

UC Berkeley

UC Berkeley Previously Published Works

Title

Decentralized capacity management and internal pricing

Permalink

<https://escholarship.org/uc/item/4r1138c6>

Journal

Review of Accounting Studies, 15(3)

ISSN

1573-7136

Authors

Dutta, Sunil
Reichelstein, Stefan

Publication Date

2010-09-01

DOI

10.1007/s11142-010-9126-3

Peer reviewed

Decentralized capacity management and internal pricing

Sunil Dutta · Stefan Reichelstein

Published online: 3 April 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract This paper studies the acquisition and subsequent utilization of production capacity in a multidivisional firm. In a setting where an upstream division provides capacity services for itself and a downstream division, our analysis explores whether the divisions should be structured as investment or profit centers. The choice of responsibility centers is naturally linked to the internal pricing rules for capacity services. As a benchmark, we establish the efficiency of an arrangement in which the upstream division is organized as an investment center, and capacity services to the downstream division are priced at full historical cost. Such responsibility center arrangements may, however, be vulnerable to dynamic hold-up problems whenever the divisional capacity assignments are fungible in the short-run, and therefore, it is essential to let divisional managers negotiate over their actual capacity assignments. The dynamic hold-up problem can be alleviated with more symmetric choice of responsibility centers. The firm can centralize ownership of capacity assets with the provision that both divisions rent capacity on a periodic basis from a central unit. An alternative and more decentralized solution is obtained by a system of bilateral capacity ownership in which both divisions become investment centers.

Keywords Performance management · Managerial incentives · Accrual accounting

JEL Classification M40 · M41 · D82

S. Dutta (✉)

Haas School of Business, University of California, Berkeley, CA, USA
e-mail: dutta@haas.berkeley.edu

S. Reichelstein (✉)

Graduate School of Business, Stanford University, Stanford, CA, USA
e-mail: reichelstein@stanford.edu

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. We identify environments in which such a decentralized structure results in efficient outcomes when the downstream division rents capacity in each period at a suitably chosen transfer price. In particular, we examine the use of full cost transfer prices that include historical cost charges for capacity assets installed by the upstream division.

The common reliance on full-cost (transfer) pricing in practice has been a challenge for research in managerial accounting. Generally, the use of full-cost prices is predicted to result in double-marginalization, as capacity related costs are considered sunk at the time internal transfers are being decided.³ However, this

¹ Capacity 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.

² The 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.

³ See, for instance, Balakrishnan and Sivaramakrishnan (2002), Goex (2002), Sahay (2003), Wei (2004), Pfeiffer et al. (2009), and Bouwens and Steens (2008).

logic requires modification when there is not just a single investment decision upfront but instead the firm undertakes a sequence of overlapping capacity investments. In a dynamic context of overlapping investments, Arrow (1964) identified the marginal cost of one unit capacity *for one period of time*, despite the fact that investments inherently create joint capacity over 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 allocated in proportion to the capacity available in a given period, relative to the total (discounted) capacity generated over the life of the asset.⁴

Building on the insights of Arrow (1964) and Rogerson (2008), we establish a benchmark result showing the efficiency of full cost transfer pricing. In particular, these prices include depreciation and imputed capital charges for past capacity investments. At the same time, these prices reflect the forward looking marginal cost of capacity services provided by the upstream division. Our benchmark result obtains in settings in which capacity is *dedicated* in the sense that the divisions' capacity usage is determined at the beginning of each period. Thus the production processes of the two divisions are sufficiently different so as to preclude redeployment of the aggregate capacity available in the short run, once the managers have received updated information on their divisional revenues.

In contrast, capacity may be *fungible* in the short run. It is then natural to allow the divisional managers to negotiate an adjustment to the initial capacity rights so as to capture any remaining trading gains that result from fluctuations in the divisional revenues. The resulting pricing rules then amount to a form of adjustable full cost transfer pricing. We find that such an organizational arrangement subjects the upstream division to a *dynamic hold-up* 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. Both divisions are then effectively regarded as profit centers with discretion to secure capacity for themselves at a full cost transfer price, which reflects the long-run marginal cost of capacity. Provided the central office can commit to such a transfer pricing policy, neither division can

⁴ The relative practical capacity rule is conceptually similar to the so-called *relative benefit rule* (Rogerson 1997), which has played a prominent role 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.

game the system by securing excessive capacity. The divisional incentives to secure capacity unilaterally arise from two sources: the autonomous use of the capacity secured and a share of the overall firm-wide revenue that is obtained with negotiated capacity adjustments.⁵ We find that the resulting divisional incentives are congruent with the firm-wide objective at least in an approximative sense, that is, to the extent that the divisional revenue functions can be approximated by quadratic functions.⁶

We finally examine the efficiency of a responsibility center arrangement that views both divisions as investment centers. Each division then attains “ownership” of capacity assets in accordance with its capacity demands, even though the downstream division may not have the technological expertise required to manage its assets. The resulting multi-period game creates a rich strategic interaction between the two divisions. Nonetheless we characterize equilibrium investment strategies that lead to efficient capacity acquisitions. The efficiency of such an organizational arrangement still relies on proper depreciation rules for capacity assets to ensure that division managers internalize the long-run marginal cost of capacity in each period. It also remains essential that divisional managers have the flexibility to negotiate over their capacity holdings in each period so as to take advantage of short-term fluctuations in product demand.

Taken together, our results suggest that, when a firm makes a sequence of overlapping investments, a symmetric responsibility center structure is more conducive to obtaining goal congruence. By centralizing ownership of all capacity assets and charging the divisions in each period a full cost transfer price that reflects the long-run marginal cost of capacity, the firm effectively decomposes the multi-period game into a sequence of disjoint one-period games. An alternative and more decentralized solution is to let the two divisions keep each other “in check” as part of the dynamic investment equilibrium that ensues between two investment centers. Either responsibility center structure avoids the dynamic hold-up problem that arises if one division can unilaterally drive up capacity acquisitions, while the other division is effectively “stuck” with the acquired assets.

The remainder of the paper is organized as follows. The model is described in Section 2. Section 3 examines an asymmetric organizational structure in which only the upstream division is an investment center. Section 4 considers an alternative organizational arrangement of centralized capacity ownership, which effectively treats the two divisions as profit centers. We explore a decentralized structure in which both divisions assume ownership of the capacity assets in Section 5. Extensions of our basic model are provided in Section 6, and we conclude in Section 7.

⁵ The incentive structure here is similar to that discussed in the incomplete contracting literature. Each party’s return to relationship-specific investments has two sources: the unilateral status-quo payoff and a share of the overall surplus available to both parties; see, for example, Bolton and Dewatripont (2005, chapter 12).

⁶ To obtain an exact solution for general revenue functions, the central office will need to coordinate the divisional capacity requests in a different manner. We extend our basic model in Section 6 and demonstrate the efficiency of a “gatekeeper” arrangement, which puts one division in charge of any new capacity acquisitions.

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 that views the upstream division as an investment center whose balance sheet reflects the historical cost of past capacity investments. In that sense, the upstream division acquires economic “ownership” of the assets. The downstream division (Division 2), in contrast, rents capacity on a periodic basis and therefore 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 units of capacity at date $t - 1$, the beginning of period t , is given by:

$$C_t = v \cdot b_t.$$

For reasons of notational and expositional parsimony, we assume that assets have a useful life of $n = 2$ periods. As argued below, all our results would be unchanged for a general useful life of n periods. If b_t units of capacity are installed at date $t - 1$, they become fully functional in period t . At the same time, the practical capacity declines to $b_t \cdot \beta$ in period $t + 1$. 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:

$$k_t = b_t + \beta \cdot b_{t-1}, \quad (1)$$

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. We denote the amount of capacity that Division 2 has reserved for itself in period t by k_{2t} . 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 $R_i(q_{it}, \theta_{it}, \epsilon_{it})$.⁷ 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.

⁷ If 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.

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} .⁸ While the random variables $\{\theta_t\}$ will be serially correlated, the transitory shocks $\{\epsilon_{it}\}$ are assumed to be identically and independently distributed across time; that is, $Cov(\epsilon_{it}, \epsilon_{i\tau}) = 0$ for each $t \neq \tau$, though in any given period these shocks may be correlated across divisions and therefore $Cov(\epsilon_{1t}, \epsilon_{2t}) \neq 0$.

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, $\theta_{i,t+1} \geq \theta_{it}$ for all t .⁹ As a consequence, the *expected* marginal revenues are nondecreasing over time, that is,

$$E_\epsilon \left[R'_i(q, \theta_{i,t+1}, \epsilon_{i,t+1}) \right] \geq E_\epsilon \left[R'_i(q, \theta_{it}, \epsilon_{it}) \right], \quad (2)$$

for all $q \geq 0$, 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 $\theta_t = (\theta_{1t}, \theta_{2t})$.¹⁰ 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 .

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.

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 earlier literature, we assume that each divisional manager is

⁸ The 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.

⁹ For 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.

¹⁰ As 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.

given a sequence of performance measures $\pi_i = (\pi_{i1}, \dots, \pi_{iT})$. Our analysis focuses on two candidate performance metrics: accounting *income* for profit centers and *residual income* for investment centers. Of course, the specifics of the accounting rules, including the depreciation- and transfer pricing rules, determine how these two measures are computed for a given set of transactions.

