999), Anctil and Dutta (1999), Wielenberg (2000) and Pfeiffer et al. (2007).

 $^{^{23}}$ A similar convex combination of investment returns arises in the analysis of Edlin and Reichelstein (1995), where the parties sign a fixed quantity contract to trade some good at a later date. While the initial contract will almost always be renegotiated, its significance is to provide the divisions with a status-quo return on their relationship-specific investments.

revenue identified in (15) includes a third term, which simply reflects its first-mover (i.e., Stackelberg leader) status.

The structure of the marginal revenues in (14) and (15) also highlights the importance of giving Division 2 the option of securing capacity rights. Without this option, the firm would face an underinvestment problem. To see this, note that if only Division 1 were to acquire capacity from the center, its marginal revenue at the efficient capacity level k_t^o would be:

$$E_{\epsilon}\left[(1-\delta)\cdot R_{1}'(k_{t}^{o},\theta_{1t},\epsilon_{1t})+\delta\cdot S(k_{t}^{o},\theta_{t},\epsilon_{t})\right].$$
(16)

Yet, this marginal revenue is less than c because:

$$E_{\epsilon}\left[S(k_t^o, \theta_t, \epsilon_t)\right] = E_{\epsilon}\left[R_1'(q_1^*(k_t^o, \theta, \epsilon_t), \theta_{1t}, \epsilon_{1t})\right] = c,$$
(17)

and $E_{\epsilon}\left[R'_{1}(q_{1}^{*}(k_{t}^{o},\theta,\epsilon_{t}),\theta_{1t},\epsilon_{1t})\right] > E_{\epsilon}\left[R'_{1}(k_{t}^{o},\theta,\epsilon_{t}),\theta_{1t},\epsilon_{1t})\right]$. Thus the upstream division would have insufficient incentives to secure the firm-wide optimal capacity level on its own. This observation speaks directly to our finding in Proposition 2. Although the dynamic hold-up problem of "strategic" excess capacity could be effectively addressed by prohibiting the downstream division from secure capacity rights on its own, such an approach would also induce the upstream division to underinvest as it would anticipate being held up on its investment in the ensuing negotiation.

We next characterize the efficient capacity level k_t^o in the fungible capacity scenario in relation to the efficient capacity level, $\bar{k}_t^o \equiv \bar{k}_{1t}^o + \bar{k}_{2t}^o$, that two stand-alone divisions should acquire in the dedicated capacity setting. To that end, it will be useful to make the following assumption regarding the divisional revenue functions:

Assumption (A1): The marginal revenue functions, $R'_i(\cdot, \epsilon_{it})$, are linear in ϵ_{it} .

A sufficient condition for this linearity condition to hold is that the revenue functions are multiplicatively separable; i.e., $R_i(q_{it}, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \hat{R}_i(q_{it}, \theta_{it})$.

Proposition 3 Given A1, the optimal capacity level k_t^o in period t satisfies:

$$k_t^o \begin{cases} \geq \bar{k}_t^o & \text{if } S(k_t, \theta_t, \epsilon_t) \text{ is convex in } \epsilon_t \\ \leq \bar{k}_t^o & \text{if } S(k_t, \theta_t, \epsilon_t) \text{ is concave in } \epsilon_t. \end{cases}$$
(18)

According to Proposition 3, the curvature of the shadow price determines whether a riskneutral central decision maker would effectively be risk-seeking or risk-averse with respect to the residual uncertainty associated with the stochastic shock ϵ_t . Relative to the benchmark setting of dedicated capacity, in which capacity reallocations are (by definition) impossible, a shadow price function, $S(\cdot)$, that is convex in ϵ_t makes the volatility inherent in ϵ_t more valuable to a risk-neutral decision maker. The central decision maker would therefore be willing to invest a larger amount in capacity. The reverse holds when the shadow price is concave. An immediate consequence of Proposition 3 is that when the shadow price is linear in ϵ_t , the optimal capacity level is precisely the same as in the scenario of dedicated capacity.²⁴ We point out in passing that assumption A1 does not imply the linearity of $S(\cdot)$ in ϵ_t , because ϵ_t enters $S(\cdot)$ not only directly but also via the ex-post efficient capacity allocation, $q_i^*(k_t, \theta_t, \epsilon_t)$.

The curvature of the shadow price functions hinges (unfortunately) on the third derivatives of the net-revenue functions. All three scenarios identified in Proposition 3 can arise for standard functional forms. For instance, it is readily checked that if $R_i(q, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot \sqrt{q}$, then the shadow price is a convex function of ϵ_t , and therefore $k_t^0 > \bar{k}_t^0$. On the other hand, $S(\cdot, \cdot, \epsilon_t)$ is concave when $R_i(q, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot (1 - e^{-q})$. Examples of revenue functions that yield linear shadow prices, and hence $k_t^o = \bar{k}_t^o$ for each t, include: (i) $R_i(q, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot \ln q$ and (ii) $R_i(q, \theta_{it}, \epsilon_{it}) = q \cdot [\epsilon_{it} \cdot \theta_{it} - h_i \cdot q]$. It should be noted that all of the above examples satisfy assumption A1, and yet $S(\cdot)$ is generally not a linear function of ϵ_t .

With centralized capacity ownership, the T-period game becomes intertemporally separable for the divisions since their moves in any given period have no payoff consequences in future periods. Given this intertemporal separability, any collection of Nash equilibria in the T "stage game" would also constitute a subgame perfect equilibrium for the T-period game. The following result identifies a necessary and sufficient condition for a stage game equilibrium to result in efficient capacity levels under adjustable full cost transfer pricing

²⁴In particular, when $S(\cdot)$ is linear in ϵ_t , we have $k_t^o(\theta_{1t}, \theta_{2t}) = \bar{k}_{1t}^o(\theta_{1t}) + \bar{k}_{2t}^o(\theta_{2t})$.

policy.

Lemma 4 Given A1, suppose capacity is fungible and capacity ownership is centralized. Adjustable full cost transfer pricing then achieves strong goal congruence if and only if the divisions respectively secure the following capacity levels:

$$k_{1t} = \bar{k}_{1t}^{o}$$
 and $k_{2t} = k_{t}^{o} - \bar{k}_{1t}^{o}$

Because of the sequential decision structure, goal congruence can be obtained only if the upstream division, in its role as the Stackelberg follower, secures precisely the same amount of capacity level as it would in the stand-alone scenario of dedicated capacity. In contrast, Division 2, as the Stackelberg leader, must have an incentive to secure the capacity level $k_{2t} = \bar{k}_{2t}^o + [k_t^o - \bar{k}_t^o]$; i.e., it must make up for any difference between k_t^o and \bar{k}_t^o .

When the shadow price $S(k_t, \theta_t, \epsilon_t)$ is linear in ϵ_t , it is readily seen that the stand-alone capacity levels $(\bar{k}_{1t}^o, \bar{k}_{2t}^o)$ are a solution to the divisional first-order conditions in (14) and (15). These choices are in fact the unique Nash equilibrium; i.e., $(\bar{k}_{1t}^o, \bar{k}_{2t}^o)$ is the unique maximizer of the divisional objective functions.

Proposition 4 Given A1, suppose capacity is fungible and capacity ownership is centralized. Adjustable full cost transfer pricing then achieves strong goal congruence provided the shadow price $S(k_t, \theta_t, \epsilon_t)$ is linear in ϵ_t .

