

# Capacity Rights and Full Cost Transfer Pricing

**Sunil Dutta**

Haas School of Business  
University of California, Berkeley

and

**Stefan Reichelstein\***

Graduate School of Business  
Stanford University

February 2018

---

\*We are grateful to Steven Mitsuda, Anna Rohlfing-Bastian and seminar participants at the Verein fuer Socialpolitik (Ausschuss fuer Unternehmenstheorie) and Columbia University (Burton Workshop) for helpful comments and suggestions. Sunil Dutta acknowledges research support provided by the Indian School of Business.

## **Abstract: Capacity Rights and Full Cost Transfer Pricing**

This paper examines the theoretical properties of the practice of full cost transfer pricing in multi-divisional firms. In our model of a multi-divisional firm, divisional managers are responsible for the initial acquisition of productive capacity as well as its utilization in subsequent periods, once operational uncertainty has been resolved. We refer to a transfer pricing rule as a full cost rule if the discounted sum of transfer payments is equal to the initial capacity acquisition cost and the present value of all subsequent variable costs of output supplied to a division. Our analysis identifies environments where a suitable variant of full cost transfer pricing induces efficiency in both the initial investments and the subsequent output levels. Our study also highlights the need for a proper integration of the divisional control rights over capacity investments and the valuation rules for intracompany transfers.

# 1 Introduction

The transfer of intermediate products and services across divisions of a firm is frequently valued at full cost. Surveys and textbooks consistently report that in contexts where a market-based approach is either infeasible or unreliable, cost-based transfer pricing is the most prevalent method for both internal managerial and tax reporting purposes.<sup>1</sup> At the same time, case studies and managerial accounting textbooks have pointed out consistently that full cost transfer pricing will frequently result in sub-optimal resource allocations. The objective of this paper is to investigate the incentive properties of full cost transfer pricing in multi-divisional firms. Specifically, we seek to identify environments in which full cost transfer pricing “works,” that is, it creates time-consistent incentives for divisional managers.<sup>2</sup>

A key feature of our model is that divisional managers are responsible for initial acquisition of productive capacity as well as its subsequent utilization in future periods after resolution of demand uncertainty. We seek to characterize transfer pricing mechanisms that induce divisional managers to make efficient capacity investment and utilization decisions. Our criterion for incentive compatibility follows the literature on goal congruent performance measures such as Rogerson (1997), Dutta and Reichelstein (2002), Baldenius et al.(2007), and Nezlobin et al.(2015). Accordingly, the divisional performance measures must in any particular time period be congruent with the objective of maximizing firm value. Put differently, regardless of the managers’ planning horizons and intertemporal preferences, a goal congruent mechanism must induce (i) the efficient levels of capacity investments upfront, and (ii) the efficient production quantities in subsequent time periods after the resolution of revenue uncertainty in those periods.

---

<sup>1</sup>See, for instance, Eccles and White (1988), Ernst & Young (1993), Tang (2002), Feinschreiber and Kent (2012), Datar and Rajan (2014), and Zimmerman (2016).

<sup>2</sup>The perspective in this paper is similar to that underlying the literature on the use of full cost measures for pricing and capacity expansion decisions. See, for example, Banker and Hughes (1994), Balachandran et al.(1997), Goex (2002), Balakrishnan and Sivaramakrishnan (2002), Gramlich and Ray (2016), and Reichelstein and Sahoo (2018). While these studies examine the role of full cost from a central planning perspective, our focus is on decentralization and management control.

Numerous theoretical and empirical studies have examined the performance of cost-based transfer pricing.<sup>3</sup> Among these studies, Dutta and Reichelstein (2010) is structurally closest to the analysis in this paper. Their findings identify conditions under which full cost transfer pricing will lead to efficient outcomes. However, while capacity investments are costly, there are no subsequent operating costs associated with producing output in their model. Unlike our analysis in this paper where it may be efficient not to exhaust the available capacity in bad states of the world, capacity is always fully utilized in Dutta and Reichelstein (2010). Their analysis thus abstracts away from one of the central points featured, for example, in the HBS case study “Polysar Limited” (Simons, 2000). A key takeaway from this case is that under full cost transfer pricing the buying division tends to reserve too much production capacity because demand for its product is uncertain and the internal pricing rule charges the division only for the share of full cost that pertains to the capacity actually utilized.

Our model considers two divisions that sell a product each in separate markets. Due to technical expertise, the upstream division installs and maintains all productive capacity. It also produces the output sold by the downstream division. For performance evaluation purposes, the upstream division is therefore viewed as an investment center, while the downstream division, having no capital assets, is merely a profit center. The periodic transfer payments from the upstream to the downstream division depend on the initial capacity choices and the current production levels. We refer to a transfer pricing rule as a full cost rule if the discounted sum of transfer payments is equal to the present value of cash outflows associated with the capacity assigned to the downstream division and all subsequent output services rendered to that division. In particular, a two-part pricing rule that charges in a lump sum fashion for capacity in each period in addition to variable charges, based on actual production volumes, will be considered a full-cost transfer price. Thus full cost transfer pricing

---

<sup>3</sup>A partial list of references includes Eccles and White (1988), Vaysman (1996), Baldenius et al. (1999), Sahay (2002), Goex and Schiller (2007), Pfeiffer et al.(2009), Baldenius (2008), and Bouwens and Steens (2016).

does not necessarily run into the problem of double marginalization that results from the buying division internalizing a unit charge based on cost components that are sunk (Datar and Rajan, 2014, and Zimmerman, 2016).