The manager of Division i is assumed to attach the non-negative weight u_{it} to his performance measure in period t . One can think of the weights $u_i = (u_{i1}, \dots, u_{iT})$ as reflecting a manager’s discount factor as well as the bonus coefficients attached to the periodic performance measures. At the beginning of period 1, manager i ’s objective function can thus be written as $\sum_{t=1}^T u_{it} \cdot E[\pi_{it}]$.

A performance measure is said to be *goal congruent* if it induces managers to make decisions that maximize the present value of firm-wide cash flows. We require the desired incentives to be *robust* in the sense that they hold regardless of the weights (u_{i1}, \dots, u_{iT}) . Formally, a performance measurement system is said to attain *strong goal congruence* if for any $u_i > 0$ the resulting game has a subgame perfect equilibrium in which the divisions choose the optimal capacity levels in each period.

3 Unilateral divisional capacity ownership

If one division in a firm has technical expertise in acquiring and maintaining production capacity, it is natural to consider a responsibility center structure that makes that division an investment center with “ownership” of the capacity assets. At the same time, this upstream division may be instructed to provide capacity service on a periodic basis to the other downstream division. Figure 1 depicts the sequence of events in a representative period.

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 financial performance of the upstream division is assumed to be measured by its residual income:

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

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 contains 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

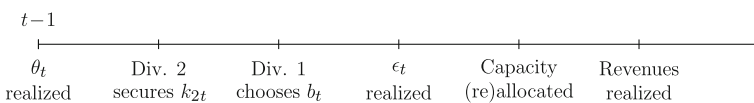


Fig. 1 Events in period t : divisional capacity ownership

asset’s acquisition cost. We let the parameter d represent the depreciation charge in period t per dollar of capacity investment undertaken in that period. The remaining book value $v \cdot b_t \cdot (1 - d)$ will be depreciated in period $t + 1$. Thus the total depreciation charge for Division 1 in period t can be written as:

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

and the historical cost value of the net assets at date $t - 1$ is given by:

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

3.1 Benchmark scenario: dedicated capacity

If the production processes of the two divisions are sufficiently different, it may be impossible for the divisions to redeploy the available capacity stock k_t in period t after the random shocks ϵ_t are realized. We refer to such a setting as one of *dedicated capacity* as 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 seeks a path of efficient investment and capacity levels so as to maximize the stream of discounted future cash flows. Suppose hypothetically that a central planner had the entire information regarding future revenues, that is, the sequence of future θ_t ’s. The optimal investment decisions $b \equiv (b_1, b_2, \dots)$ would be chosen so as to maximize the net present value of the firm’s expected future cash flows:

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

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

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

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

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

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

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.¹¹ The non-negativity constraints for new investments, $b_t \geq 0$, will not bind provided the corresponding sequence of capacity levels $k = (k_1, k_2, \dots)$ satisfy the monotonicity requirement $k_{t+1} \geq k_t$ for all t . This latter condition will be met whenever the expected marginal revenues satisfy the monotonicity condition in (2).¹²

Lemma 1 also identifies c as the effective long-run marginal cost of capacity. An intuitive argument for this characterization is that the firm can increase its capacity in period 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 in period t , buy β unit less in period $t + 1$, buy β^2 more unit in period $t + 2$, and so on. The cost of this variation, evaluated in terms of its present value as of date $t - 1$, is given by:

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

and therefore the present value of the variation at date t (i.e., the end of period 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 time period*. It is useful to note that c is exactly the price that a hypothetical supplier would charge for renting out capacity for one period, 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*.¹³

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. 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:

¹¹ The statement in Lemma 1 assumes implicitly that $\bar{k}_{it}^0 > 0$. A sufficient condition for this to hold is the following boundary condition: $R'_i(0, \theta_{it}, \epsilon_{it}) > c$ for all $\theta_{it}, \epsilon_{it}$.

¹² Condition (2) is obviously sufficient for the targeted capacity levels to be non-decreasing over time. The lower the productivity persistence parameter β , the more condition (2) could be relaxed while preserving the finding in Proposition 1. As shown by Rogerson (2008), the unit cost of capacity made available at date t for one period of time will generally exceed the value c in (7) if condition (2) does not hold. The designer then faces the potential problem of acquiring capacity today that may not be valuable given next period’s lower expected marginal revenue.

¹³ It is straightforward to extend the above analysis to a setting where assets have a useful life of n periods. For an investment undertaken at date t , the practical capacity available at date $t + i$, $1 \leq i \leq n$ is β_i , with $\beta_i \leq 1$. Provided the capacity levels are (weakly) decreasing over time i.e., $\beta_i \geq \beta_{i+1}$, the marginal cost of obtaining one unit of capacity for one period is then given by:

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

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

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's residual income measure in period t then becomes:

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

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

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

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:

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

We note that $0 < d < 1$ and:

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

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.¹⁴ 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 $r = 0.1$, the relative practical capacity rule amounts to straight line depreciation if the practical capacity in the second period declines to 91%.

Suppose now the firm depreciates investments according to the relative practical capacity rule and the transfer price for capacity services charged to Division 2 is based on the full historical cost (which includes the imputed interest charges).¹⁵ As

¹⁴ The term *relative practical capacity* rule has been coined in Rajan and Reichelstein (2009), 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 based only on the anticipated pattern of an asset's productivity over time but not 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) and the *economic depreciation* rule proposed by Hotelling (1925). As demonstrated in Rajan and Reichelstein (2009), these depreciation rules are generally different, though they coincide in certain special cases, most notably if all investments have zero NPV.

¹⁵ 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

a consequence, both divisions will be charged the competitive rental price c per unit of capacity in each period. The key difference in the treatment of the two divisions is that the downstream division can rent capacity on an “as needed” basis, while capacity investments entail a multi-period 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 charged against 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; that is, each division chooses its capacity level.

Proposition 1 When capacity is dedicated, a system of unilateral capacity ownership combined with full cost transfer pricing achieves strong goal congruence.

As demonstrated in the proof of Proposition 1, the divisional managers face a T -period game with a unique subgame perfect equilibrium.¹⁶ Irrespective of past decisions, the downstream division has a dominant strategy incentive to secure the optimal capacity level because it is charged the relevant 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; that is, condition (2) is met, the upstream division will not find itself in a position of excess capacity, provided the downstream division follows its dominant strategy.¹⁷

The result in Proposition 1 makes a strong case for full-cost transfer pricing, that is, 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 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), Bouwens and Steens (2008) and others, it has been difficult for the academic accounting literature to justify the use 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

Footnote 15 continued

consistent with the practice of full cost transfer pricing that features prominently in most surveys on transfer pricing; see, for instance, Ernst and Young (2003), Tang (2002).

¹⁶ The game has other Nash equilibria, which are not perfect and may result in inefficient capacity levels.

¹⁷ It should be noted that the result in Proposition 1 in no way requires 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. However, the formal claim in Proposition 1 needs to be modified if the division managers have private information, since the resulting game then has no proper subgames. Specifically, the concept of subgame perfect equilibrium could be replaced by another equilibrium concept requiring sequential rationality, such as Bayesian perfect equilibrium.

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. This argument can be made in a “clockwork environment” in which there are no random disturbances (i.e., $\epsilon_t \equiv \bar{\epsilon}_t$). At date t , the cost of capacity investments for that period is sunk, yet the opportunity cost of capacity is equal to c , precisely because at date $t - 1$ each division secured capacity up to the point where its marginal revenue is equal to c . Once there are random fluctuations in the divisional revenues, however, there is no reason to believe that the opportunity costs at date t relate systematically to the historical fixed costs at the earlier date $t - 1$.

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.¹⁸

3.2 Fungible capacity

We now relax the assumption of dedicated capacity. A plausible alternative scenario, which we maintain throughout the remainder of this paper, is that the demand shocks ϵ_t are realized sufficiently early in any given period and the 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, $\sum_{i=1}^2 R_i(q_{it}, \theta_{it}, \epsilon_{it})$, 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:

$$R'_1(q_{1t}^*, \theta_{1t}, \epsilon_{1t}) = R'_2(k_t - q_{1t}^*, \theta_{2t}, \epsilon_{2t}). \quad (10)$$

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

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

¹⁸ Banker 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 is 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.

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 now given by:

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

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:

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

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, since 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, a system of unilateral capacity ownership combined with 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 equilibrium that 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 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.¹⁹

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 organizational 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 capacity provider 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. As a consequence, the downstream division will then no longer be able to “hold-up” the upstream division as this division is no longer the residual claimant of capacity rights.

¹⁹ The proof of Proposition 2 exploits that, in accordance with the notion of strong goal congruence, the downstream division may not put much weight on the loss of first-period profit that results from strategically exaggerating its capacity needs in that period.

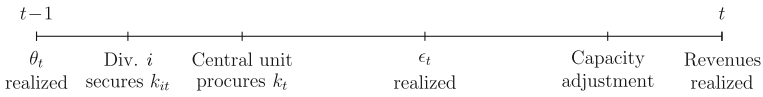


Fig. 2 Events in period t : centralized capacity ownership

Initially, we suppose that the central capacity provider charges the divisions the full cost c per unit of capacity. Since this charge coincides with 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. The sequence of events in a representative period is depicted on the following timeline (Fig. 2).

After the two managers have observed the realization of the demand 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

$$R_2^*(k_{1t}, k_{2t} | \theta_t, \epsilon_t) = \delta \cdot R_2(k_{2t}, \theta_{2t}, \epsilon_{2t}) + (1 - \delta) \cdot [M_f(k_t, \theta_t, \epsilon_t) - R_1(k_{1t}, \theta_{1t}, \epsilon_{1t})].$$

Taking division j 's capacity request k_{jt} as given, division i will choose k_{it} to maximize:

$$E_\epsilon[\pi_{it}] \equiv E_\epsilon[R_i^*(k_{it}, k_{jt} | \theta_t, \epsilon_t)] - c \cdot k_{it}. \tag{13}$$

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 .