Linearity of the shadow price $S(\cdot)$ in ϵ_t implies that the level of investment that is desirable from an ex-ante perspective is the same as in the dedicated capacity setting. This parity holds despite the fact that the expected profit of the integrated firm is higher than the sum of expected profits of two stand-alone divisions. From the divisional return perspective, the δ and $(1-\delta)$ expressions in (14) and (15) are exactly the same at the stand-alone capacity levels $(\bar{k}_{1t}^o \text{ and } \bar{k}_{2t}^o)$. We recall that the shadow price will indeed satisfy the linearity condition identified in Proposition 4 provided the divisional revenue functions can be described by a quadratic function of the form:

$$R_i(q,\theta_{it},\epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot q - h_{it} \cdot q^2$$
(19)

for some constants $h_{it} > 0.2^5$ The quadratic form in (19) might serve as a reasonable approximation of the "true" revenue functions. Although our model presumes that the functions $R_i(\cdot)$ are known *precisely*, it might be unrealistic to expect that managers have such detailed information in most real world contexts. To that end, a second-order polynomial approximation of the form in (19) might prove adequate. We conclude that an internal pricing system which allows the divisions to rent capacity at full, historical cost achieves effective coordination, subject to the qualification that the divisional revenues can be approximated "sufficiently well" by quadratic revenue functions.²⁶

Propositions 3 and 4 strongly suggest that if the shadow price function $S(\cdot)$ is not linear in ϵ_t , adjustable full cost transfer pricing will no longer result in efficient capacity investments because of a *coordination failure* in the divisional capacity requests. The following result characterizes the directional bias of the resulting capacity levels.

Proposition 5 Given A1, suppose capacity is fungible and capacity ownership is centralized. Adjustable full cost transfer pricing results in over-investment (under-investment) if the shadow price $S(k_t, \theta_t, \epsilon_t)$ is concave (convex) in ϵ_t .

Transfers at cost lead each division to properly internalize the incremental cost of capacity. However, as noted above, the divisional investment incentives are essentially a convex combination of two forces: the benefits of capacity that a division receives on its own and the optimized revenue that the two divisions can attain jointly by reallocating capacity. When $k_t^o > \bar{k}_t^o$ (because the shadow price is convex in ϵ_t), the efficient capacity level k_t^o cannot emerge in equilibrium. The marginal benefit that the downstream division receives from the autonomous use of its capacity is not sufficiently large at the quantity $k_t^o - \bar{k}_{1t}^o$ since:

$$E_{\epsilon}\left[R_{2}^{'}(k_{t}^{o}-\bar{k}_{1t}^{o},\theta_{2t},\epsilon_{2t})\right] < E_{\epsilon}\left[R_{2}^{'}(\bar{k}_{2t}^{o},\theta_{2t},\epsilon_{2t})\right] = c$$

Yet as shown in Lemma 4, it must be the downstream division, in its role as Stackelberg leader, that must have an incentive to secure the additional capacity that is desired with a convex shadow price.

²⁵While linearity of the shadow price $S(\cdot)$ in ϵ_t is certainly a non-generic case, we note that functional forms other than a quadratic yield the same conclusion; e. g., $R_i(q, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot \ln q$.

²⁶This result will also hold if each division has private information regarding θ_{it} . The choice of $\bar{k}_{it}^o(\theta_{it})$ then forms a Bayesian equilibrium in each period, provided the parties anticipate to negotiate the ultimate capacity usage with symmetric and complete information, i.e., (θ_t, ϵ_t) will be known to both parties.

To counteract the bias identified in Proposition 5, one would expect that an appropriately chosen mark-up (or discount) to the unit cost of capacity c could restore goal congruence. To that end, suppose that the central capacity provider charges the divisions a transfer price of $TP = (1 + m) \cdot c$. We interpret m > 0 as a mark-up, while m < 0 would constitute a discount from full cost.

Corollary 1 Given A1, suppose the shadow price $S(k_t, \theta_t, \epsilon_t)$ is concave (convex) in ϵ_t . If capacity is fungible and capacity ownership is centralized, adjustable cost-based transfer pricing can achieve goal congruence only if the transfer price includes a mark-up on (discount from) the full cost of capacity.

Transfer pricing surveys indicate that cost-plus transfer prices are widely used in practice. Some authors have suggested that this policy reflects fairness considerations in the sense that both profit centers should view a transaction as profitable (Eccles,1985, and Eccles and White, 1988). In contrast, our result here points to mark-ups as an essential tool for correcting the bias resulting from the fact that neither division fully internalizes the externality associated with uncertain returns from capacity investments. From the perspective of the firm's central office, a major obstacle of course, is that the optimal mark-up depends on the information variables, θ_t , which reside with the divisional managers.

To conclude this subsection, we demonstrate that our findings in this section are not sensitive to the sequential decision structure in which the downstream division is given a first-mover advantage. This sequence was natural when Division 1 was an investment center with ultimate responsibility for acquiring and maintaining new capacity. Suppose now that the central office instructs the two divisions to announce their desired capacity levels simultaneously. Subsequently, the central office installs the necessary amount of new capacity so as to meet the announced demands of the two divisions. For this alternative game form, the capacity investments will reflect the corresponding Cournot equilibrium, rather than the Stackelberg equilibrium identified in the preceding propositions.

Corollary 2 Proposition 4 and 5 continue to hold if the divisions secure their capacity rights simultaneously rather than sequentially.

The proof of this claim mirrors that of Propositions 4 and 5. The only substantive

difference is that the last term on the left-hand side of (15) disappears when the parties make simultaneous rather than sequential moves.

4.2 Negotiated Transfer Pricing

The conclusion of the previous subsection is that even if the firm implements a system of intertemporal cost charges that reflect the competitive rental prices of capacity, such a cost-based system will generally not lead to efficient coordination in the divisional capacity choices. In this section, we examine the possibility of improved coordination by appointing the upstream division a "gatekeeper" who must approve capacity rights secured by the other division.

Suppose that instead of having the right to secure capacity unilaterally for the current period, the downstream division can now only do so through a mutually acceptable negotiation with the upstream division. If the two divisions reach an upfront agreement, it specifies Division 2's capacity rights k_{2t} and a corresponding transfer payment TP_t that it must make to Division 1 for obtaining these rights. The parties report the outcome of this agreement (k_{2t}, TP_t) to the central office, which commits to honor it as the status quo point in any subsequent renegotiations. Division 1 then secures enough capacity from the central office to meet its own capacity needs as well as fulfill its obligation to the downstream division. As before, Division 1 is charged the historical full cost of capacity under the Relative Practical Capacity depreciation rule (i.e., c) for each unit of capacity that it acquires from the central owner.

If the parties fail to reach a mutually acceptable agreement, the downstream division would have no claim on capacity in that period. The upstream division would then secure capacity for its own use at the price of c per unit.²⁷

To summarize, we refer to the following decision rules as *negotiated transfer pricing*:

• Division 2 can secure capacity rights, k_{2t} , for period t through negotiation with Division 1. The parties report the outcome of this negotiation (k_{2t}, TP_t) to the central office.

 $^{^{27}}$ As argued in the previous subsection, in the absence of an upfront agreement, the upstream division would under-invest because it anticipates a hold-up on its investment.

- Without an upfront agreement, Division 2 would have no claim on capacity.
- Division 1 is in charge of securing capacity for both divisions. Division 1 is charged the full cost c for capacity.
- In response to new information, division managers are free to renegotiate Division 2's initial capacity assignment with a corresponding transfer payment.

Since ownership of the capacity assets is again centralized, the divisional capacity choice problems are again separable across time periods. Therefore, a system of negotiated transfer pricing will attain our strong goal congruence requirement if it induces the two divisions to acquire collectively the capacity level k_t^o in each period.

Proposition 6 Suppose capacity is fungible and capacity ownership is centralized. Negotiated transfer pricing then achieves strong goal congruence.

The proof of Proposition 6 demonstrates that in order to maximize their joint expected surplus, the divisions will agree on a particular amount of capacity level k_{2t} that the downstream can claim for itself in any subsequent renegotiation. Thereafter the upstream division has an incentive to acquire the optimal amount of capacity k_t^o for period t.²⁸ By taking away Division 2's unilateral right to rent capacity at some transfer price, the central office will generally make Division 2 worse off. We note, however, that this specification of the default point for the initial negotiation is of no importance for the efficiency of a negotiated capacity management system. The same capacity level, albeit with a different transfer payment, would result if the central office stipulated that in the absence of an agreement at the initial stage of period t, Division 2 could unilaterally rent capacity at some transfer price p_t (for instance, $p_t = c$).