We distinguish two alternative scenarios depending on whether the divisions' products can share the same capacity assets. In the *dedicated capacity* scenario, the products require different productive assets, and hence the capacity cannot be shared across the divisions. Private information at the divisional level then makes it natural to give each division unilateral capacity rights. We identify production and information environments where a suitable variant of full cost transfer pricing induces efficient outcomes. Under certain conditions, the simplistic full cost transfer pricing rule featured in the Polysar case can be modified to obtain a goal congruent solution. Essential to this finding is that the buying division now also faces excess capacity charges.<sup>4</sup> While such excess capacity charges will not be imposed in equilibrium, the potential threat is sufficient to correct for the bias inherent in simplistic full cost transfer pricing.

In the scenario of dedicated capacity, we identify production and market environments where some variant of full cost transfer pricing induces efficient outcomes. We find the preferred transfer pricing rule varies depending on whether the value of capacity is expected to change over time and whether, given an efficient capacity choice in the first place, it will at times be advantageous to idle some of the available capacity.<sup>5</sup> Common to these pricing rules is that the fixed cost charges for capacity must be equal to what earlier literature has referred to as the "user cost of capital"<sup>6</sup>

---

<sup>4</sup>Our solution here is consistent with prescriptions in the managerial accounting literature on how to allocate the overhead costs associated with excess capacity, e.g., Kaplan (2006) and Martinez-Jerez (2007).

<sup>5</sup>The technical condition here will be referred to as the "limited volatility condition" which plays a central role in Reichelstein and Rohlfsing-Bastian (2015) in characterizing the relevant cost to be imputed for capacity expansion decisions.

<sup>6</sup>In contrast to our framework here, the derivation of the user cost of capital has been derived in models with overlapping investments in an infinite horizon setting, e.g., Arrow (1964), Carlton and Perloff (2005), Rogerson (2008, 2011), Rajan and Reichelstein (2009) and Reichelstein and Sahoo (2017).

For stationary environments in which the expected value of capacity remains constant over time, a standard two-part full cost transfer pricing rule will provide the downstream division with appropriate capacity investment incentives. At the same time, the periodic capacity cost charges do not interfere with the subsequent capacity utilization decisions.

When the two products in question can share the installed capacity, it suggests itself to allow the divisions to negotiate ex-post over the utilization of the available capacity. In such fungible capacity settings, the cost-based transfer price defines the parties' status quo payoffs in the subsequent negotiations. If the capacity acquisition decision were to be delegated to the upstream division in its role as an investment center, the resulting outcome would generally entail under-investment. The upstream division would then anticipate not earning the full expected return on its investment because gains from the optimized total contribution margin would be shared in the negotiation between the two divisions, when the initial acquisition cost would already be sunk.<sup>7</sup> Under certain conditions, we find that the coordination and hold-up problem associated with the initial capacity choice can be resolved by giving both divisions the unilateral right to reserve capacity, charging the downstream division for its capacity reservation by means of full cost transfer prices, and allowing the divisions to negotiate the actual use of the available capacity in subsequent time periods.

A coordination mechanism that works in a broader class of environments is obtained in the fungible capacity scenario if the downstream division must obtain approval from the investment center manager for any capacity it wants to reserve for its own use. The upstream division then becomes essentially a "gatekeeper" that will agree to let the downstream division reserve capacity for itself in exchange for a stream of lump-sum payments determined through initial negotiation. The upstream

---

<sup>7</sup>Even though investments are verifiable in our model, the hold-up problem that arises when only the upstream division makes capacity investments is essentially the same as in earlier incomplete contracting literature. One branch of that literature has explored how transfer pricing can alleviate hold-up problems when investments are "soft" (unverifiable); see, for example, Baldenius et al. (1999), Edlin and Reichelstein (1995), Sahay (2000), Baldenius (2008), and Pfeiffer et al. (2009).

division will thereafter have an incentive to invest in additional capacity on its own up to the efficient level. The resulting mechanism can be viewed as a hybrid between cost-based and negotiated transfer pricing rules such that the downstream division is charged the full cost of the total capacity acquired and total output produced.

Aside from the work of Dutta and Reichelstein (2010), this paper is closely related to Reichelstein and Rohlfsing-Bastian (2015). They examine the relevant cost measure for capacity investments in a centralized setting, but do not consider any performance evaluation and management control issues. Baldenius, Nezlobin and Vaysman (2016) is another precursor to the present paper insofar as they study managerial performance evaluation in a setting where capacity may remain idle in unfavorable states of the world. Their analysis, however, confines attention to a single division firm, and thus coordination and internal pricing issues do not arise in their model.

The remainder of the paper proceeds as follows. The basic model is described in Section 2. Section 3 examines a setting in which the divisions' products require different production facilities and therefore capacity is dedicated. Propositions 1 - 4 delineate environments in which full cost transfer pricing can induce the divisions to choose initial capacity levels and subsequent production levels that are efficient from the overall firm perspective. Section 4 considers the alternative arrangement in which capacity is fungible and can be traded across divisions. Propositions 5 and 6 demonstrate the need for allowing the downstream division to secure capacity rights for itself initially, even if the entire available capacity can be reallocated through negotiations in subsequent periods. We conclude in Section 5.