Consider now a Nash-equilibrium (k_{1t}^*, k_{2t}^*) equilibrium of the stage game played in period t .²⁰ If $k_{it}^* > 0$ for each i , then by the Envelope Theorem the following first-order conditions are met:

$$E_\epsilon \left[(1 - \delta) \cdot R_1'(k_{1t}^*, \theta_{1t}, \epsilon_{1t}) + \delta \cdot S(k_{1t}^* + k_{2t}^*, \theta_t, \epsilon_t) \right] = c \tag{14}$$

and

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

We note that since $R_i'(\cdot)$ and $S(\cdot)$ are decreasing functions of k_{it} , Division i 's objective function is globally concave, and therefore there is a unique best response k_{it}^* for any given conjecture regarding k_{jt} .

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-

²⁰ The proof of Proposition 6 below shows that a pure strategy Nash equilibrium indeed exists.

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" hold-up problem.²¹ Yet the divisions also derive *autonomous* value from the capacity available to them, even if the overall capacity were not to be reallocated ex post. 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.²²

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

$$E_\epsilon \left[(1 - \delta) \cdot R'_1(k_t^o, \theta_{1t}, \epsilon_{1t}) + \delta \cdot S(k_t^o, \theta_t, \epsilon_t) \right]. \quad (16)$$

Yet, this marginal revenue is less than c because:

$$c = E_\epsilon [S(k_t^o, \theta_t, \epsilon_t)] = E_\epsilon \left[R'_1(q_1^*(k_t^o, \theta, \epsilon_t), \theta_{1t}, \epsilon_{1t}) \right] > E_\epsilon \left[R'_1(k_t^o, \theta, \epsilon_t), \theta_{1t}, \epsilon_{1t} \right]. \quad (17)$$

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 securing capacity rights on its own, such an approach would also induce the upstream division to underinvest as it would anticipate a traditional hold-up on its investment in the ensuing negotiation.

With centralized capacity ownership and fungible capacity, 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 "stage games" would also constitute a subgame perfect equilibrium for the T -period game. 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:

²¹ 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. (2009).

²² 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 return on their relationship-specific investments, even if the status quo were to be implemented.

Assumption(A1) $R_i(q_{it}, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \hat{R}_i(q_{it}, \theta_{it})$ and the shadow price $S(\cdot, \epsilon_t)$ is linear in ϵ_t .

A sufficient condition for linearity of $S(\cdot)$ is that the divisional revenues can be described by quadratic functions:

$$R_i(q, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot q - h_{it} \cdot q^2 \tag{18}$$

for some constants $h_{it} > 0$. For other standard functional forms of $R_i(\cdot)$, one obtains shadow prices that are nonlinear in ϵ_t . We discuss this aspect in more detail in Section 6.1 below and for now note that the following efficiency results apply only in an approximate sense, that is, to the extent that the divisional revenue functions can be approximated sufficiently well by second-order polynomials.²³

A shadow price of capacity linear in ϵ_t implies that the efficient capacity level with *fungible* capacity is equal to the sum of the efficient capacity levels in the *dedicated* capacity scenario. Formally, $k_t^o \equiv \bar{k}_{1t}^o + \bar{k}_{2t}^o$ ²⁴. Furthermore, 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; that is, $(\bar{k}_{1t}^o, \bar{k}_{2t}^o)$ is the unique maximizer of the divisional objective functions.

Proposition 3 Given A1, centralized capacity ownership combined with adjustable full cost transfer pricing achieves strong goal congruence.

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 and \bar{k}_{2t}^o .

The quadratic form in (18) 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 contexts. To that end, a second-order polynomial approximation of the form in (18) 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.²⁵

²³ It is readily verified that a logarithmic functional form, $R_i(q, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot \ln q$, also yields a shadow price $S(\cdot, \epsilon_t)$ linear in ϵ_t .

²⁴ A formal proof of this claim follows from the more general Proposition 5 in Section 6.1 below. It is useful to note that when $S(\cdot)$ is linear in ϵ_t , the function $k_t^o(\theta_{1t}, \theta_{2t})$ is separable such that $k_t^o(\theta_{1t}, \theta_{2t}) = \bar{k}_{1t}^o(\theta_{1t}) + \bar{k}_{2t}^o(\theta_{2t})$.

²⁵ Proposition 4 extends to settings in which each division has private information regarding θ_{it} . Replicating the steps in the proof of the proposition, it is readily verified that the strategies $k_{it}^*(\theta_{it}) \equiv \bar{k}_{it}^o(\theta_{it})$ then form 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 at the negotiation stage.

5 Bilateral divisional capacity ownership

The dynamic hold-up problem identified in Proposition 2 arose in an asymmetric responsibility center arrangement in which Division 2 is a profit center with rights to secure capacity from the upstream division on a period-by-period basis. This arrangement effectively enables Division 2 to force the upstream division to acquire long-term assets, whose subsequent utilization will then be renegotiated. Proposition 3 demonstrates that centralization of capacity ownership rights effectively addresses the hold-up problem, provided the central unit can commit to renting out capacity at the long-run marginal cost, c , regardless of the divisions' past requests for capacity.

A natural organizational alternative is to structure both divisions as independent investment centers, even though technical expertise may make it necessary for Division 1 to have physical control of the capacity assets. From an incentive perspective, the issue is that the ensuing game is then no longer separable, either intertemporally or cross-sectionally. We investigate whether in equilibrium the divisions will make the desired multi-period investment choices given that they anticipate periodic negotiations to adjust their current capacity holdings. In keeping with the structure of our model, the two divisions are assumed to acquire capacity simultaneously at the unit cost of v at the beginning of each period. The sequence of events in a representative period is depicted in the following timeline (Fig. 3).

Extending the investments center scenario described in Section 3, we assume that performance for each division is measured by its residual income:

$$\pi_{it} = Inc_{it} - r \cdot A_{i,t-1},$$

where A_{it} denotes book value of Division i 's capacity related assets at the end of period t . As before, investments are depreciated according to the relative practical capacity rule. Consequently, both divisions are effectively charged the competitive rental price c per unit of capacity in each period. Suppose Division i begins period t with

$$h_{it} \equiv \beta \cdot b_{i,t-1}$$

units of capacity and acquires b_{it} additional units in period t . It will then have a capacity stock of $k_{it} = h_{it} + b_{it}$ for use in period t . Upon observing the realization of the demand shock ϵ_t , the two managers are assumed to divide the total capacity available, $k_t = k_{1t} + k_{2t}$, so as to maximize the sum of the divisional revenues. As a consequence, Division i 's expected payoff in period t is given by:

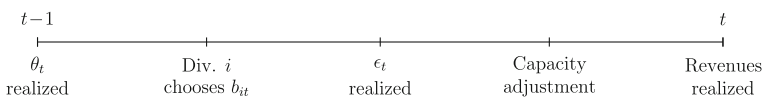


Fig. 3 Events in period t : bilateral capacity ownership

$$\pi_{it}(b_{1t}, b_{2t}) = E_\epsilon \left[(1 - \delta_i) \cdot R_i(k_{it}, \theta_{it}, \epsilon_{it}) + \delta_i \cdot \{M_f(k_t, \theta_t, \epsilon_t) - R_j(k_{jt}, \theta_{jt}, \epsilon_{jt})\} \right] - c \cdot k_{it}, \tag{19}$$

with $\delta_1 \equiv \delta$ and $\delta_2 \equiv (1 - \delta)$. For any given weights u_t that the divisions attach to their periodic performance measures, we obtain a multi-stage game in which each division makes one move in each period, the choice of b_{it} . The investment choices $b_t = (b_{1t}, b_{2t})$ and the state parameter θ_t are the new information variables in period t . However, each division’s future payoffs depend on the entire history of information variables only through $h_t \equiv (h_{1t}, h_{2t})$. Accordingly, a pure (behavior) strategy for Division i consists of T mappings $\{b_{it} : (h_t, \theta_t) \rightarrow \mathbb{R}_+\}$.

Proposition 4 Given A1, a system of bilateral capacity ownership combined with negotiated capacity adjustments achieves strong goal congruence.

To establish strong goal congruence, we show that a particular profile of strategies, which we term as “no escalation” strategies, constitutes a subgame perfect equilibrium of the T -stage game and leads each division to procure the efficient capacity stock in each period. These no-escalation strategies $b^* \equiv \{b_1^*(h_1, \theta_1), \dots, b_T^*(h_T, \theta_T)\}$ are defined as follows:

- If Division i finds itself with excess capacity in period t , that is, $h_{it} \geq \bar{k}_{it}^o(\theta_t)$, then $b_{it}^*(h_t, \theta_t) = 0$
- If both divisions are below their efficient capacity levels, that is, $h_{jt} \leq \bar{k}_{jt}^o(\theta_t)$, then Division i will invest up to that level: $b_{it}^*(h_t, \theta_t) = \bar{k}_{it}^o(\theta_t) - h_{it}$
- If $h_{it} < \bar{k}_{it}^o(\theta_t)$ but Division j starts with a capacity stock that exceeds the threshold level h_{jt}^+ , given by the first-order condition:

$$(1 - \delta_i) \cdot R_i'(h_{it}, \theta_{it}, \bar{\epsilon}_{it}) + \delta_i \cdot S(h_{it} + h_{jt}^+, \theta_t, \bar{\epsilon}_t) = c, \tag{20}$$

then $b_{it}^*(h_t, \theta_t) = 0$.