Our finding that a two-stage negotiation allows the divisions to achieve an efficient outcome is broadly consistent with the results in Edlin and Reichelstein (1995) and Wielenberg (2000). The main difference is that in the present setting the divisions bargain over the downstream division's unilateral capacity rights. As observed above, the upstream division would acquire too little capacity from a firm-wide perspective, if the downstream division

 $^{^{28}\}mathrm{We}$ note that the separability condition A1 is of no importance in establishing the claim in proposition 6.

could not stake an initial capacity claim. On the other hand, Proposition 5 demonstrated that simply giving the downstream division the right to acquire capacity at the relevant cost, c, could result in either over- or under-investment. By appointing Division 1 a gatekeeper for Division 2's unilateral capacity claims, the firm effectively balances the divisional rights and responsibilities so as to obtain goal congruence.

To conclude this subsection we note that our analysis has focused on the upstream division as an effective gatekeeper essentially because this division was assumed to have unique technological expertise in installing and maintaining production capacity. Yet, the preceding analysis makes clear that our conclusion also applies to a more symmetric setting. Suppose each divisions individually relies on production capacity that is fungible. The firm would then achieve effective coordination by appointing one of the divisions as a gatekeeper: requests for capacity must be routed through the gatekeeper which authorizes all capacity acquisitions and charges the other division a transfer price that is mutually acceptable to both parties.

5 Extensions

This section seeks to demonstrate that the findings of this paper are robust to several variations of the base model examined in the preceding sections. As pointed out above, our specification that assets have a useful life of two periods was made entirely for reasons of notational convenience. Suppose now that assets have a useful life of n periods. For an investment undertaken at date t, the practical capacity available at date t+i, $1 \le i \le n$ is β_i , with $\beta_i \le 1$. Provided the capacity levels are (weakly) decreasing over time i.e., $\beta_i \ge \beta_{i+1}$, Rogerson (2008) has shown that the marginal cost of obtaining one unit of capacity for one period is given by:

$$c = \frac{v}{\sum_{i=1}^{n} \gamma^i \cdot \beta_i}.$$
(20)

This characterization of the competitive rental price of capacity obviously extends the formula given in Lemma 1, where it was assumed that n = 2. It is clear that all of our results in the previous sections would remain intact once the marginal cost in the two period setting is replaced with its *n*-period counterpart in (20). As one would expect, the marginal cost c in (20) is monotone decreasing in the useful life of the asset, n, and in each of the persistence parameters β_i .

In many industry contexts, it is plausible that the acquisition cost of new capacity assents changes over time, possibly because of technological progress. Accordingly, suppose that the cost of acquiring one unit of capacity at date t is $c \cdot \alpha^t$. A scenario of $\alpha \leq 1$ would reflect that it over time it becomes cheaper for the firm to replace its capacity assets. Rogerson (2008) shows that in this scenario of geometrically declining capacity costs the marginal cost of one unit of capacity available at date t is:

$$c_t = \frac{v \cdot \alpha^t}{\sum_{i=1}^n (\alpha \cdot \gamma)^i \cdot \beta_i}$$

As before, there exists a unique depreciation schedule and a corresponding historical cost charges which impute a total cost of c_t per unit of capacity made available at date t. Adopting the notation introduced in Section 3, let $z_{t-i,t}$ denote the historical cost charge at date t per dollar of capacity investment undertaken at date t-i. In order for the full cost of capacity (depreciation plus imputed interest) to be equal to c_t , the modified historical cost charges must satisfy:

$$z_{t-i,t} = c_t \cdot \beta_i.$$

Clearly, this characterization reduces to our earlier finding in (9), if $\alpha = 1$, n = 2 and $\beta_1 = 1$. For the case n = 2, it is instructive to consider the modified depreciation charge d, which now amounts to:

$$d = \frac{1}{\gamma + \gamma^2 \cdot \beta \cdot \alpha} - r.$$
⁽²¹⁾

Direct comparison with (8) shows that the impact of technological progress ($\alpha \leq 1$) is that capacity assets should now be written off in a more accelerated fashion, i.e., a higher depreciation percentage in the first period. This change reflects that assets become economically obsolete faster when it is cheaper to replace them with new assets in future periods. To summarize, our results extend seamlessly to a scenario where asset acquisition costs change over time provided the divisions are charged the full cost of capacity c_t , either directly via the transfer price, or, if the upstream division is an investment center, via the modified depreciation schedule in (21).

Our goal congruence framework has abstracted from managerial incentive problems related to moral hazard. One possible approach to incorporating actions that are personally costly to managers is to let the divisional cash flows in each period also be functions of unobservable managerial effort. In the context of a one-period model, Edlin and Reichelstein (1995) identify separability conditions which imply that goal congruent performance measures can also serve as the basis of optimal second-best contracts. Essential to this argument is that the capacity investments are not a source of additional contracting frictions and therefore the optimal second-best incentive scheme does not need to balance a tradeoff between productive efficiency and higher managerial compensation. In contrast, in the models of Christensen et al. (2002), Dutta and Reichelstein (2002) and Baldenius et al. (2007) the investment decisions are a source of informational rent for the manager, and as a consequence, optimality require a departure from the first-best investment levels. In these papers, the second- best policy entails lower investment levels which can be induced through a suitable increase in the "hurdle rate", that is, the capital charge rate applied to the book value of assets.

6 Conclusion

The acquisition and subsequent utilization of capacity poses challenging incentive and coordination problems for multidivisional firms. Our model has examined the incentive properties of two different responsibility center arrangements. A decentralized ownership structure views the supplier of capacity services (the upstream division) as an investment center that is responsible for the acquisition of new capacity assets. In contrast, the downstream division rents capacity at a transfer price based on the historical cost of capacity, which includes depreciation and imputed interest charges. A suitable depreciation schedule for capacity investment expenditures ensures that both divisions internalize the firm's marginal cost of capacity. Transfer prices set at the full historical cost of capacity then lead the divisional managers to choose capacity levels that are efficient from the firm-wide perspective, provided capacity is dedicated, that is, the divisional capacity assignments are fixed in the short run. If the production processes of the two divisions have enough commonalities so that capacity becomes fungible in the short run, it is essential to give divisional managers discretion to negotiate a reallocation of the aggregate capacity available. This flexibility allows the firm to optimize the usage of aggregate capacity in response to fluctuations in the divisional revenues. Yet our analysis shows that the corresponding system of adjustable full cost transfer pricing can cause a dynamic hold-up problem in which the downstream division drives up its capacity demands opportunistically. It does so in anticipation of obtaining the corresponding excess capacity at a lower cost through negotiations in future periods.

The dynamic hold-up problem can be addressed by centralizing capacity ownership. Both divisions are then structured as profit centers that secure capacity from a central office on a period-by-period basis. However, even when these cost based transfer prices reflect the competitive rental price of capacity, the firm generally faces a remaining coordination problem. Making their capacity choices individually, the divisional managers generally do not internalize the firm-wide benefits of additional capacity acquisitions. A negotiated capacity management in which the downstream division no longer secures capacity rights unilaterally but through negotiations with the upstream division can solve this coordination problem. The upstream division then becomes effectively a gatekeeper for all new capacity investments. One implication of our analysis is that even if neither division has a technological advantage in providing capacity services, there would be coordination advantages to assigning the ultimate authority for new capacity acquisitions to a single gatekeeper.

Our analysis leaves open the possibility of other organizational solutions to the dynamic holdup and coordination problems. For instance, since the essence of the dynamic holdup problem is that the downstream division is not held accountable for the long-term effects of irreversible capacity investments, a natural response would be to structure both divisions as investment centers, even though the physical control over capacity assets rests with the upstream division. Rather than renting capacity on a period-by-period basis, the downstream division would now be required to carry "its share" of past capacity acquisitions on its balance sheet with a corresponding set of depreciation and capital charges. It would be instructive to assess the efficiency of such a responsibility center structure in future research.