## **2 Model Description**

Consider a vertically integrated firm comprised of two divisions and a central office. Both divisions sell a marketable product (possibly a service) in separate and unrelated markets. In order for either division to deliver its product in subsequent periods, the firm needs to make upfront capacity investments. Because of technical expertise, only

the upstream division (Division 1) is in a position to install and maintain the productive capacity for both divisions. Division 1 also carries out the production for both divisions, and therefore incurs all periodic production costs.<sup>8</sup> Our analysis considers an organizational structure which views the upstream division as an investment center whose balance sheet reflects the historical cost of the initial capacity investments. In that sense, the upstream division acquires economic “ownership” of the capacity related assets.

Capacity could be measured either in hours or the amount of output produced. New capacity is acquired at time  $t = 0$ . Our analysis considers the two distinct scenarios of *dedicated* and *fungible* capacity. In the former scenario, the two products are sufficiently different so as to require separate production facilities. With fungible capacity, in contrast, both products can utilize the same capacity infrastructure. The upfront cash expenditure for one unit of capacity for Division  $i$  is  $v_i$  in the dedicated capacity setting. If Division  $i$  acquires  $k_i$  units of capacity, it has the option to produce up to  $k_i$  units of output in each of the next  $T$  periods.<sup>9</sup> In case of fungible capacity, the cost of acquiring one unit of capacity is  $v$ , which allows either division to produce one unit of output in each of the next  $T$  periods.

The actual production levels for Division  $i$  in period  $t$  are denoted by  $q_{it}$ . We assume that sales in each period are equal to the amount of production in that period; i.e., the divisions do not carry any inventory. Aside from requisite capacity resources, the delivery of one unit of output for Division  $i$  requires a unit variable cost of  $w_{it}$  in period  $t$ . These unit variable costs are anticipated upfront by the divisional managers with certainty, though they may become known and verifiable to the firm’s accounting system only when incurred in a particular period. The divisional contribution margins

---

<sup>8</sup>It is readily verified that our findings would be unchanged if the upstream division were to transfer an intermediate product which is then completed and turned into a final product by the downstream division.

<sup>9</sup>We thus assume that physical capacity does not diminish over time, but instead follows the “one-hoss shay” pattern, commonly used in the capital accumulation and regulation literature. See, for example, Rogerson (2008) and Nezlobin, Rajan and Reichelstein (2012).



are given by

$$CM_{it}(q_{it}, \epsilon_{it}) = x_{it} \cdot R_i(q_{it}, \epsilon_{it}) - w_{it} \cdot q_{it}.$$

The first term above,  $x_{it} \cdot R_i(q_{it}, \epsilon_{it})$ , denotes Division  $i$ 's revenues in period  $t$  with  $x_{it} \geq 0$  representing intertemporal parameters that allow for the possibility of declining, or possibly growing, revenues over time.

In addition to varying with the production quantities  $q_{it}$ , the periodic revenues are also subject to one-dimensional transitory shocks  $\epsilon_{it}$ . These random shocks are realized at the beginning of period  $t$  before the divisions choose their output levels for the current period, and prior to any capacity trades in the fungible capacity setting. We assume that the random shocks  $\epsilon_{it}$  are distributed according to density functions  $f_i(\cdot)$  with support on the interval  $[\underline{\epsilon}_i, \bar{\epsilon}_i]$ . The random variables  $\{\epsilon_{it}\}$  are also assumed to be independently distributed across time; i.e.,  $Cov(\epsilon_{it}, \epsilon_{i\tau}) = 0$  for each  $t \neq \tau$ , though they may be correlated across the two divisions; i.e., it is possible to have  $Cov(\epsilon_{1t}, \epsilon_{2t})$  to be non-zero in any given period  $t$ .

The exact shape of the revenue revenue functions,  $R_i(q_{it}, \epsilon_{it})$ , is private information of the divisional managers. These revenue functions are assumed to be increasing and concave in  $q_{it}$  for each  $i$  and each  $t$ . At the same time, the marginal revenue functions:

$$R'_i(q, \epsilon_{it}) \equiv \frac{\partial R_i(q, \epsilon_{it})}{\partial q}$$

are assumed to be increasing in  $\epsilon_{it}$ .

In any given period, the actual production quantity for a division may differ from its initial capacity rights for two reasons. First, for an unfavorable realization of the revenue shock  $\epsilon_{it}$ , a division may decide not to exhaust the entire available capacity because otherwise marginal revenues would not cover the incremental cost  $w_{it}$ . Second, in the case of fungible capacity, a division may want to yield some of its capacity rights to the other division if that division has a higher contribution margin.

Our model is in the tradition of the earlier goal congruence literature which does not explicitly address issues of moral hazard and managerial compensation. Instead the focus is on the choice of *goal congruent* performance measures for the divisions.