- If $h_{it} < \bar{k}_{it}^o(\theta_t)$ but Division j starts with “moderate” excess capacity $h_{jt} \in (\bar{k}_{jt}^o(\theta_t), h_{jt}^+)$, then $b_{it}^*(h_t, \theta_t) = b_{it}^+$, defined implicitly by:

$$(1 - \delta_i) \cdot R_i'(h_{it} + b_{it}^+, \theta_{it}, \bar{\epsilon}_{it}) + \delta_i \cdot S(h_{1t} + h_{2t} + b_{it}^+, \theta_t, \bar{\epsilon}_t) = c. \tag{21}$$

Figure 4 illustrates the no-escalation strategies. If both divisions adhere to $b_{it}^*(\cdot)$, it is readily seen that the efficient capacity level will be procured in each period, given that the firm starts out with zero capacity at date 0. We also note that for any history h_t , $b_{it}^*(h_t, \theta_t)$ is a Nash equilibrium strategy for the single-stage game that would be played if the divisions were to myopically maximize their current payoffs. In addition, the amount of new capacity acquisition b_{it}^* is (weakly) decreasing in the level of a division’s own old capacity, h_{it} , as well as in the level of old capacity of the other division, h_{jt} . Finally, strategy b^* has the property that a unilateral deviation from b^* in period t affects divisional capacity choices only in the current and next periods but has no effect on capacity choices in periods beyond $t + 1$.

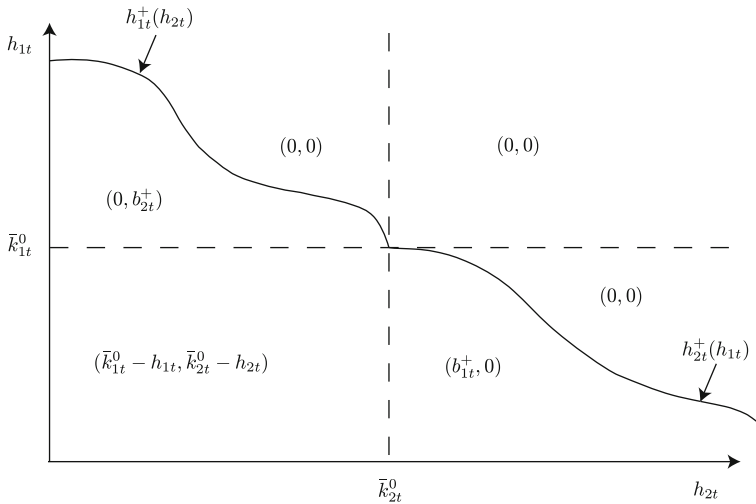


Fig. 4 Illustration of the no-escalation strategies

The proof of Proposition 4 exploits that any unilateral single-stage deviation from b^* in any given period t leaves the deviating division worse off in each of the two relevant periods, i.e., periods t and $t + 1$.²⁶ That the deviating division cannot benefit in the current period follows from the property that $b_t^*(h_t, \theta_t)$ is a Nash equilibrium strategy in the hypothetical static game in which divisions choose their capacities to myopically maximize their current payoffs. While Division i can effectively weaken Division j 's default bargaining status in the next period by acquiring excess capacity in the current period, and thereby inducing Division j to acquire less capacity in the next period, the proof of Proposition 4 shows that Division i cannot improve its payoff in the next period from such a deviation from the posited equilibrium strategy b^* .²⁷ Since any deviation from b^* makes the deviating division worse off not only in the aggregate over the entire planning horizon but also on a period-by-period basis, b^* is a subgame perfect equilibrium strategy for *all* values of $u_i \in \mathbb{R}_{++}^T$.

6 Extensions

6.1 Nonlinear shadow prices

Our first extension concerns the effect of relaxing the linearity assumption in A1. One important implication of A1 is that the optimal aggregate capacity level is the

²⁶ In our finite multi-stage games with observed actions, to prove that b^* is a subgame perfect equilibrium, it suffices to show that no division can do better by deviating from b^* in a single-stage. See the one-stage deviation principle in Theorem 4.1 of Fudenberg and Tirole (1991).

²⁷ Recall that Division i 's net revenue, $(1 - \delta_i) \cdot R_i(k_{it}, \theta_{it}, \epsilon_{it}) + \delta_i \cdot [M_f(k_t, \theta_t, \epsilon_t) - R_j(k_{jt}, \theta_{jt}, \epsilon_{jt})]$, depends on Division j 's default status at the bargaining stage, i.e., $R_j(k_{jt}, \theta_{jt}, \epsilon_{jt})$. Since b_{jt}^* is decreasing in $h_{it} = \beta \cdot b_{i,t-1}$, Division i can weaken Division j 's default status by over-investing in the previous period. However, the proof of Proposition 4 shows that such a strategy cannot be beneficial because of its "direct" cost $c \cdot k_{it}$.

same k_t^o in both the dedicated and the fungible capacity scenario. The following result shows that the curvature of the shadow price in ϵ is crucial in shaping the comparison between k_t^o and \bar{k}_t^o .

Proposition 5 Suppose that $R_i(q_{it}, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \hat{R}_i(q_{it}, \theta_{it})$. The optimal capacity level k_t^o in period t then satisfies:

$$k_t^o \begin{cases} \geq \bar{k}_t^o & \text{if } S(k_t, \theta_t, \epsilon_t) \text{ is convex in } \epsilon_t \\ \leq \bar{k}_t^o & \text{if } S(k_t, \theta_t, \epsilon_t) \text{ is concave in } \epsilon_t. \end{cases} \tag{22}$$

According to Proposition 5, the curvature of the shadow price determines whether a risk-neutral 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 more in capacity. The reverse holds when the shadow price is concave. We point out in passing that the assumption that $R_i(q_{it}, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \hat{R}_i(q_{it}, \theta_{it})$ 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 5 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 $k_t^o > \bar{k}_t^o$. On the other hand, $S(\cdot, \cdot, \epsilon_t)$ is concave when $R_i(q, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \theta_{it} \cdot (1 - e^{-q})$. As mentioned above, 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 that the noise term ϵ_{it} enters $R(\cdot)$ multiplicatively and yet $S(\cdot)$ is generally not a linear function of ϵ_t .

Returning to an organizational structure of centralized capacity ownership, Propositions 3 and 5 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 6 Suppose that $R_i(q_{it}, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \hat{R}_i(q_{it}, \theta_{it})$. Centralized capacity ownership combined with adjustable full cost transfer pricing then 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.²⁸

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).²⁹ 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.

6.2 Appointing a gatekeeper

One conclusion of Proposition 6 above is that even if the firm centralizes capacity ownership and charges the divisions the competitive rental prices of capacity, there will be a remaining coordination problem if the shadow price capacity is nonlinear in the random disturbances, ϵ_t . We now examine the idea of improved coordination by appointing the upstream division a "gatekeeper" who must approve capacity rights secured by the other division.

Suppose that, as in Section 3, a central unit procures new capacity as needed. However, 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.³⁰ 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

²⁸ Our findings here stand in contrast to earlier incomplete contracting models on transfer pricing, for instance, 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 only one of mitigating hold-ups and avoiding under-investment.

²⁹ The notion that firms may want to bias internal prices deliberately is, of course, also 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, a central planner "distorts" the internal price to achieve pre-commitment in the firm's competition with external rivals.

³⁰ We focus on the upstream division as a gatekeeper because this division was assumed to have unique technological expertise in installing and maintaining production capacity. Yet the following analysis makes clear that the role of the two divisions could be switched.

division would have no claim on capacity in that period, though it may, of course, obtain capacity ex post through negotiation with the other division.

Proposition 7 With centralized capacity ownership, a gatekeeper arrangement achieves strong goal congruence.

Since ownership of capacity assets is centralized, the divisional capacity choice problems are again separable across time periods. Therefore, a gatekeeper arrangement will attain strong goal congruence if it induces the two divisions to acquire collectively the capacity level k_t^o in each period. The proof of Proposition 7 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 .³¹ 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 gatekeeper arrangement. 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 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.

6.3 Optimal incentive contracting

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. Specifically, suppose that the divisional cash flows in period t are given by:

$$CF_{it} = R_i(k_{it}, \theta_{it}, \epsilon_{it}) + Y_{it}, \quad (23)$$

where Y_{it} can be contributed by manager i in period t at a personal cost $C_{it}(Y_{it}, \tau_i)$. The parameter τ_i is a time-invariant productivity parameter known only to the

³¹ We note that the multiplicative separability condition $R_i(q_{it}, \theta_{it}, \epsilon_{it}) = \epsilon_{it} \cdot \hat{R}_i(q_{it}, \theta_{it})$ is not required in establishing the claim in Proposition 7.

manager of Division i . Furthermore, suppose that each manager is risk neutral and discounts the future at the same rate as the firm's owners and that each manager's utility payoff in period t is given by his period t compensation less the personal cost $C_{it}(Y_{it}, \tau_i)$. Adapting the arguments in Edlin and Reichelstein (1995), it can be shown that the performance measures identified in Propositions 1, 4, and 6 are also the basis of optimal second-best incentive contracts. In particular, it is optimal, under certain conditions, to pay each manager a share of his performance measure π_{it} . Even if managers are equally patient, a proper allocation of the investment expenditures remains essential if the cost functions $C_{it}(Y_{it}, \tau_i)$ vary over time and, as a consequence, the intensity of the desired incentive provision varies over time.