Appendix

Proof of Lemma 1 We first show that for any sequence of capacity investments $b = (b_0, b_1, b_2...)$, with $b_t \ge 0$:

$$\sum_{t=0}^{\infty} v \cdot b_t \cdot \gamma^t = \sum_{t=1}^{\infty} c \cdot k_t \cdot \gamma^t$$

where $k_t = b_{t-1} + \beta \cdot b_{t-2}$. Since $k_0 = 0$, we can write

$$\sum_{t=0}^{\infty} v \cdot b_t \cdot \gamma^t = v \cdot [k_1 + \gamma[k_2 - \beta \cdot k_1] + \gamma^2[k_3 - \beta[k_2 - \beta \cdot k_1] + \gamma^3[k_4 - \beta[k_3 - \beta \cdot [k_2 - \beta \cdot k_1]]] + \dots$$

This expression is linear in each k_i and the coefficient on k_1 is:

$$v \left[1 - \gamma\beta + \gamma^2 \cdot \beta^2 - \gamma^3 \cdot \beta^3 + \gamma^4 \cdot \beta^4 \dots \right]$$
$$= v \left[\sum_{i=0}^{\infty} (\gamma \cdot \beta)^{2i} - \sum_{i=0}^{\infty} (\gamma \cdot \beta)^{2i+1} \right]$$
$$= v \cdot \sum_{i=0}^{\infty} (\gamma \cdot \beta)^{2i} [1 - \gamma \cdot \beta]$$
$$= v \cdot \frac{1}{1 + \gamma \cdot \beta}.$$

Similarly, the coefficient on k_t is

$$v \cdot \frac{1}{1 + \gamma \cdot \beta} \cdot \gamma^{t-1} = c \cdot \gamma^t.$$

In terms of future capacity levels, the firm's discounted future cash outflows can therefore be expressed as:

$$\sum_{t=1}^{\infty} E_{\epsilon_t} \left[M_d(\theta_t, \epsilon_t, k_t) - c \cdot k_t \right] \cdot \gamma^t.$$

This problem is intertemporally separable and the optimal \bar{k}_t^o are given by $\bar{k}_t^o = \bar{k}_{1t}^o + \bar{k}_{2t}^o$, where \bar{k}_{it}^o satisfies the first order conditions:

$$E_{\epsilon_i}[R'_i(\bar{k}^o_{it},\theta_{it},\epsilon_{it})] = c$$

The monotonicity condition in (2) ensures that the optimal \bar{k}_{it}^o are weakly increasing over time. Therefore the non-negativity constraints $b_t \ge 0$ do not bind. \Box

Proof of Proposition 1:

Using backward induction, consider the decision made by the downstream division in the last period. Independent of the current capacity stock and past decisions, its objective is to maximize:

$$E_{\epsilon_2}[R_2(k_{2T}, \theta_{2T}, \epsilon_{2T})] - c \cdot k_{2T}.$$
(22)

Let $\bar{k}_{2T}^o(\theta_{2T})$ denote the maximizer of (22). Division 1 faces the constrained optimization problem:

$$E_{\epsilon_1}[R_1(k_{1T}, \theta_{1T}, \epsilon_{1T})] - c \cdot k_{1T}$$

$$\tag{23}$$

subject to:

$$k_{1T} + \bar{k}_{2T}^o(\theta_{2T}) \ge \beta \cdot b_{T-2}.$$

Since Division 1's objective function in (23) is concave, it follows that the optimal capacity level installed at date T - 1 is:

$$k_T^* = \max\{\bar{k}_{1T}^o(\theta_{1T}) + \bar{k}_{2T}^o(\theta_{2T}), \beta \cdot b_{T-2}\},\$$

where $\bar{k}_{1T}^o(\theta_{1T})$ is the unconstrained maximizer of (23). In particular, the upstream division would invest $b_T = 0$ if $k_T^* = \beta \cdot b_{T-2}$.

In a subgame perfect equilibrium, Division 2 must select its capacity choice in period T-1 according to $\bar{k}^o_{2,T-1}(\theta_{2,T-1})$, irrespective of past decisions. In response, Division 1 will install a capacity level:

$$k_{T-1}^* = \max\{\bar{k}_{1,T-1}^o(\theta_{1,T-1}) + \bar{k}_{2,T-1}^o(\theta_{2,T-1}), \beta \cdot b_{T-3}\}.$$

Proceeding inductively, we conclude that in any period, the downstream division will rent the myopically optimal quantity $\bar{k}_{2t}^o(\theta_{2t})$. In response, the upstream division cannot do better than to select the capacity level k_t^* in period t. The assumption that marginal revenues are increasing for each division ensures that

$$\bar{k}_{1,t+1}^{o}(\theta_{1,t+1}) + \bar{k}_{2,t+1}^{o}(\theta_{2,t+1}) \ge \bar{k}_{1t}^{o}(\theta_{1t}) + \bar{k}_{2t}^{o}(\theta_{2t}).$$

As a consequence, the non-negativity constraint on new investments will not bind and:

$$k_t^* = \bar{k}_{1t}^o(\theta_{1t}) + \bar{k}_{2t}^o(\theta_{2t}) \equiv \bar{k}_t^o.$$

Proof of Proposition 2:

Let T = 2 and suppose that the random shocks ϵ_t assume their expected value $\bar{\epsilon}$ for sure. Furthermore, suppose that the divisional revenue functions are identical both crosssectionally and intertemporally; that is, $\theta_{it} = \theta$ for each $i \in \{1, 2\}$ and each $t \in \{1, 2\}$. Let $k \equiv \bar{k}_t^o$ denote the efficient capacity level for each division. Thus, $R'_i(k, \theta, \bar{\epsilon}) = c$. For simplicity, we also set the decay factor β equal to 1. Absent any growth in revenues and absent any decay in capacity, the optimal investment levels are $b_0 = 2 \cdot k$ and $b_1 = 0$, respectively. To show that there is no sub-game perfect equilibrium which results in efficient capacity investments for some weights u_t , suppose the downstream division has the entire bargaining power at the negotiation stage; that is, $\delta = 0.2^{9}$

Step 1: For any $b_0 \ge 2 \cdot k$, Division 2 will secure $k_{22} = 0$ in the second period and Division will set $b_1 = 0$.

At the beginning of the second period, Division 1 will choose b_1 so as to maximize:

$$R_1(b_0 - k_{22} + b_1, \theta, \bar{\epsilon}) - c \cdot b_1$$

subject to the constraints $b_1 \ge 0$ and $b_0 - k_{22} + b_1 \ge 0$. We note that the charges corresponding to b_0 are sunk costs. Division 2's second period capacity demand induces the following optimal response from Division 1:

$$b_1(k_{22}, b_0) = \begin{cases} 0 & \text{if } k_{22} \le b_0 - k \\ k_{22} + k - b_0 & \text{if } k_{22} \ge b_0 - k. \end{cases}$$
(24)

Anticipating this response, Division 2' second-period profit is given by:

$$\Gamma(k_{22},\theta,\bar{\epsilon}) = M_f(b_0,\theta,\bar{\epsilon}) - R_1(b_0 - k_{22},\theta,\bar{\epsilon}) - c \cdot k_{22}$$

for any $k_{22} \leq b_0 - k$. Thus, we find that $\Gamma'(k_{22}, \theta, \bar{\epsilon}) = R'_1(b_0 - k_{22}, \theta, \bar{\epsilon}) - c \leq R'_1(k, \theta, \bar{\epsilon}) - c < 0$ for all $k_{22} \leq b_0 - k$.

For any $k_{22} \ge b_0 - k$ the downstream division's payoff is:

 $^{^{29}}$ This specification does simplify the algebra considerably, yet as will become clear below, is in no way essential for the following argument.

$$\Gamma(k_{22},\theta,\bar{\epsilon}) = M_f(k+k_{22},\theta,\bar{\epsilon}) - R_1(k,\theta,\bar{\epsilon}) - c \cdot k_{22}.$$

Since by definition $S(k_{22} + k, \theta, \bar{\epsilon}) \leq c$, it follows that $\Gamma'(k_{22}, \theta, \bar{\epsilon}) < S(k_{22} + k, \theta, \bar{\epsilon}) - c < 0$. This completes the proof of Step 1.