Accordingly, we assume that each divisional manager is evaluated by a performance measure  $\pi_{it}$  in each of the  $T$  time periods. The downstream division, which has only operational responsibilities for procuring and selling output, is treated as a profit center whose performance measure is measured by its divisional profit. In contrast, the upstream division, which also has control over capacity assets, is viewed as an investment center with *residual income* as its performance measure.<sup>10</sup> The remaining design variables of the internal managerial accounting system then consist of divisional capacity rights, depreciation schedules, and the transfer pricing rule.

Figure 1 illustrates the structure of the multi-divisional firm and its two constituent responsibility centers.

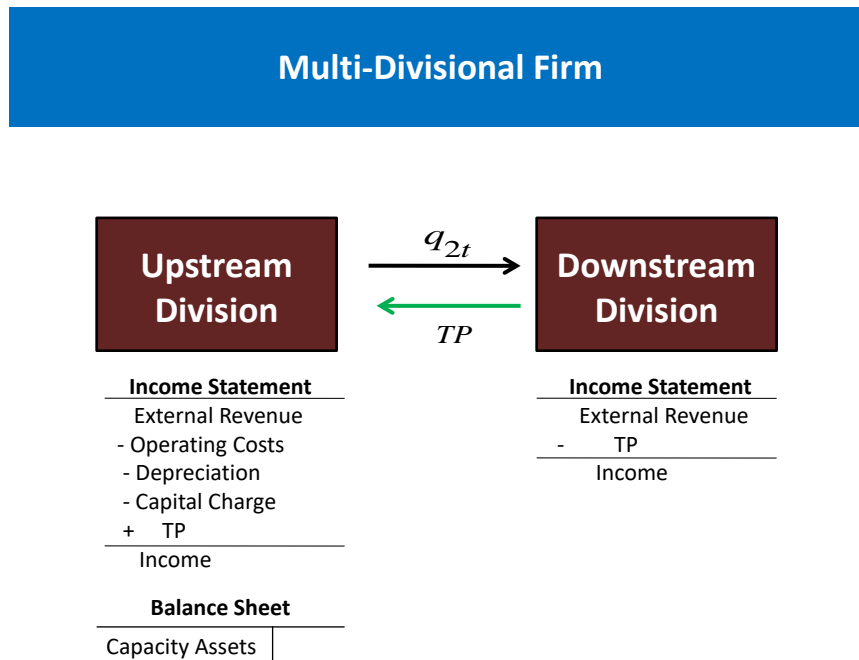


Figure 1: Divisional Structure of the Firm

<sup>10</sup>Earlier literature, including Reichelstein (1997), Dutta and Reichelstein (2002), and Baldenius et al.(2007), has argued that among a particular class of accounting based metrics only residual income can achieve the requisite goal congruence requirements.

The downstream division's performance measure (i.e., its operating income) in period  $t$  is given by

$$\pi_{2t} = Inc_{2t} = x_{2t} \cdot R_2(q_{2t}, \epsilon_{2t}) - TP_t(k_2, q_{2t}),$$

where  $TP_t(k_2, q_{2t})$  denotes the transfer payment to the upstream division in period  $t$  for securing  $k_2$  units of capacity and obtaining  $q_{2t}$  units of output. The residual income measure for the upstream division is given by

$$\pi_{1t} = Inc_{1t} - r \cdot BV_{t-1}, \quad (1)$$

where  $BV_t$  denotes book value of capacity 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 residual income measure in (1) depends on two accruals: the transfer price received from the downstream division and the depreciation charges corresponding to the initial capacity investments. Specifically,

$$Inc_{1t} = x_{1t} \cdot R_1(q_{1t}, \epsilon_{1t}) - w_{1t} \cdot q_{1t} - w_{2t} \cdot q_{2t} - D_t + TP_t(k_2, q_{2t}),$$

where  $D_t$  is the total depreciation expense in period  $t$ . Let  $d_{it}$  denote the depreciation charge in period  $t$  per dollar of initial capacity investment undertaken for Division  $i$ . Thus,

$$D_t = d_{1t} \cdot v_1 \cdot k_1 + d_{2t} \cdot v_2 \cdot k_2.$$

The depreciation schedules satisfy the usual tidiness requirement that  $\sum_{\tau=1}^T d_{i\tau} = 1$ ; i.e, the depreciation charges sum up to an asset's historical acquisition cost over its useful life. Book values evolve according to the simple iterative process:  $BV_t = BV_{t-1} - D_t$ , with  $BV_0 = v_1 \cdot k_1 + v_2 \cdot k_2$  and  $BV_T = 0$ .

Under the residual income measure, the overall *capital charge* imposed on the upstream division is the sum of depreciation charges plus imputed interest charges. Given the depreciation schedules  $\{d_{it}\}_{t=1}^T$ , the overall capital charge becomes:

$$D_t + r \cdot BV_{t-1} = z_{1t} \cdot v_1 \cdot k_1 + z_{2t} \cdot v_2 \cdot k_2, \quad (2)$$

where  $z_{it} \equiv d_{it} + r \cdot (1 - \sum_{\tau=1}^{t-1} d_{i\tau})$ . It is well known from the general properties of the residual income metric that regardless of the depreciation schedule, the present value of the  $z_{it}$  is equal to one; that is,  $\sum_{t=1}^T z_{it} \cdot \gamma^t = 1$  (Hotelling, 1925).