The crucial feature of the cash flow specification in (25) is, of course, the assumed additive separability of divisional revenues and the managers' productive contributions Y_{it} . This separability implies that 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 the models of Baiman and Rajan (1995), Christensen et al. (2002), Dutta and Reichelstein (2002), Baldenius et al. (2007), and Pfeiffer and Schneider (2007) in contrast, the investment decisions are a source of informational rent for the manager, and as a consequence, optimality requires 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. It remains an open question for future research whether similar results can be obtained in the context of multiple overlapping investment decisions.

7 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 alternative responsibility center arrangements that differ in the economic ownership of capacity assets. One natural responsibility center structure is to make the supplier of capacity services (the upstream division) an investment center that is fully accountable for all of its capacity assets. In contrast, the downstream division is viewed as a profit center that rents capacity on a periodic basis. If these rentals are based on a transfer price that reflects the historical cost of capacity, composed of depreciation and imputed interest charges, and capacity investments are depreciated in accordance with the underlying utilization pattern, both divisions will 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.

For the benchmark setting of dedicated capacity, we conclude that all responsibility center arrangements considered in the above analysis are equally effective. However, this conclusion no longer applies if the production processes of the two divisions have enough commonalities so that capacity becomes fungible in

the short run. It is then 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 we find that with unilateral capacity ownership the corresponding system of adjustable full cost transfer pricing is generally vulnerable to a dynamic hold-up problem: the downstream division drives up its capacity demands opportunistically in anticipation of obtaining the corresponding excess capacity at a lower cost through negotiations in future periods.

One approach to alleviating dynamic hold-up problems is to have a central capacity provider with the ability to commit to capacity rentals on a period-by-period basis at the appropriate transfer price. While the divisions retain flexibility in negotiating adjustments to their capacity rights, neither profit center can gain from exaggerating or low-balling the desired initial capacity targets. We finally demonstrate that efficient investment decisions can also emerge for a fully decentralized structure that views both divisions as investment centers with residual control rights over their own assets. With symmetric rights and obligations, the divisions keep each other “in check” in equilibrium, even though both anticipate periodic negotiations over their current capacity rights.

Moving further afield, there appear to be several promising directions for extending the analysis in this paper. One set of extensions relates to the informational structure and the extent of private information that divisional managers have at various stages of the multi-period game. In terms of organizational design, we note that in many firms one upstream division provides capacity services to multiple downstream users. Auction mechanisms, rather than bilateral negotiations, then become natural candidates for allocating scarce capacity resources in the short-run. Such mechanisms have been examined in both the academic and practitioner literatures in one-shot settings; see, for example, Malone (2004), Plambeck and Taylor (2005) and Baldenius et al. (2007). The presence of multiple capacity buyers may mitigate the dynamic holdup problem encountered in our analysis. Intuitively, a downstream division derives fewer benefits from opportunistically driving up capacity acquisitions in one period if any excess capacity in future periods is sold off competitively, instead of being appropriated through bilateral negotiation.

Acknowledgments We would like to thank seminar participants at the following universities: Yale, Minnesota, Berlin (Humboldt), Tilburg, Toronto (Rotman), Ohio State, Texas at Austin, Harvard, Carnegie Mellon, Munich (LMU), London (Cass), Bern, Copenhagen and the Indian School of Business for their comments on an earlier version of this paper, titled “Decentralized Capacity Management.” We are particularly indebted to Bob Kaplan, Miles Gietzman, Robert Goex and two anonymous reviewers for their detailed comments. Finally, Alexander Nezlobin and Yanruo Wang provided valuable research assistance.

Open Access This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

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

$$\sum_{t=1}^{\infty} v \cdot (1+r) \cdot b_t \cdot \gamma^t = \sum_{t=1}^{\infty} c \cdot k_t \cdot \gamma^t$$

where $k_t = b_t + \beta \cdot b_{t-1}$. Direct substitution yields:

$$\begin{aligned} \sum_{t=1}^{\infty} v \cdot (1+r) \cdot b_t \cdot \gamma^t &= v \cdot [k_1 + \gamma[k_2 - \beta \cdot k_1] + \gamma^2[k_3 - \beta[k_2 - \beta \cdot k_1]] \\ &\quad + \gamma^3[k_4 - \beta[k_3 - \beta \cdot [k_2 - \beta \cdot k_1]]] + \dots \end{aligned}$$

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

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

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 flows can therefore be expressed as:

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

This optimization 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:

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

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

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:

$$E_{\epsilon} [R_2(k_{2T}, \theta_{2T}, \epsilon_{2T})] - c \cdot k_{2T}. \tag{24}$$

Let $\bar{k}_{2T}^o(\theta_{2T})$ denote the maximizer of (24). Division 1 faces the constrained optimization problem:

$$E_{\epsilon} [R_1(k_{1T}, \theta_{1T}, \epsilon_{1T})] - c \cdot k_{1T} \tag{25}$$

subject to the constraint $k_{1T} + \bar{k}_{2T}^o(\theta_{2T}) \geq \beta \cdot b_{T-1}$. Since Division 1’s objective

function in (25) 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 (25). In particular, the upstream division would invest $b_T = 0$ if $k_T^* = \beta \cdot b_{T-1}$.

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

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

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

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

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

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

Proof of Proposition 2 Let $T = 2$ and suppose that the random shocks ε_t assume their expected value $\bar{\varepsilon}$ for sure. Furthermore, suppose that the divisional revenue functions are identical both cross-sectionally 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, $R'_i(k, \theta, \bar{\varepsilon}) = c$. 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 $b_1 = 2 \cdot k$ and $b_2 = 0$, 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, $\delta = 0$.³²

Step 1: For any $b_1 \geq 2 \cdot k$, Division 2 will secure $k_{22} = 0$ in the second period, and Division 1 will set $b_2 = 0$.

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

$$R_1(b_1 - k_{22} + b_2, \theta, \bar{\varepsilon}) - c \cdot b_2,$$

subject to the constraints $b_2 \geq 0$ and $b_1 - k_{22} + b_2 \geq 0$. We note that the charges corresponding to b_1 are sunk costs. Division 2's second period capacity demand induces the following optimal response from Division 1:

³² This specification does simplify the algebra considerably yet, as will become clear below, it is in no way essential for the following argument.

$$b_2(k_{22}, b_1) = \begin{cases} 0 & \text{if } k_{22} \leq b_1 - k \\ k_{22} + k - b_1 & \text{if } k_{22} \geq b_1 - k. \end{cases} \tag{26}$$

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

$$\Gamma(k_{22}, \theta, \bar{\epsilon}) = M_f(b_1, \theta, \bar{\epsilon}) - R_1(b_1 - k_{22}, \theta, \bar{\epsilon}) - c \cdot k_{22}$$

for any $k_{22} \leq b_1 - k$. Thus, we find that $\Gamma'(k_{22}, \theta, \bar{\epsilon}) = R'_1(b_1 - k_{22}, \theta, \bar{\epsilon}) - c \leq R'_1(k, \theta, \bar{\epsilon}) - c < 0$ for all $k_{22} \leq b_1 - k$. For any $k_{22} \geq b_1 - k$ the downstream division's payoff is

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

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 $b_1 \geq 2 \cdot k$, Division 2's second period payoff is increasing in b_1 .

From Step 1 we know that neither division will obtain additional capacity rights if $b_1 \geq 2 \cdot k$. As a consequence, Division 2's payoff becomes

$$\Gamma(0, b_1, \theta, \bar{\epsilon}) = M_f(b_1, \theta, \bar{\epsilon}) - R_1(b_1, \theta, \bar{\epsilon}).$$

This expression is increasing in b_1 because

$$\frac{\partial}{\partial b_1} \Gamma(0, b_1, \theta, \bar{\epsilon}) = R'_1(q_1^*(b_1, \theta, \bar{\epsilon}), \theta, \bar{\epsilon}) - R'_1(b_1, \theta, \bar{\epsilon}) > 0,$$

as $q_1^*(b_1, \theta, \bar{\epsilon}) > 0$. 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_1 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

$$E_\epsilon[S(k, \theta_{t+1}, \epsilon_{t+1})] \geq E_\epsilon[S(k, \theta_t, \epsilon_t)]. \tag{27}$$

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

$$S(k, \theta_{t+1}, \epsilon) \geq S(k, \theta_t, \epsilon). \tag{28}$$

Suppose first $0 < q_1^*(k, \theta_{t+1}, \epsilon) \leq q_1^*(k, \theta_t, \epsilon) < k$. Since $R'_1(q, \theta_{1t}, \epsilon_{1t})$ is increasing in θ_{1t} and $\theta_{1,t+1} \geq \theta_{1t}$, the definition of the shadow price in (11) implies the inequality in (28). Suppose now $q_1^*(k, \theta_{t+1}, \epsilon) \geq q_1^*(k, \theta_t, \epsilon)$. Since the shadow price can be expressed as

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

$\theta_{2,t+1} \geq \theta_{2t}$ and $R'_2(q, \theta_{2t}, \epsilon_{2t})$ is increasing in θ_{2t} , we conclude that (28) 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

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

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

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

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

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

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 $b_t \geq 0$ will not be binding. \square

Proof of Proposition 3 The proof of Proposition 5 below 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:

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

and

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

Furthermore,

$$S(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t) = c = R'_i(\bar{k}_{it}^o, \theta_{it}, \bar{\epsilon}_{it}).$$

As a consequence, the divisional first-order conditions:

$$(1 - \delta) \cdot R'_1(\bar{k}_{1t}^o, \theta_{1t}, \bar{\epsilon}_{1t}) + \delta \cdot S(\bar{k}_{1t}^o + \bar{k}_{2t}^o, \theta_t, \bar{\epsilon}_t) = c \tag{29}$$

and

$$\delta \cdot R'_2(\bar{k}_{2t}^o, \theta_{2t}, \bar{\epsilon}_{2t}) + (1 - \delta) \cdot S(\bar{k}_{1t}^o + \bar{k}_{2t}^o, \theta_t, \bar{\epsilon}_t) = c \tag{30}$$

are met at $(\bar{k}_{1t}^o, \bar{k}_{2t}^o)$. Since, each division’s objective function is globally concave in its choice variable k_{it} , the pair $(\bar{k}_{1t}^o, \bar{k}_{2t}^o)$ constitutes a Nash equilibrium. Finally, it is straightforward to check that there cannot be any other pure strategy equilibria. \square

Proof of Proposition 4 Conditional on history h_t and state information θ_t , Division i chooses b_{it} in period t to maximize:

$$\sum_{\tau=t}^T u_{it} \cdot E[\pi_{it}(b_{it}, b_{jt} | h_t, \theta_t)],$$

where $\pi_{it}(b_{it}, b_{jt} | h_t, \theta_t)$ is as given in (19) and u_{it} is the weight that Division i attaches to its performance measure in period t . Given that each divisions begins period 1

with zero capacity, i.e., $h_1 = (0, 0)$, it can be readily verified that strategy $b^* \equiv \{b_1^*(h_1, \theta_1), \dots, b_T^*(h_T, \theta_T)\}$ with:

$$b_{it}^*(h_t, \theta_t) = \begin{cases} 0 & \text{for } h_{it} \geq \bar{k}_{it}^o(\theta_t) \text{ and all } h_{jt} \geq 0, \\ 0 & \text{for } h_{it} < \bar{k}_{it}^o(\theta_t) \text{ and } h_{jt} \geq h_{jt}^+, \\ b_{it}^+ & \text{for } h_{it} < \bar{k}_{it}^o(\theta_t) \text{ and } h_{jt} \in (\bar{k}_{jt}^o(\theta_t), h_{jt}^+), \\ \bar{k}_{it}^o(\theta_t) - h_{it} & \text{for } h_{it} < \bar{k}_{it}^o(\theta_t) \text{ and } h_{jt} \leq \bar{k}_{jt}^o(\theta_t), \end{cases} \quad (31)$$

leads to efficient capacity levels in each period. To prove the proposition, we will therefore show that b^* is a subgame perfect equilibrium strategy for arbitrary weights $u_i > 0$. We first prove the following property of b^* :

Step 1: In any given period t , suppose each division chooses b_{it} to myopically maximize its current payoff $\pi_{it}(b_{it}, b_{jt}|h_t, \theta_t)$. Then $b_t^*(h_t, \theta_t)$ in (31) constitutes a Nash equilibrium strategy in the resulting one-stage game for any $h_t \in \mathbb{R}_+^2$. \square

Proof If b_{it} maximizes $\pi_{it}(b_{it}, b_{jt}|h_t, \theta_t)$, then the following first-order condition must hold:

$$(1 - \delta_i) \cdot R'_i(h_{it} + b_{it}, \theta_{it}, \bar{e}_i) + \delta_i \cdot S(h_{it} + b_{it} + h_{jt} + b_{jt}, \theta_t, \bar{e}_t) - c \leq 0. \quad (32)$$

The above conditions will hold as an equality whenever the corresponding $b_{it} > 0$. Since each division's payoff function is globally concave in its choice variable b_{it} , the first-order conditions in (32) are necessary as well as sufficient.

Case I: $h_{it} \geq \bar{k}_{it}^o$ for each i .

In this case, the first-order conditions in (32) hold simultaneously at $b_{1t} = b_{2t} = 0$, and hence $(0, 0)$ is a Nash equilibrium strategy.

Case II: $h_{it} \leq \bar{k}_{it}^o$ for each i .

In this case, it can be readily verified from the first-order conditions in (32) that $(\bar{k}_{1t}^o - h_{1t}, \bar{k}_{2t}^o - h_{2t})$ is a Nash equilibrium strategy.

Case III: $h_{it} \leq \bar{k}_{it}^o$ and $h_{jt} \in (\bar{k}_{jt}^o, h_{jt}^+)$.

We need to show that $b_{it} = b_{it}^+$ and $b_{jt} = 0$ is a Nash equilibrium strategy, where b_{it}^+ is as given by (21). Given that $h_{jt} < h_{jt}^+$, b_{it}^+ is an optimal response to $b_{jt} = 0$ by definition of b_{it}^+ in (21). We now claim that $h_{it} + b_{it}^+ < \bar{k}_{it}^o$. To the contrary, suppose $h_{it} + b_{it}^+ \geq \bar{k}_{it}^o$. Then

$$R'_i(h_{it} + b_{it}^+, \theta_{it}, \bar{e}_i) \leq c$$

and

$$S(h_{it} + b_{it}^+ + h_{jt}, \theta_t, \bar{e}_t) < S(\bar{k}_{it}^o + k_{jt}^o, \theta_t, \bar{e}_t) = c$$

since $h_{jt} > \bar{k}_{jt}^o$. But then the first-order condition in (21) cannot hold. This proves the claim that $h_{it} + b_{it}^+ < \bar{k}_{it}^o$. Since $h_{it} + b_{it}^+ < \bar{k}_{it}^o$, we have $R'_i(h_{it} + b_{it}^+, \theta_{it}, \bar{e}_i) > c$. Therefore, equation (32) implies that

$$S(h_{it} + b_{it}^+ + h_{jt}, \theta_t, \bar{e}_t) < c.$$

Given the above inequality and the fact that $R'_j(h_{jt}, \theta_{jt}, \bar{e}_j) < c$, it follows that $b_{jt} = 0$ is the best response to b_{it}^+ because

$$(1 - \delta_j) \cdot R'_j(h_{jt}, \theta_{jt}, \bar{e}_{jt}) + \delta_j \cdot S(h_{it} + b_{it}^+ + h_{jt}, \theta_t, \bar{e}_t) < c.$$

Case IV: $h_{it} \leq \bar{k}_{it}^o$ and $h_{jt} \geq h_{jt}^+$.

In this case, we need to show that $(0, 0)$ is a Nash equilibrium. Since $h_{jt} \geq h_{jt}^+$, equation (20) implies that $b_{it} = 0$ is Division i 's best response to $b_{jt} = 0$. Furthermore, since $h_{it} \leq \bar{k}_{it}^o$, we have $R'_i(h_{it}, \theta_{it}, \bar{e}_{it}) \geq c$, and therefore (32) implies that:

$$S(h_{it} + h_{jt}, \theta_t, \bar{e}_t) \leq c.$$

Given the above inequality and the fact that $R'_j(h_{jt}, \theta_{jt}, \bar{e}_{jt}) < R'_j(\bar{k}_{jt}^o, \theta_{jt}, \bar{e}_{jt}) = c$, it follows that

$$(1 - \delta_j) \cdot R'_j(h_{jt}, \theta_{jt}, \bar{e}_{jt}) + \delta_j \cdot S(h_{it} + h_{jt}, \theta_t, \bar{e}_t) < c.$$

Therefore, $b_{jt} = 0$ is a best response to $b_{it} = 0$. This completes the proof of Step 1.

Since the two divisions play a finite multi-stage game with observed actions, the one-stage deviation principle of Theorem 4.1 in Fudenberg and Tirole (1991) applies. To prove that b^* is a subgame perfect equilibrium strategy, it thus suffices to show that there is no division i and no strategy \hat{b}_i that agrees with b^* except in a single period such that \hat{b}_i is a better response to b_j^* than b_i^* . Since the two divisions are symmetrical, to prove this result, we will show that Division 1 cannot benefit from any unilateral one-stage deviation.

Suppose the two divisions play the strategy as specified by b^* in (31) for each of the first $t - 2$ periods. Let $h_{t-1}^* \equiv (\beta \cdot b_{1,t-2}^*, \beta \cdot b_{2,t-2}^*)$ denote the resulting history at the beginning of period $t - 1$. We note that each division will begin period $t - 1$ with less than $\bar{k}_{i,t-1}^o$ units of capacity, i.e., $h_{i,t-1}^* < \bar{k}_{i,t-1}^o$. In period $t - 1$, suppose Division 2 plays $b_{2,t-1}^*(h_{t-1}, \theta_{t-1})$, but Division 1 deviates to $\hat{b}_{1,t-1}$. Let $\hat{h}_t \equiv (\hat{h}_{1t}, \hat{h}_{2t})$ with

$$\hat{h}_{1t} = \beta \cdot \hat{b}_{1,t-1},$$

and

$$\hat{h}_{2t} = \beta \cdot b_{2,t-1}^*(h_{t-1}^*, \theta_{t-1})$$

denote the history generated by these choices. In subsequent periods, both divisions revert back to strategy b^* .