Step 2: For any $b_0 \ge 2 \cdot k$, Division 2's second period profit is increasing in b_0 From Step 1 we know that neither division will obtain additional capacity rights if $b_0 \ge 2 \cdot k$. As a consequence, Division 2's profit becomes

$$\Gamma(0, b_0, \theta, \bar{\epsilon}) = M_f(b_0, \theta, \bar{\epsilon}) - R_1(b_0, \theta, \bar{\epsilon}).$$

That expression is increasing in b_0 because

$$\frac{\partial}{\partial b_0} \Gamma(0, b_0, \theta, \bar{\epsilon}) = R_1'(q_1^*(b_0, \theta, \bar{\epsilon}), \theta, \bar{\epsilon}) - R_1'(b_0, \theta, \bar{\epsilon}) > 0$$

as $q_2^*(b_0, \theta, \bar{\epsilon}) > 0$. We conclude that Division 2 has an incentive to force Division 1 to acquire excess capacity in the first period, that is, to drive b_0 beyond the efficient level $2 \cdot k$. Division 2 can do so unilaterally by increasing k_{21} . Doing so is, of course, costly in period 1. Yet, it will be an optimal strategy for the downstream division provided the performance measure weights are such that u_{21} is sufficiently small relative to u_{22} .

Proof of Lemma 2:

Step 1: For a given capacity level k, the expected shadow price of capacity is increasing over time, that is:

$$E_{\epsilon}[S(k,\theta_{t+1},\epsilon_{t+1})] \ge E_{\epsilon}[S(k,\theta_t,\epsilon_t)].$$
(25)

For any fixed pair (k, ϵ) , we claim that:

$$S(k, \theta_{t+1}, \epsilon) \ge S(k, \theta_t, \epsilon) \tag{26}$$

Suppose first $0 < q_1^*(k, \theta_{t+1}, \epsilon) \le q_1^*(k, \theta_t, \epsilon) < k$. Since $R'_1(q, \theta_{1t}, \epsilon_{1t})$ is increasing in θ_{1t} and $\theta_{1,t+1} \ge \theta_{1t}$, the definition of the shadow price in (11) implies the inequality in (26). Suppose now $q_1^*(k, \theta_{t+1}, \epsilon) \ge q_1^*(k, \theta_t, \epsilon)$. Since the shadow price can be expressed as:

$$S(k, \theta_t, \epsilon) = R'_2(k - q_1^*(k, \theta_t, \epsilon), \theta_{2t}, \epsilon_2),$$

 $\theta_{2,t+1} \ge \theta_{2t}$ and $R'_2(q, \theta_{2t}, \epsilon_{2t})$ is increasing in θ_{2t} , we conclude that (26) holds. The claim now follows because ϵ_t and ϵ_{t+1} are *iid*.

If $q_i^*(k, \theta_t, \epsilon) = 0$, a similar argument can be made, keeping in mind that $S(k, \theta_t, \epsilon) = R'_j(k, \theta_{jt}, \epsilon_j)$ if $q_i^*(k, \theta_t, \epsilon) = 0$.

Step 2: Proceeding exactly as in the proof of Lemma 1, the firm's expected future cash flows are:

$$\sum_{t=1}^{\infty} E_{\epsilon_t} \left[M_f(k_t, \theta_t, \epsilon_t) - c \cdot k_t \right] \cdot \gamma^t.$$

This problem is intertemporally separable and the optimal k_t^o satisfy the first order conditions:

$$E_{\epsilon_t}\left[\frac{\partial}{\partial k_t}M_f(k_t^o,\theta_t,\epsilon_t)\right] = R'_i(q_i^*(k_t^o,\theta_t,\epsilon),\theta_{it},\epsilon_{it}) = c_i$$

provided $q_i^*(k_t^o, \theta_t, \epsilon) > 0$. By definition,

$$R'_i(q^*_i(k^o_t, \theta_t, \epsilon), \theta_{it}, \epsilon_{it}) = E_{\epsilon_t}[S_t(k^o_t, \theta_t, \epsilon_t)] = c.$$

The claim therefore follows after observing that, by Step 1, the optimal capacity levels, k_t^o are increasing over time and, as a consequence, the non-negativity constraints $b_t \ge 0$ will not be binding.

Proof of Lemma 3:

We first show that if k_{1t}^* is interior, then $-1 < \frac{\partial k_{1t}^*}{\partial k_{2t}} < 0$. Implicitly differentiating (13) with respect to k_{2t} yields

$$\frac{\partial k_{1t}^*}{\partial k_{2t}} = \frac{-E_{\epsilon_t}[\delta \cdot S'(k_t^*, \theta_t, \epsilon_t)]}{E_{\epsilon_t}\left[\delta \cdot S'(k_t^*, \theta_t, \epsilon_t) + (1 - \delta) \cdot R_1''(k_{1t}^*, \theta_{1t}, \epsilon_{1t})\right]}$$

where $S'(k_t^*, \theta_t, \epsilon_t) \equiv \frac{\partial S(k_t^*, \theta_t, \epsilon_t)}{\partial k_t}$. The result then follows because $S'(k_t^*, \theta_t, \epsilon_t)$ and $R''_1(k_{1t}^*, \theta_{1t}, \epsilon_{1t})$ are both negative and $\delta \in (0, 1)$.

We now establish that $k_{1t} = 0$ cannot be part of an equilibrium. To the contrary, suppose there exists a k_{2t} such that the pair $(0, k_{2t})$ is an equilibrium. Since $k_{1t}^* = 0$ is Division 1's optimal response to k_{2t} , we must have:

$$E_{\epsilon}\left[(1-\delta)\cdot R_1'(0,\theta_{1t},\epsilon_{1t})+\delta\cdot S(k_{2t},\theta_t,\epsilon_t)\right] \le c.$$

Since $R'_1(0, \theta_{1t}, \epsilon_{1t}) < c$, the above inequality implies that:

$$E_{\epsilon}[S(k_{2t}, \theta_t, \epsilon_t)] < c, \tag{27}$$

which in turn implies that $k_{2t} > k_t^o$.

Let Γ denote Division 2's marginal revenue at $k_{1t} = 0$; i.e.,

$$\Gamma \equiv E_{\epsilon} \left[\delta \cdot R_2'(k_{2t}, \theta_{2t}, \epsilon_{2t}) + (1 - \delta) \cdot S(k_{2t}, \theta_t, \epsilon_t) + \frac{\partial k_{1t}^*}{\partial k_{2t}} \cdot \left[S(k_{2t}, \theta_t, \epsilon_t) - c \right] \right].$$

We note that the third term of the above expression is strictly less than $c - E_{\epsilon}[S(k_{2t}, \theta_t, \epsilon_t)]$ because of the inequality in (27) and the fact that $-1 < \frac{\partial k_{1t}^*}{\partial k_{2t}} \leq 0$. Consequently,

$$\Gamma < \delta \cdot E_{\epsilon} \left[R'_2(k_{2t}, \theta_{2t}, \epsilon_{2t}) - S(k_{2t}, \theta_t, \epsilon_t) \right] + c.$$

The first-term on the right hand side of the above inequality is non-positive from the definition of the shadow price. Therefore, $\Gamma < c$, and hence $k_{2t} > k_t^o$ cannot be an optimal strategy, which contradicts the hypothesis that $(0, k_{2t})$ is an equilibrium.

Proof of Proposition 3: We first note that the efficient capacity level in the dedicated capacity setting, \bar{k}_t^o , can be alternatively defined by the following equation:

$$S(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t) = c. \tag{28}$$

This holds because (i) $S(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t) \equiv R'_1(q_1^*(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t), \theta_{1t}, \bar{\epsilon}_{1t})$, and (ii) $q_1^*(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t) = \bar{k}_{1t}^o$, since $\bar{k}_t^o = \bar{k}_{1t}^o + \bar{k}_{2t}^o$ and, given assumption A1, \bar{k}_{it}^o satisfies:

$$E_{\epsilon_i}\left[R'_i(\bar{k}^o_{it},\theta_{it},\epsilon_{it})\right] = R'_i(\bar{k}^o_{it},\theta_{it},\bar{\epsilon}_{it}) = c.$$

The efficient capacity level in the fungible capacity setting is given by:

$$E_{\epsilon}[S(k_t^o, \theta_t, \epsilon_t)] = c \tag{29}$$

When $S(\cdot)$ is linear in ϵ_t , $E_{\epsilon}[S(k_t^o, \theta_t, \epsilon_t)] = S(k_t^o, \theta_t, \bar{\epsilon}_t)$. Equations (28) and (29) therefore imply that $k_t^o = \bar{k}_t^o$.