The manager of Division  $i$  is assumed to attach non-negative weights  $\{u_{it}\}_{i=1}^T$  to her performance measure in different time periods. The weights  $u_i = (u_{i1}, \dots, u_{iT})$  reflect both the manager's discount factor as well as the bonus coefficients attached to the periodic performance measures. 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 equilibrium decisions that maximize the net present value of firm-wide future cash flows. Consistent with the earlier literature, we impose the criterion of *strong goal congruence*, which requires that managers have incentives to make efficient production and investment decisions for any combination of the coefficients  $u_{it} \geq 0$ . Strong goal congruence requires that desirable managerial incentives must hold not only over the entire planning horizon, but also on a period-by-period basis. That is, each manager must have incentives to make efficient production and capacity decisions *even* if that manager were solely focused on maximizing her performance measure  $\pi_{i\tau}$  in any given single period  $\tau$ .<sup>11</sup>

The criterion of strong goal congruence can be applied with one of several alternative non-cooperative equilibrium concepts, e.g., dominant strategies or Nash equilibrium. An additional property identified in some of our subsequent results is the notion of a *separable* performance measure. A performance measure is said to be separable if it remains unaffected by the decisions made by the other manager. Clearly, separability can only be met if the divisions have dominant strategies.

---

<sup>11</sup>The concept of goal congruence dates back to the early work of Solomons (1964). Dutta (2008) identifies settings in which the accrual accounting rules that emerge as goal congruent are also part of optimal contracting arrangements in agency problems.

### 3 Dedicated Capacity

We first investigate a setting in which the divisional products require different capacity infrastructures. Since the divisional managers have private information about their future revenues, it is natural to consider an arrangement in which each division has unilateral rights to procure capacity for its own use. The analysis in this section focuses on identifying the depreciation schedules and transfer pricing rules that provide incentives for the divisional managers to choose efficient levels of capacity upfront and make optimal production decisions in subsequent periods. The following time line illustrates the sequence of events at the initial investment date and in a generic period  $t$ .

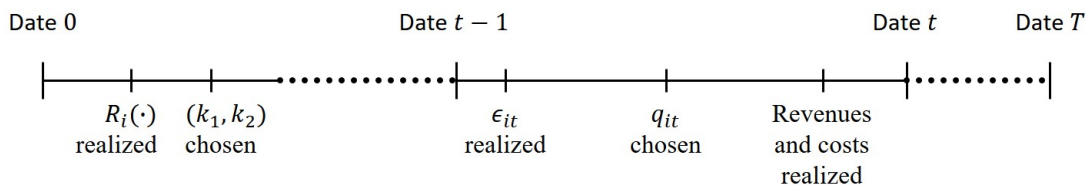


Figure 2: Sequence of Events in the Dedicated Capacity Scenario

If a central planner had full information regarding future revenues, the optimal investment decisions  $(k_1, k_2)$  would be chosen so as to maximize the net present value of the firm's expected future cash flows

$$\Gamma(k_1, k_2) = \Gamma_1(k_1) + \Gamma_2(k_2), \quad (3)$$

where

$$\Gamma_i(k_i) = \sum_{t=1}^T E_{\epsilon_i} [CM_{it}(k_i | x_{it}, w_{it}, \tilde{\epsilon}_{it})] \cdot \gamma^t - v_i \cdot k_i, \quad (4)$$

and  $CM_{it}(\cdot)$  denotes the maximized value of the expected future contribution margin in period  $t$ :

$$CM_{it}(k_i|x_{it}, w_{it}, \epsilon_{it}) \equiv x_{it} \cdot R_i(q_{it}^o(k_i, \cdot), \epsilon_{it}) - w_{it} \cdot q_{it}^o(k_i, \cdot),$$

with

$$q_{it}^o(k_i, \cdot) = \operatorname{argmax}_{q_{it} \leq k_i} \{x_{it} \cdot R_i(q_{it}, \epsilon_{it}) - w_{it} \cdot q_{it}\}.$$

The notation  $q_{it}^o(k_i, \cdot)$  above is short-hand for the sequentially optimal quantity  $q_{it}^o(k_i, x_{it}, w_{it}, \epsilon_{it})$  that maximizes the divisional contribution margin in period  $t$ , given the initial capacity choice  $k_i$ , current revenue and variable cost parameters (i.e.,  $x_{it}$  and  $w_{it}$ ), and the realization of the current shock  $\epsilon_{it}$ . To avoid laborious checking of boundary cases, we assume throughout our analysis that the marginal revenue at zero exceeds the unit variable cost of production for all  $\epsilon_{it}$ ; i.e.,

$$R'_i(0, \epsilon_{it}) - w_i > 0$$

for all realizations of  $\epsilon_{it}$ .

### 3.1 Stationary Environments

One significant simplification for the resource allocation problem we study obtains if the firm anticipates that the economic fundamentals are, at least in expectation, identical over the next  $T$  periods. Formally, an environment is said to be *stationary* if  $x_{it} = 1$ ,  $w_{it} = w_i$  and the  $\{\epsilon_{it}\}$  are i.i.d. for each  $i$ . For the setting of stationary environments, we drop subscript  $t$  from  $CM_{it}(\cdot)$  and  $q_{it}^o(\cdot)$ . The result below characterizes the efficient capacity levels,  $k_i^o$ , for this setting.