We now claim that this one-stage deviation by Division 1 affects its payoffs, and of Division 2, only in periods $t - 1$ and t . To prove this claim, we first note that any capacity acquired in period $t - 1$ is gone by the end of period t . Furthermore, strategy profile b^* has the property that $b_{it}^*(h_t, \theta_t) \leq \bar{k}_{it}^o$ for each i and any history h_t . Since $\bar{k}_{i,t+1}^o \geq \bar{k}_{it}^o$, a one-stage deviation in period $t - 1$ will never cause either of the two divisions to acquire more than $\bar{k}_{i,t+1}^o$ units of capacity in period t . Consequently, following a one-stage deviation in period $t - 1$, continuation of strategy b^* from period $t + 1$ onwards will generate the same outcomes in all subsequent periods as when strategy b^* is followed during the entire game.

To prove that b^* is a subgame perfect equilibrium strategy, we will therefore show that for all $u_{1,t-1} \geq 0$ and all $u_{1t} \geq 0$,

$$\begin{aligned}
 & u_{1,t-1} \cdot E \left[\pi_{1,t-1}(\hat{b}_{1,t-1}, b_{2,t-1}^*(h_{t-1}^*, \theta_{t-1})) \right] + u_{1t} \cdot E \left[\pi_{1t}(b_{1t}^*(\hat{h}_t, \theta_t), b_{2t}^*(\hat{h}_t, \theta_t)) \right] \\
 & \leq \sum_{\tau=t-1}^t u_{1\tau} \cdot E \left[\pi_{1\tau}(b_{1\tau}^*(h_{\tau}^*, \theta_{\tau}), b_{2\tau}^*(h_{\tau}^*, \theta_{\tau})) \right].
 \end{aligned}$$

To show that the above inequality holds for all $u_{1,t-1} \geq 0$ and all $u_{1t} \geq 0$, we prove that it holds on a period-by-period basis, that is,

$$\pi_{1,t-1}(\hat{b}_{1,t-1}, b_{2,t-1}^*(h_{t-1}^*, \theta_{t-1})) \leq \pi_{1,t-1}(b_{1,t-1}^*(h_{t-1}^*, \theta_{t-1}), b_{2,t-1}^*(h_{t-1}^*, \theta_{t-1})), \tag{33}$$

and

$$\pi_{1t}(b_{1t}^*(\hat{h}_t, \theta_t), b_{2t}^*(\hat{h}_t, \theta_t)) \leq \pi_{1t}(b_{1t}^*(h_t^*, \theta_t), b_{2t}^*(h_t^*, \theta_t)). \tag{34}$$

for each θ_{t-1} and θ_t . The inequality in (33) holds, since Step 1 shows that $b_{1,t-1}^*(h_{t-1}^*, \theta_{t-1})$ is a best response to $b_{2,t-1}^*(h_{t-1}^*, \theta_{t-1})$ in a hypothetical single-stage game in which divisions choose their period $t - 1$ capacities to myopically maximize their current payoffs.

To establish inequality (34), we prove the following result:

Step 2: $\frac{\partial \pi_{1t}^*(\hat{h}_{1t}|\theta_t)}{\partial \hat{h}_{1t}} \leq 0$ for all θ_t , where $\pi_{1t}^*(\hat{h}_{1t}|\theta_t) \equiv \pi_{1t}(b_{1t}^*(\hat{h}_{1t}, \theta_t), b_{2t}^*(\hat{h}_{1t}, \theta_t))$.

Proof Note that $\hat{h}_{2t} \leq \bar{k}_{2t}^o$. Therefore, if $\hat{h}_{1t} \leq \bar{k}_{1t}^o$, then Division 1 will acquire $\bar{k}_{1t}^o - \hat{h}_{1t}$ number of units in period t and hence $\frac{\partial \pi_{1t}^*(\hat{h}_{1t}|\theta_t)}{\partial \hat{h}_{1t}} = 0$.

Consider now the case when $\hat{h}_{1t} > \bar{k}_{1t}^o$. Differentiating $\pi_{1t}^*(\hat{h}_{1t}|\theta_t)$ with respect to \hat{h}_{1t} yields

$$\frac{\partial \pi_{1t}^*(\hat{h}_{1t}|\theta_t)}{\partial \hat{h}_{1t}} = \Delta + \delta \cdot [S - R_2'] \cdot \frac{\partial b_{2t}^*(\hat{h}_{1t}, \theta_t)}{\partial \hat{h}_{1t}}, \tag{35}$$

where, for brevity, we have suppressed the arguments of S and R_2' , and

$$\Delta \equiv (1 - \delta) \cdot R_1'(\hat{h}_{1t}, \theta_{1t}, \bar{\epsilon}_{1t}) + \delta \cdot S(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^*, \theta_t, \bar{\epsilon}_t) - c. \tag{36}$$

We note from Step 1 that $(0, b_{2t}^*)$ is a Nash equilibrium in the single-stage game in period t . Consequently, Division 1's first-order condition in (32) applies and therefore $\Delta \leq 0$. If $\hat{h}_{1t} \geq h_{1t}^+$, then $b_{2t}^* = 0$ and hence (35) implies that $\frac{\partial \pi_{1t}^*}{\partial \hat{h}_{1t}} \leq 0$ because $\Delta \leq 0$ and $\frac{\partial b_{2t}^*}{\partial \hat{h}_{1t}} = 0$. Consider now the case $\hat{h}_{1t} \in (\bar{k}_{1t}^o, h_{1t}^+)$ and $b_{2t}^* = b_{2t}^+ > 0$. Recall that b_{2t}^+ is defined by equation (21), which yields

$$\frac{\partial b_{2t}^+}{\partial \hat{h}_{1t}} = \frac{-(1 - \delta) \cdot S'}{\delta \cdot R_2'' + (1 - \delta) \cdot S'} \in (-1, 0).$$

Equation (21) can be rearranged to yield

$$S - c = \delta \cdot (S - R_2').$$

Substituting this into (35) gives

$$\frac{\partial \pi_{1t}^*(\hat{h}_{1t}|\theta_t)}{\partial \hat{h}_{1t}} = \Delta + [S - c] \cdot \frac{\partial b_{2t}^+}{\partial \hat{h}_{1t}}.$$

Lemma A2 follows from the above equation if $S > c$. Consider now the case when $S \leq c$. Substituting for Δ from (36) and simplifying yields

$$\begin{aligned} \frac{\partial \pi_{1t}^*(\hat{h}_{1t}|\theta_t)}{\partial \hat{h}_{1t}} = & (1 - \delta) \cdot \left[R'_1(\hat{h}_{1t}, \theta_{1t}, \bar{\epsilon}_{1t}) - S(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t) \right] \\ & + [S(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t) - c] \cdot \left[1 + \frac{\partial b_{2t}^+}{\partial \hat{h}_{1t}} \right]. \end{aligned}$$

The above expression is negative because

- (i) $S(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t) \leq c$ by assumption,
- (ii) $1 + \frac{\partial b_{2t}^+}{\partial \hat{h}_{1t}} \geq 0$, and
- (iii) $R'_1(\hat{h}_{1t}, \theta_{1t}, \bar{\epsilon}_{1t}) < S(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t)$.

To see why inequality (iii) holds, we recall that

$$S(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t) = R'_1(q_{1t}^*(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t), \theta_{1t}, \bar{\epsilon}_{1t}).$$

Furthermore, since $\hat{h}_{1t} > \bar{k}_{1t}^o$,

$$R'_1(\hat{h}_{1t}, \theta_{1t}, \bar{\epsilon}_{1t}) < c.$$

The proof of Step 1 showed that $\hat{h}_{2t} + b_{2t}^+ < \bar{k}_{2t}^o$, and therefore

$$R'_2(\hat{h}_{2t} + b_{2t}^+, \theta_{2t}, \bar{\epsilon}_{2t}) > c.$$

The above two inequalities imply that subsequent to the realization of period t demand shock, the total capacity will be reallocated such that Division 1 transfers some of its capacity to Division 2 when the realized shock satisfies $\epsilon_t = \bar{\epsilon}_t$. That is,

$$q_{1t}^*(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t) < \hat{h}_{1t}.$$

This implies that

$$R'_1(q_{1t}^*(\cdot), \theta_{1t}, \bar{\epsilon}_{1t}) \equiv S(\hat{h}_{1t} + \hat{h}_{2t} + b_{2t}^+, \theta_t, \bar{\epsilon}_t) > R'_1(\hat{h}_{1t}, \theta_{1t}, \bar{\epsilon}_{1t}),$$

and therefore inequality (iii) holds. This completes the proof of Step 2 and Proposition 4. □

Proof of Proposition 5 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:

$$S(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t) = c. \tag{37}$$

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

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

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

$$E_\epsilon [S(k_t^o, \theta_t, \epsilon_t)] = c. \tag{38}$$

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 (37) and (38) 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:

$$E_\epsilon [S(\bar{k}_t^o, \theta_t, \epsilon_t)] < S(\bar{k}_t^o, \theta_t, \bar{\epsilon}_t) = c. \tag{39}$$

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. \square

Proof of Proposition 6 We first note that a pure strategy Nash equilibrium (k_{1t}^*, k_{2t}^*) always exists. To see this, let $f_{it}(k_{jt})$ denote the best response by Division i if it conjectures that Division j chooses k_{jt} . For convenience, we drop the subscript t . Let $\bar{k}_i = f_i(0)$. The functions $f_i(\cdot)$, defined on $[0, \infty)$, are continuous and monotonically decreasing, because each party’s payoff is concave in its own capacity level. Consider now the function

$$\Delta(k_2) \equiv f_1(k_2) - f_2^{-1}(k_2), \tag{40}$$

defined on the interval $[0, \bar{k}_2]$. We have: $\Delta(0) \leq 0$ because $f_2(\bar{k}_1) \geq 1$ and $f_2(\cdot)$ is (weakly) decreasing. Furthermore $\Delta(\bar{k}_2) \geq 0$ because $f_1(\bar{k}_2) \geq 0$, yet $f_2(0) = \bar{k}_2$. Since $\Delta(\cdot)$ is continuous, the Intermediate Value Theorem ensures the existence of a value k_2^* such that $\Delta(k_2^*) = 0$. Letting $k_1^* = f_1(k_2^*)$, we conclude that (k_1^*, k_2^*) is a Nash equilibrium.