If $S(\cdot)$ is concave in ϵ_t , the application of Jensen's inequality yields:

$$E_{\epsilon}[S(\bar{k}^{o}_{t},\theta_{t},\epsilon_{t})] < S(\bar{k}^{o}_{t},\theta_{t},\bar{\epsilon}_{t}) = c.$$
(30)

The result $k_t^o < \bar{k}_t^o$ then follows because $S(k, \theta_t, \epsilon_t)$ is decreasing in k. A similar argument proves that $k_t^o > \bar{k}_t^o$ when $S(\cdot)$ is convex.

Proof of Lemma 4:

Division 1's optimal response, $k_t^*(k_{2t})$, is unique and given by the first-order condition in (13). Therefore, Division 1 will choose the first-best capacity k_t^o if and only if k_{2t} satisfies:

$$E_{\epsilon}[R_{1}^{'}(k_{t}^{o}-k_{2t},\theta_{1t},\epsilon_{1t})]=c.$$

Since $E_{\epsilon}[R'_1(\bar{k}^o_{1t}, \theta_{1t}, \epsilon_{1t})] = c$ and $R'_1(k_1, \cdot, \cdot)$ is decreasing in k_1 , the above equation can be satisfied if and only if $k^o_t - k_{2t} = \bar{k}^o_{1t}$. To finish the proof, we need to show that the resulting capacity level for Division 2 is non-negative; i.e., $k_{2t} = k^o_t - \bar{k}^o_{1t} \ge 0$. We note that:

$$c = E_{\epsilon}[R_1'(\bar{k}_{1t}^o, \theta_{1t}, \epsilon_{1t})] = E_{\epsilon}[S(k_t^o, \theta_t, \epsilon_t)] \ge E_{\epsilon}[R_1'(k_t^o, \theta_{1t}, \epsilon_{1t})],$$

where the last inequality follows from the definition of shadow price. Since $R'_i(k, \cdot, \cdot)$ is decreasing in k, it follows that $k^o_t \ge \bar{k}^o_{1t}$. \Box

Proof of Proposition 4: Proposition 3 shows that $k_t^o = \bar{k}_t^o \equiv \bar{k}_{1t}^o + \bar{k}_{2t}^o$ when the shadow price $S(k_t, \theta_t, \epsilon_t)$ is linear in ϵ_t . Consequently, it suffices to show that the adjusted full cost transfer pricing induces division *i* to secure \bar{k}_{it}^o units of capacity for each $i \in \{1, 2\}$. Since $R'_i(\cdot, \cdot, \epsilon_{it})$ and $S(\cdot, \cdot, \epsilon_t)$ are linear in ϵ_t , it follows that:

$$E_{\epsilon_i}[R'_i(k_{it},\theta_{it},\epsilon_{it})] = R'_i(k_{it},\theta_{it},\bar{\epsilon}_{it})$$

and

$$E_{\epsilon}[S(k_t, \theta_t, \epsilon_t)] = S(k_t, \theta_t, \bar{\epsilon}_t).$$

As a consequence, the divisional first-order conditions under the full cost transfer system become:

$$(1-\delta) \cdot R_1'(k_{1t}^*, \theta_{1t}, \bar{\epsilon}_{1t}) + \delta \cdot S(k_{1t}^* + k_{2t}, \theta_t, \bar{\epsilon}_t) = c$$
(31)

$$\delta \cdot R_2'(k_{2t}^*, \theta_{2t}, \bar{\epsilon}_{2t}) + (1 - \delta) \cdot S(k_t^*, \theta_t, \bar{\epsilon}_t) + \frac{\partial k_{1t}^*}{\partial k_{2t}} \cdot \{S(k_t^*, \theta_t, \bar{\epsilon}_t) - c\} = c.$$
(32)

We recall from the proof of Proposition 2 that when the shadow price $S(k_t, \theta_t, \epsilon_t)$ is linear in ϵ_t , the efficient capacity level is given by:

$$S(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t) = c.$$

Substituting this in (31) and (32) shows that $(\bar{k}_{1t}^o, \bar{k}_{2t}^o)$ is indeed a solution to the first-order conditions.

Given that Division 1's optimal response is uniquely given by $k_{1t}^*(k_{2t})$, it suffices to show that \bar{k}_{2t}^o is the unique maximizer of Division 2's objective function. To prove this, we will show that Division 2's profit function is single-peaked, i.e., $\pi_2'(k_{2t}) < 0$ for all $k_{2t} > \bar{k}_{2t}^o$ and $\pi_2'(k_{2t}) > 0$ for all $k_{2t} < \bar{k}_{2t}^o$.

Case I: $k_{2t} > \bar{k}_{2t}^o$ Suppose $k_{2t} = \bar{k}_{2t}^o + \Delta_2$ and $\Delta_2 > 0$. Let $k_{1t}^* = \bar{k}_{1t}^o - \Delta_1$ denote division 1's optimal response. We note that $0 < \Delta_1 < \Delta_2$, since $-1 < \frac{\partial k_{1t}^*}{\partial k_{2t}} < 0$. Consequently,

$$k_t^*(\bar{k}_{2t}^o + \Delta_2) = \bar{k}_{1t}^o - \Delta_1 + \bar{k}_{2t}^o + \Delta_2 = \bar{k}_t^o + \Delta,$$

where $\Delta = \Delta_2 - \Delta_1 > 0$. Suppressing θ_t and ϵ_t as arguments, we can write:

$$\pi_{2}'(\bar{k}_{2t}^{o} + \Delta_{2}) = \delta \cdot R_{2}'(\bar{k}_{2t}^{o} + \Delta_{2}) + (1 - \delta) \cdot S(\bar{k}_{t}^{o} + \Delta) + \frac{\partial k_{1t}^{*}}{\partial k_{2t}} \cdot \left\{ S(\bar{k}_{t}^{o} + \Delta) - c \right\} - c.$$
(33)

Since $0 > \frac{\partial k_{1t}^*}{\partial k_{2t}} > -1$ and $S(\bar{k}_t^o + \Delta) < S(\bar{k}_t^o) = c$, it follows that:

$$\frac{\partial k_{1t}^*}{\partial k_{2t}} \cdot \left\{ S(\bar{k}_t^o + \Delta) - c \right\} < c - S(\bar{k}_t^o + \Delta)$$

Therefore,

$$\pi_{2}^{'}(\bar{k}_{2t}^{o} + \Delta_{2}) < \delta \cdot \left[R_{2}^{'}(\bar{k}_{2}^{o} + \Delta_{2}) - S(\bar{k}_{t}^{o} + \Delta) \right]$$
(34)

Ex-post efficiency requires that any given capacity stock k must be reallocated so that $R'_1(q_1^*(k)) = R'_2(k - q_1^*(k))$. Furthermore, we note that:

$$R_1'(\bar{k}_{1t}^o - \Delta_1) > R_1'(\bar{k}_{1t}^o) = c$$

and

$$R_2'(\bar{k}_{2t}^o + \Delta_2) < R_2'(\bar{k}_{2t}^o) = c$$

Since $q_1^*(\bar{k}_t^o) = \bar{k}_{1t}^o$, it follows that

$$q_1^*(\bar{k}_t^o + \Delta) > \bar{k}_{1t}^o - \Delta_1$$
$$\Leftrightarrow \bar{k}_t^o + \Delta - q_1^*(\bar{k}_t^o + \Delta) < \bar{k}_{2t}^o + \Delta_2$$

Since $S(\bar{k}_t^o + \Delta) \equiv R'_2(\bar{k}_t^o + \Delta - q_1^*(\bar{k}_t^o + \Delta))$, the above inequality implies that:

$$S(\bar{k}_t^o + \Delta) > R_2'(\bar{k}_{2t}^o + \Delta_2).$$

It then follows from (34) that $\pi'_2(\bar{k}^o_{2t}+\kappa_2)<0$.