**Lemma 1** *Suppose capacity is dedicated and the divisional environments are stationary. If the optimal capacity level  $k_i^o$  is greater than zero, it is given by the unique solution to the equation:*

$$E_{\epsilon_i} \left[ R'_i(q_i^o(k_i^o, w_i, \tilde{\epsilon}_{it}), \tilde{\epsilon}_{it}) \right] = c_i + w_i, \quad (5)$$

where

$$c_i = \frac{v_i}{\sum_{t=1}^T \gamma^t}. \quad (6)$$

**Proof:** All proofs are in the Appendix.

Earlier literature, including Rogerson (2008) and Rajan and Reichelstein (2009), refers to  $c_i$  as the user cost of capital or the unit cost of capacity. The user cost of capital  $c_i$  is obtained by “annuitizing” the unit cost of capacity  $v_i$  (i.e., dividing  $v_i$  by  $\sum_{t=1}^T \gamma^t$ , which is the present value of \$1 annuity over  $T$  periods). It is readily verified that  $c_i$  is the price that a hypothetical supplier would charge for renting out capacity for one period of time if the rental business breaks even.

Lemma 1 says that the optimal capacity level  $k_i^o$  is such that the expected marginal revenue at the *sequentially optimal* production levels,  $q_i^o(k_i^o, \cdot)$  is equal to the sum of the unit cost of capacity  $c$  and the variable cost  $w_i$ . We shall subsequently refer to this sum,  $c_i + w_i$ , as the *full cost* per unit of output. As observed in Reichelstein and Rohlfsing-Bastian (2015),  $c_i + w_i$  will generally exceed the traditional measure of full cost in managerial accounting. The reason is that this measure does not include the imputed interest charges for capital. For instance, if the depreciation charges are uniform, the traditional measure of full cost in each period is given by  $\frac{v_i}{T} + w_i$ , which is less than  $\frac{v_i}{\sum_{t=1}^T \gamma^t} + w_i \equiv c_i + w_i$ .

In the context of our model, one common representation of full cost transfer pricing is that the downstream division is charged in the following manner for intra-company transfers:

1. Division 2 has the unilateral right to reserve capacity at the initial date.
2. Division 2 can choose the quantity,  $q_{2t}$ , to be transferred in each period subject to the initial capacity limit.
3. In period  $t$ , Division 2 is charged the full cost of output delivered, that is:

$$TP_t(k_2, q_{2t}) = (w_2 + c_2) \cdot q_{2t}.$$

This variant of full cost transfer pricing is essentially the one featured in the Harvard case study “Polysar” (Simons, 2000). The downstream division is charged for capacity only to the extent that it actually utilizes that capacity. A key takeaway from the Polysar case study is that the buying division will tend to reserve too much capacity upfront in the face of uncertain demand for its product. Such a strategy preserves the division’s option to meet market demand if it turns out to be strong, while it incurs no penalty for idling capacity if market conditions turn out to be unfavorable.

In contrast to the conclusion emerging from the Polysar case study, Dutta and Reichelstein (2010, Proposition 1) argue that with dedicated capacity full cost transfer pricing will result in efficient capacity investments. In their setting, however, the issue of capacity under-utilization does not arise because, by assumption, there are no variable costs of production (i.e.,  $w_i = 0$ ). Divisions may face uncertainty regarding the value of capacity, though given any investment they will sequentially always prefer to exhaust the capacity available.

An additional issue with the variant of cost-based transfer pricing described above is that unless  $q_{2t} = k_2$  in each period, the discounted value of the transfer pricing charges is not equal to the total discounted cost of the capacity investment and subsequent operating costs. While this is arguably not a crucial issue for an internal accounting rule, we nonetheless introduce the following balancing constraint:

**Definition** *A transfer pricing rule is said to be a full cost pricing rule if, in equilibrium:*

$$\sum_{t=1}^T TP_t(k_2, q_{2t}) \cdot \gamma^t = v_2 \cdot k_2 + \sum_{t=1}^T w_{2t} \cdot q_{2t} \cdot \gamma^t$$

The qualifier “in equilibrium” in the preceding definition refers to the notion that the transfer payments needs to be balanced only for the equilibrium investment and operating decisions. The precise notion of equilibrium will vary with the particular setting considered, specifically whether capacity is dedicated or fungible.



One natural way to deter divisional managers from reserving “excessive” amounts of capacity is the imposition of *excess capacity* charges.<sup>12</sup> In addition to the full cost of units delivered, the buying decision will then be charged in proportion to the amount of capacity not utilized at some rate  $\mu$ . A full cost transfer pricing rule subject to the excess capacity charges will entail the following transfer payments:

$$TP_t(k_2, q_{2t}) = (w_2 + c_2) \cdot q_{2t} + \mu \cdot (k_2 - q_{2t}). \quad (7)$$