Suppose now the shadow price $S(k_t, \theta_t, \epsilon_t)$ is a concave function of ϵ_t , and hence $k_t^o < \bar{k}_t^o$. Suppose also, the contrapositive of the claim, that (k_{1t}^*, k_{2t}^*) induces under-investment in the sense that $k_t^* \equiv k_{1t}^* + k_{2t}^* \leq k_t^o$. Since $S(k, \cdot, \cdot)$ is a decreasing function, this implies:

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

Suppose first that $(k_{1t}^*, k_{2t}^*) > 0$. The first-order conditions for a Nash equilibrium then yield:

$$(1 - \delta) \cdot R'_1(k_{1t}^*, \theta_{1t}, \bar{\epsilon}_{1t}) + \delta \cdot E_\epsilon [S(k_t^*, \theta_t, \epsilon_t)] = c \tag{41}$$

$$\delta \cdot R'_2(k_{2t}^*, \theta_{2t}, \bar{\epsilon}_{2t}) + (1 - \delta) \cdot E_\epsilon [S(k_t^*, \theta_t, \epsilon_t)] = c. \tag{42}$$

In order for these conditions to be met, the following inequalities would have to hold:

$$R'_1(k_{1t}^*, \theta_{1t}, \bar{\epsilon}_{1t}) \leq c = R'_1(\bar{k}_{1t}^o, \theta_{1t}, \bar{\epsilon}_{1t})$$

and

$$R'_2(k_{2t}^*, \theta_{2t}, \bar{e}_{2t}) \leq c = R'_2(\bar{k}_{2t}^o, \theta_{2t}, \bar{e}_{2t}).$$

That, however, would contradict the hypothesis:

$$k_{1t}^* + k_{2t}^* \leq k_t^o < \bar{k}_t^o = \bar{k}_{1t}^o + \bar{k}_{2t}^o.$$

If the Nash equilibrium (k_{1t}^*, k_{2t}^*) , involves a boundary value, say $k_{1t}^* = 0$, the first-order condition in (41) is replaced by:

$$(1 - \delta) \cdot R'_1(0, \theta_{1t}, \bar{e}_{1t}) + \delta \cdot E_\epsilon [S(k_t^*, \theta_t, \epsilon_t)] \leq c.$$

But that is impossible as well since $E_\epsilon [S(k_t^*, \theta_t, \epsilon_t)] = c$ and $R'_1(0, \theta_{1t}, \bar{e}_{1t}) > c$. A parallel argument shows that there will be under-investment when $S(\cdot, \cdot, \epsilon_t)$ is convex in ϵ_t . □

Proof of Proposition 7 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:

$$(1 - \delta) \cdot E_\epsilon [R_1(k_{1t}, \theta_{1t}, \epsilon_{1t})] + \delta \cdot E_\epsilon [M_f(k_t^o + k_{1t} - \bar{k}_{1t}^o, \theta_t, \epsilon_t) - R_2(k_t^o - \bar{k}_{1t}^o, \theta_{2t}, \epsilon_{2t})] - c \cdot (k_t^o + k_{1t} - \bar{k}_{1t}^o). \tag{43}$$

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

$$E_\epsilon [(1 - \delta) \cdot R'_1(k_{1t}, \theta_{1t}, \epsilon_{1t}) + \delta \cdot S(k_t^o + k_{1t} - \bar{k}_{1t}^o, \theta_t, \epsilon_t)] = c,$$

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 fail 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

$$\hat{\pi}_{1t} = E_\epsilon [(1 - \delta) \cdot R_1(\hat{k}_t, \theta_{1t}, \epsilon_{1t}) + \delta \cdot M_f(\hat{k}_t, \theta_t, \epsilon_t)] - c \cdot \hat{k}_t, \tag{44}$$

while Division 2’s default payoff is

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

By agreeing to transfer $k_t^o - \bar{k}_{1t}^o$ units of capacity rights to Division 2, the two divisions can increase 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

$$\begin{aligned} E_c & \left[(1 - \delta) \cdot R_1(\bar{k}_{1t}^o, \theta_{1t}, \epsilon_{1t}) + \delta \cdot [M_f(k_t^o, \theta_t, \epsilon_t) - R_2(k_t^o - \bar{k}_{1t}^o, \theta_{2t}, \epsilon_{2t})] \right] + TP_t \\ & = \hat{\pi}_{1t} + \Delta \cdot \delta M. \end{aligned}$$

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. \square

References

- Alles, M., & Datar, S. (1998). Strategic transfer pricing. *Management Science*, 44, 451–461.
- Anctil, R., & Dutta, S. (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*, 16, 21–30.
- Arya, A., & Mittendorf, B. (2008). Pricing internal trade to get a leg up on external rivals. *Journal of Economics and Management Strategy*, 17(3), 709–731.
- Baiman, S., & Rajan, M. (1995). Centralization, delegation, and shared responsibility in the assignment of capital investment decision rights. *Journal of Accounting Research*, 33(Suppl), 135–164.
- Balakrishnan, R., & Sivaramakrishnan, S. (2002). A critical overview of full-cost data for planning and pricing. *Journal of Management Accounting Research*, 3–31.
- Baldenius, T., Dutta, S., & Reichelstein, S. (2007). Cost allocations for capital budgeting decisions. *The Accounting Review*, 82(4), 837–867.
- Baldenius, T., Reichelstein, S., & Sahay, S. (1999). Negotiated versus cost-based transfer pricing. *The Review of Accounting Studies*, 4(2), 67–91.
- Banker, R., & Hughes, J. (1994). Product costing and pricing. *The Accounting Review*, 69(3), 479–494.
- Bastian, N., & Reichelstein, S. (2004). *Transfer pricing at Timken*. Stanford GSB, Case # A-190.
- Bolton, P., & Dewatripont, M. (2005). *Contract theory*. Cambridge, MA: MIT Press.
- Bouwens, J., & Steens, B. (2008). *The economics of full-cost transfer pricing*, Working Paper, Tilburg University.
- Christensen, P., Feltham, G., & Wu, M. (2002). 'Cost of capital' in residual income measurement under moral hazard. *The Accounting Review*, 77(1), 1–23.
- Dutta, S., & Reichelstein, S. (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., & White, H. (1988). Price and authority in inter-profit center transactions. *American Journal of Sociology*, 94(Supplement), 17–51.
- Edlin, A., & Reichelstein, S. (1995). Negotiated transfer pricing: An efficiency result. *The Accounting Review*, 69(3), 479–494.
- Ernst & Young (2003). *Global transfer pricing survey*. <http://www.ey.com>.
- Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: MIT Press.
- Goex, R. (2002). Capacity planning and pricing under uncertainty. *Journal of Management Accounting Research*, 59–79.
- Hotelling, H. (1925). A general mathematical theory of depreciation. *Journal of the American Statistical Association*, 20, 340–353.
- Hughes, J., & Kao, J. (1997). Cross subsidization, cost allocation and tacit collusion. *Review of Accounting Studies*, 2(3), 265–293.
- Kouvelis, P., & Lariviere, M. (2000). Decentralizing cross-functional decisions: Coordination through internal markets. *Management Science*, 46(8), 1049–1058.
- Malone, T. (2004). Bringing the market inside. *Harvard Business Review* April, 1–9.

- Martinez-Jerez, A. (2007). Understanding customer profitability at Charles Schwab, Harvard Business School, Case # 106-002.
- Pfeiffer, T., & Schneider, G. (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., & Wagner J. (2009). Cost-based transfer pricing, Working paper, University of Vienna.
- Plambeck, E., & Taylor, T. (2005). Sell the plant: The impact of contract manufacturing on innovation, capacity and profitability. *Management Science*, 51, 133–150.
- Porteus, E., & Wang, S. (1991). On manufacturing/marketing incentives. *Management Science*, 37(9), 1166–1181.
- Rajan, M., & Reichelstein, S. (2009). Depreciation rules and the relation between marginal and historical cost. *Journal of Accounting Research*, 47(3), 1–43.
- 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*, 116, 931–950.
- Sahay, S. (2002). Transfer pricing based on actual cost. *Journal of Management Accounting Research*, 15, 177–192.
- Tang, R. (2002). *Current trends and corporate cases in transfer pricing*. Westport, CT: Quorum Books.
- Mieghem, V. J. (2003). Capacity management, investment and hedging: Review and recent developments. *Manufacturing and Service Operation Management*, 5(4), 269–302.
- Wei, D. (2004). Interdepartmental cost allocation and investment incentives. *Review of Accounting Studies*, 9, 97–116.
- Wielenberg, S. (2000). Negotiated transfer pricing, specific investment and optimal capacity choice. *Review of Accounting Studies*, 5, 197–216.
- Zimmerman, J. (1979). The costs and benefits of cost allocations. *The Accounting Review*, 54, 504–521.