Case II: $k_{2t} < \bar{k}_{2t}^o$ Now suppose $k_{2t} = \bar{k}_{2t}^o - \Delta_2$ and $\Delta_2 > 0$ and let $k_t^* = \bar{k}_t^o + \Delta$, where $0 < \Delta < \Delta_2$. Since $\frac{\partial k_{1t}^*}{\partial k_{2t}} > -1$ and $S(\bar{k}_t^o - \Delta) > S(\bar{k}_t^o) = c$, it follows that:

$$\frac{\partial k_{1t}^*}{\partial k_{2t}} \cdot \left\{ S(\bar{k}_t^o - \Delta) - c \right\} > c - S(\bar{k}_t^o - \Delta)$$

Therefore,

$$\pi'_{2}(\bar{k}^{o}_{2t} + \Delta_{2}) > \delta \cdot \left[R'_{2}(\bar{k}^{o}_{2} - \Delta_{2}) - S(\bar{k}^{o}_{t} - \Delta) \right]$$

Following the similar arguments as used in case I above, it can be shown that

$$R_{2}'(\bar{k}_{2t}^{o} - \Delta_{2}) > R_{2}'(\bar{k}_{t}^{o} - \Delta - q_{1}^{*}(\bar{k}_{t}^{o} - \Delta)) \equiv S(\bar{k}_{t}^{o} - \Delta)$$

and hence $\pi'_{2}(\bar{k}^{o}_{2t} - \Delta_{2}) > 0.$

Proof of Proposition 5:

The proof is by contradiction. Consider first the case when the shadow price $S(k_t, \theta_t, \epsilon_t)$ is a concave function of ϵ_t , and hence $k_t^o < \bar{k}_t^o$. Lemma 3 shows that a cost-based system will induce the efficient capacity level if and only if $k_{2t} = k_t^o - \bar{k}_{1t}^o < \bar{k}_{2t}^o$. Since $k_t^*(k_{2t})$ is a monotonically increasing function, this implies that:

$$k_t^*(k_{2t}) > k_t^o \tag{35}$$

for all $k_{2t} \geq \bar{k}_{2t}^o$.

To generate the contradiction, suppose that the full cost transfer system induces underinvestment; i.e., $k_t^* \leq k_t^o$. Since $S(k, \cdot, \cdot)$ is a decreasing function, this implies that:

$$E_{\epsilon}[S(k_t^*, \theta_t, \epsilon_t)] \ge E_{\epsilon}[S(k_t^o, \theta_t, \epsilon_t)] = c.$$

Let $E_{\epsilon}[S(k_t^*, \theta_t, \epsilon_t)] = c + \Delta$ for some $\Delta \ge 0$. Substituting this in Division 2's first-order condition in (15) yields:

$$\delta \cdot E_{\epsilon}[R'_{2}(k_{2t}, \theta_{2t}, \epsilon_{2t})] + (1 - \delta) \cdot (c + \Delta) - \frac{\partial k_{1t}^{*}}{\partial k_{2t}} \cdot \Delta = c,$$

which implies that:

$$E_{\epsilon}[R_2'(k_{2t}, \theta_{2t}, \epsilon_{2t})] \le c, \tag{36}$$

because $\Delta \ge 0$ and $\frac{\partial k_{1t}^*}{\partial k_{2t}} < o$. The inequality in (36) implies:

$$k_{2t} \ge \bar{k}_{2t}^o$$

which, combined with (35), yields $k_t^* > k_t^o$. This, however, contradicts the maintained assumption that $k_t^* \leq k_t^o$. A similar argument proves that transfer at full cost induces under-investment when $S(\cdot, \cdot, \epsilon_t)$ is convex in ϵ_t .

Proof of Corollary 1: Consider first the case when the shadow price $S(k_t, \theta_t, \epsilon_t)$ is a concave function of ϵ_t . We wish to show that $p = c \cdot (1+m)$ can induce efficient investment only if m > 0. To the contrary, suppose $m \leq 0$ induces efficient investment k_t^o , where $k_t^o < \bar{k}_t^o$ by Proposition 2. Given Lemma 3, this requires that $p = c \cdot (1+m)$ induces Division 2 to choose:

$$k_{2t} = k_t^o - \bar{k}_{1t}^o < \bar{k}_{2t}^o.$$
(37)

Since $R'_i(\cdot, \cdot, \epsilon_{it})$ is linear, Division 2's first-order condition, when evaluated at $k_t^* = k_t^o$, yields:

$$\delta \cdot R_2'(k_{2t}, \theta_{2t}, \bar{\epsilon}_{2t}) + (1-\delta) \cdot c = c \cdot (1+m),$$

which implies that:

$$R_2'(k_{2t}, \theta_{2t}, \bar{\epsilon}_{2t}) \le c,$$

since $m \leq 0$ by assumption. As a consequence, we get:

$$k_{2t} \geq \bar{k}_{2t}^o$$

which contradicts (37), thereby proving the result. A similar argument proves that when $S(\cdot, \cdot, \epsilon_t)$ is convex in ϵ_t , a cost-based system can induce efficient investment only if m < 0.

Proof of Proposition 6: We first claim that if the two divisions reach an upfront agreement under which Division 2 receives $k_t^o - \bar{k}_{1t}^o$ units of capacity for some lump-sum transfer payment of TP_t , Division 1 will choose the efficient capacity level k_t^o . To prove this, note that, given $k_{2t} = k_t^o - \bar{k}_{1t}^o$, Division 1 will choose k_{1t} to maximize:

$$(1-\delta) \cdot E_{\epsilon} \left[R_1(k_{1t}, \theta_{1t}, \epsilon_{1t}) \right] + \delta \cdot E_{\epsilon} \cdot \left[M_f(k_t^o + k_{1t} - \bar{k}_{1t}^o, \theta_t, \epsilon_t) - R_2(k_t^o - \bar{k}_{1t}^o, \theta_{2t}, \epsilon_{2t}) \right] - c \cdot (k_t^o + k_{1t} - \bar{k}_{1t}^o).$$

We note that TP_t is a sunk payment, and hence irrelevant to Devision 1's capacity decision. The above maximization problem's first-order condition, which is necessary as well as sufficient, yields

$$E_{\epsilon}\left[(1-\delta)\cdot R_1'(k_{1t},\theta_{1t},\epsilon_{1t})+\delta\cdot S(k_t^o+k_{1t}-\bar{k}_{1t}^o,\theta_t,\epsilon_t)\right]=c,$$

which shows that Division 1 will indeed choose $k_{1t} = \bar{k}_{1t}^o$, and hence $k_t = k_t^o$.