In any given period, the available capacity will generally be fully utilized in good states of the world with high marginal revenues (high realizations of  $\epsilon_{it}$ ). On the other hand, capacity may be left idle under unfavorable market conditions (low realizations of  $\epsilon_{it}$ ). To state our first formal result, we introduce a notion of limited volatility in the revenue shocks  $\epsilon_{it}$  such that capacity will be fully utilized on the equilibrium path. Following Reichelstein and Rohlfling-Bastian (2015), the *limited volatility condition* is said to hold if  $q_i^o(k_i^o, \cdot) = k_i^o$  for all realizations of  $\epsilon_{it}$  where  $k_i^o$  again denotes the efficient capacity level. We note that the limited volatility condition will be met if and only if the inequality:

$$R'_i(k_i^o, \epsilon_{it}) - w_i \geq 0$$

holds for *all* realizations of  $\epsilon_{it}$ . Intuitively, the available capacity will always be exhausted in environments with relatively low volatility in terms of the range and impact of the  $\tilde{\epsilon}_{it}$ , or alternatively, if the unit variable cost,  $w_i$ , is small relative to the full cost,  $w_i + c_i$ . The limited volatility condition is thus a joint condition on the range of ex-post uncertainty and the relative magnitude of the unit variable cost relative to the full cost. If the separability condition  $R_i(q_i, \epsilon_{it}) = \epsilon_{it} \cdot \hat{R}_i(q_i)$  with  $E(\tilde{\epsilon}_{it}) = 1$  is met, the limited volatility condition holds if and only if  $\underline{\epsilon}_{it} \geq \frac{w_i}{w_i + c_i}$ .

---

<sup>12</sup>See, for instance, Kaplan (2006) and Martinez-Jerez (2007) on alternative rules for charging products and divisions for unused capacity costs.

**Proposition 1** *Suppose capacity is dedicated, the environment is stationary, and the limited volatility condition holds. Full cost transfer pricing subject to excess capacity charges, as given in (7), then achieves strong goal congruence provided  $\mu \geq c_2$  and capacity assets are depreciated according to the annuity rule.*

Excess capacity charges restore the efficiency of full cost transfers for two reasons. First, double marginalization is not an issue as the downstream division will internalize an incremental production cost of  $w_2 + c_2 - \mu \leq w_2$ . We note that the buying division will not have a short-run incentive to overproduce because the limited volatility condition ensures that the division would have exhausted the efficient capacity level,  $k_i^o$  for all realizations of  $\epsilon_{it}$  if it had imputed an incremental cost of  $w_i$  per unit of output. The downstream division will therefore also exhaust the available capacity for all  $\epsilon_{it}$  when it imputes a marginal cost less than  $w_2$ . Second, in making its initial capacity choice, the buying division will only internalize the actual unit cost of capacity,  $c_2$ , because, given the limited volatility condition, it does not anticipate excess capacity charges in equilibrium.<sup>13</sup>

Full cost transfer pricing subject to suitably chosen excess capacity charges provides the divisional managers with dominant strategy choices with regard to both their initial capacity and subsequent production decisions. The annuity depreciation schedule ensures that the financial consequences of the downstream division's choices merely "pass-through" the upstream division's performance measure because, in equilibrium, the transfer payment from Division 2 is precisely equal to the sum of depreciation, imputed capital charges, and variable production costs incurred by Division 1. Therefore, the performance evaluation system satisfies our criterion of separability.

We stress that for the above goal congruence result, it is essential that the excess capacity charge,  $\mu$ , be at least as large as the unit cost of capacity  $c_2$ . Otherwise,

---

<sup>13</sup>We note parenthetically that there would have been no need for excess excess capacity charges if either there is no periodic volatility in divisional revenues (the  $\epsilon_{it}$  are always equal to their average values) or there are no incremental costs to producing output ( $w_i = 0$ ).

the issues observed in connection with the transfer pricing policy in the Polysar case (where  $\mu = 0$ ) would resurface. Specifically, there would be a double marginalization problem in each period, since the downstream division would impute a marginal cost higher than  $w_2$ . In addition, this division would have incentives to procure excessive capacity because it is charged for the capacity only when actually utilized.

If the limited volatility condition for the buying division is not met, it will be essential to precisely calibrate the excess capacity charges. The obvious choice here is  $\mu = c_2$ , which results in the following *two-part full cost transfer pricing* rule:

$$TP_t(k_2, q_{2t}) = c_2 \cdot k_2 + w_2 \cdot q_{2t} \quad (8)$$

This pricing rule satisfies our criterion of a full cost transfer pricing rule insofar as the sum of the discounted transfer payments is identically equal to the initial capacity acquisition cost plus the discounted sum of the subsequent variable production costs. The transfer pricing rule in (8) also ensures that the performance measures are separable.

**Proposition 2** *With dedicated capacity and a stationary environment, the two-part full cost transfer pricing rule in (8) achieves strong congruence, provided capacity assets are depreciated according to the annuity rule.*

The two-part full cost transfer pricing rule charges the downstream division separately for (i) the amount of capacity that it reserves initially, and (ii) the variable cost of output that it procures actually in each period. This form of full cost transfer pricing rule eliminates the downstream division's incentives to reserve too much capacity upfront as well as the double marginalization problem associated with the naive full cost transfer pricing rule. In fact, it can be verified that absent any restrictions on the amount of volatility, the two-part transfer pricing mechanism in (8) is unique among the class of linear transfer pricing rules of the form  $TP_t(k_2, q_{2t}) = a_1 \cdot k_2 + a_2 \cdot q_{2t}$ ; i.e.,  $a_1 = c_2$  and  $a_2 = w_2$  are not only sufficient but also necessary for strong goal congruence.