To complete the proof, we need to show that there exists a transfer payment TP_t such that the ex-ante contract $(k_t^o - \bar{k}_{1t}^o, TP_t)$ will be preferred by both divisions to the default point of no agreement. If the two divisions fails to reach an agreement, Division 1 will choose its capacity level unilaterally, and Division 2 will receive no capacity rights (i.e., $k_{2t} = 0$). Let \hat{k}_t denote Division 1's optimal choice of capacity under this "default" scenario. Division 1's expected payoff under the default scenario is then given by:

$$\hat{\pi}_{1t} = E_{\epsilon} \left[(1 - \delta) \cdot R_1(\hat{k}_t, \theta_{1t}, \epsilon_{1t}) + \delta \cdot M_f(\hat{k}_t, \theta_t, \epsilon_t) \right] - c \cdot \hat{k}_t,$$
(38)

while Division 2's default payoff is:

$$\hat{\pi}_{2t} = (1 - \delta) \cdot E_{\epsilon} \left[M_f(\hat{k}_t, \theta_t, \epsilon_t) - R_1(\hat{k}_t, \theta_{1t}, \epsilon_{1t}) \right]$$

By agreeing to transfer $k_t^o - \bar{k}_{1t}^o$ units of capacity rights to Division 2, the two divisions can increases their ex-ante joint surplus by:

$$\Delta M \equiv E_{\epsilon}[M_f(k_t^o, \theta_t, \epsilon_t) - c \cdot k_t^o] - E_{\epsilon}[M_f(\hat{k}_t, \theta_t, \epsilon_t) - c \cdot \hat{k}_t].$$

The two divisions can then split this additional surplus between them in proportion to their relative bargaining power. The transfer price that implements this is given by:

$$E_{\epsilon} \left[(1-\delta) \cdot R_1(\bar{k}^o_{1t}, \theta_{1t}, \epsilon_{1t}) + \delta \cdot \left[M_f(k^o_t, \theta_t, \epsilon_t) - R_2(k^o_t - \bar{k}^o_{1t}, \theta_{2t}, \epsilon_{2t}) \right] \right] + TP_t = \hat{\pi}_{1t} + \delta \cdot \Delta M,$$

Division 2's expected payoff with this choice of transfer payment will be equal to $\hat{\pi}_{2t} + (1 - \delta) \cdot \Delta M$. Therefore, both divisions will prefer the upfront contract $(k_t^o - \bar{k}_{1t}^o, TP_t)$ to the default point of no agreement.

References

- Alles, M., and S. Datar (1998), "Strategic Transfer Pricing," Management Science, 44, 451-461.
- Anctil, R. and S. Dutta (1999), "Negotiated Transfer Pricing and Divisional vs. Firm-Wide Performance Evaluation," The Accounting Review 74(1), 87-104.
- Arrow, K. (1964). "Optimal Capital Policy, The Cost of Capital and Myopic Decision Rules," Annals of the Institute of Statistical Mathematics 1-2, 21-30.
- Arya, A. and Mittendorf B. (2008), "Pricing Internal Trade to Get a Leg Up on External Rivals," Journal of Economics and Management Strategy 17(3), 709-731.
- Arya, A., J. Glover, and R. Young (1996), "Capital Budgeting in a Multidivisional Firm," Journal of Accounting, Auditing and Finance 11(4), 519-533.
- Baiman, S. and M. Rajan (1995), "Centralization, Delegation, and Shared Responsibility in the Assignment of Capital Investment Decision Rights," *Journal of Accounting Research* 33 (Supplement), 135-164.
- Balakrishnan, R. and S. Sivaramakrishnan (2002), "A Critical Overview of Full-Cost Data for Planning and Pricing," *Journal of Management Accounting Research* 3-31.
- Baldenius, T., Dutta, S. and S. Reichelstein (2007), "Cost Allocations for Capital budgeting Decisions," The Accounting Review 82(4), 837-867.
- Baldenius, T. and S. Reichelstein, and S. Sahay (1999), "Negotiated versus Cost-Based Transfer Pricing," *The Review of Accounting Studies* 4(2), 67-91.
- Banker, R. and J.Hughes (1994), "Product Costing and Pricing," *The Accounting Review* 69(3), 479-494.
- Bastian, N. and S. Reichelstein (2004), "Transfer Pricing at Timken," Stanford GSB, Case # A-190.
- Christensen, P., G. Feltham and M. Wu (2002), " 'Cost of Capital' in Residual Income Measurement Under Moral Hazard," The Accounting Review 77(1), 1-23.
- Dutta, S., and S. Reichelstein (2002), "Controlling Investment Decisions: Depreciation and Capital Charges," *Review of Accounting Studies* 7, 253-281.
- Eccles, R. (1985), *Transfer Pricing: A Theory for Practice*, Lexington, MA: Lexington Books.

- Eccles, R., and H. White (1988), Price and Authority in Inter-Profit Center Transactions, American Journal of Sociology 94 (Supplement), 17-51.
- Edlin, A. and S. Reichelstein (1995), "Negotiated Transfer Pricing: An Efficiency Result," *The Accounting Review* 69(3), 479-494.
- Ernst & Young (2003), Global Transfer Pricing Survey, www.ey.com
- Goex, R. (2002), "Capacity Planning and Pricing under Uncertainty," Journal of Management Accounting Research 59-79.
- Grossman, S. and O. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy* 94 (August), 691-719.
- Hotelling, H. (1925). "A General Mathematical Theory of Depreciation," Journal of the American Statistical Association 20, 340-353.
- Hughes, J. and J. Kao (1997), "Cross Subsidization, Cost Allocation and Tacit Collusion," *Review of Accounting Studies* 2(3), 265-293.
- Kaplan, R. and A. Atkinson (1996), Advanced Management Accounting, Englewood Cliffs, NJ: Prentice Hall.
- Kouvelis, P. and M. Lariviere (2000), "Decentralizing Cross-functional Decisions: Coordination through Internal Markets" *Management Science* Vol. 46(8), 1049-1058.
- Martinez-Jerez, A. (2007), "Understanding Customer Profitability at Charles Schwab," Harvard Business School, Case # 106-002.
- Merchant, K. (2000), *Modern Management Control Systems*, 11th ed., Englewood Cliffs, NJ: Prentice Hall.
- Pfeiffer, T., and G. Schneider (2007), "Residual Income Based Compensation Schemes for Controlling Investment Decisions under Sequential Private Information", *Management Science* 53(3), 495-507.
- Pfeiffer, T., Lengsfeld, S., Schiller, U. and J. Wagner (2008), "Cost-based Transfer Pricing," Working paper, University of Vienna.
- Plambeck, E. and T. Taylor (2005), "Sell the Plant: The Impact of Contract Manufacturing on Innovation, Capacity and Profitability," *Management Science* 51, 133-150.
- Porteus, E. and Wang, S. (1991), "On Manufacturing/ Marketing Incentives" Management Science Vol. 37(9), 1166-1181.
- Rajan, M. and S. Reichelstein (2008). "Depreciation Rules and The Relation Between

Marginal and Historical Cost," Working paper, Stanford University.

- Reichelstein, S. (1997), "Investment Decisions and Managerial Performance Evaluation," *Review of Accounting Studies* 2, 157-180.
- Rogerson, W. (1997), "Inter-Temporal Cost Allocation and Managerial Investment Incentives: A Theory Explaining the Use of Economic Value Added as a Performance Measure," *Journal of Political Economy* 105, 770-795.
- Rogerson, W. (2008). "Inter-Temporal Cost Allocation and Investment Decisions," *Journal* of *Political Economy*, forthcoming.
- Sahay, S. (2002), "Transfer Pricing Based on Actual Cost," Journal of Management Accounting Research 15, 177-192.
- Schnell, R. (1995), Die Transferpreisrechnung als Voraussetzung fuer Dezentrales Controlling, in *Rechnungswesen und EDV*, A. Scheer(ed.), Physica Verlag, Heidelberg.
- Solomons, D. (1965), *Divisional Performance Measurement and Control*, Irwin, Homewood, Illinois.
- Tang, R. (2002), Current Trends and Corporate Cases in Transfer Pricing, Westport, CT: Quorum Books.
- Tirole, J. (2002), "Incomplete Contracts: Where Do We Stand?" *Econometrica* Vol. 52 (4).
- V. Mieghem, J. (2003), "Capacity Management, Investment and Hedging: Review and Recent Developments." *Manufacturing and Service Operation Management* Vol. 5 (4), 269-302.
- Wei, D. (2004), "Interdepartmental Cost Allocation and Investment Incentives," Review of Accounting Studies Vol. 9, 97-116.
- Wielenberg, S. (2000), "Negotiated Transfer Pricing, Specific Investment and Optimal capacity Choice" Review of Accounting Studies Vol. 5, 197-216.
- Williamson, O. (1985), Economic Institutions of Capitalism, New York London-The Free Press.
- Zimmerman, J. (1979), "The Costs and Benefits of Cost Allocations," The Accounting Review 54, 504-521.