## 3.2 Non-Stationary Environments

We have thus far restricted our analysis to stationary environments in which each division's costs and expected revenues are identical across periods. In this subsection, we investigate depreciation and transfer pricing rules that can achieve strong goal congruence for certain non-stationary environments. The following result characterizes the efficient capacity choices by generalizing Lemma 1 for non-stationary environments:

**Lemma 2** *If capacity is dedicated and the optimal capacity level,  $k_i^o$ , in (5) is greater than zero, it is given by the unique solution to the equation:*

$$\sum_{t=1}^T E_{\epsilon_{it}} \left[ x_{it} \cdot R'_i(q_{it}^o(k_i^o, \tilde{\epsilon}_{it}, x_{it}, w_{it}), \tilde{\epsilon}_{it}) \right] \cdot \gamma^t = v_i + \bar{w}_i \quad (9)$$

where

$$\bar{w}_i = \sum_{t=1}^T w_{it} \cdot \gamma^t.$$

It is readily seen that the claim in Lemma 2 reduces to that in Lemma 1 whenever  $x_{it} = 1$ ,  $w_{it} = w_i$  and  $\{\epsilon_{it}\}$  are i.i.d. Beginning with the work of Rogerson (1997), earlier work on goal congruent performance measures has shown that if the revenues attained vary across time periods, proper intertemporal cost allocation of the initial investment expenditure requires that depreciation be calculated according to the *relative benefit* rule rather than the simple annuity rule. This insight extends to the setting of our model provided the variable costs of production change in a coordinated fashion over time. Formally, the relative benefit depreciation charges are the ones defined by the requirement that the overall capital charge in period  $t$  (i.e., the sum of depreciation and imputed interest charges), as introduced in equation (2), be given by:<sup>14</sup>

---

<sup>14</sup>As pointed out by earlier studies, the corresponding relative benefit depreciation charges will coincide with straight-line depreciation if the  $x_{it}$  decline linearly over time at a particular rate (Nezlobin et al. 2012).

$$\hat{z}_{it} \equiv \frac{x_{it}}{\sum_{\tau=1}^T x_{i\tau} \cdot \gamma^\tau}$$

**Corollary to Proposition 2:** *If capacity is dedicated and  $w_{it} = x_{it} \cdot w_i$ , a two-part full cost transfer pricing rule of the form*

$$TP_t(k_2, q_{2t}) = \hat{z}_{2t} \cdot v_2 \cdot k_2 + w_{2t} \cdot q_{2t}$$

*achieves strong congruence, provided capacity assets are depreciated according to the relative benefit depreciation rule.*

The preceding result generalizes the result in Proposition 2 to a class of non-stationary environments in which expected revenues and variable costs are different across periods. However, the settings to which the above result applies is rather restrictive. Specifically, the result requires that intertemporal variations in periodic revenues and variable production costs follow identical patterns (i.e.,  $w_{it} = x_{it} \cdot w_i$ ).

With limited volatility, the result below shows that the finding of Proposition 2 can be extended to a class of non-stationary environments.

**Proposition 3** *Suppose capacity is dedicated, the limited volatility condition holds, and the  $\{\epsilon_{it}\}$  are i.i.d. The full cost transfer pricing rule*

$$TP_t(k_2) = \hat{z}_{2t} \cdot (v_2 + \bar{w}_2) \cdot k_2$$

*achieves strong goal congruence, provided the anticipated variable production costs of each division,  $\bar{w}_i \cdot k_i$ , are capitalized and the divisional capitalized costs,  $(v_i + \bar{w}_i) \cdot k_i$  are depreciated according to the respective relative benefit rule.*

The above transfer pricing rule does not charge the downstream division for actual variable costs incurred in connection with the actual production volume. Instead, the buying division is charged for the “budgeted” variable costs that will be incurred in future time periods assuming that the initially chosen capacity chosen will be fully

exhausted in all future periods. Such a policy is indeed efficient if (i) the limited volatility condition holds, and (ii) the downstream division has an incentive to choose the efficient capacity level in the first place. Given the above transfer pricing rule, the downstream division will choose  $k_2$  to maximize:

$$\sum_{t=1}^T E_{\epsilon_2} [x_{2t} \cdot R_2(k_2, \tilde{\epsilon}_{2t})] \cdot \gamma^t - (v_2 + \bar{w}_2) \cdot k_2.$$

Thus, the downstream division's objective function coincides with that of the firm for any  $k_2 \leq k_2^o$ .

We note that  $TP_t(k_2) = \hat{z}_{2t} \cdot (v_2 + \bar{w}_2) \cdot k_2$  is a full-cost transfer pricing rule because in equilibrium, Division 2 initially procures  $k_2^o$  and subsequently exhausts the available capacity. However, this transfer pricing rule no longer achieves separability because the upstream division's variable costs of production are balanced by the transfer payments received from the buying division only over the entire  $T$  period horizon, but not on a period-by-period basis.

To extend the preceding result to environments where the limited volatility condition may not be satisfied, we adopt the binary investment level model in Baldenius, Nezlobin and Vaysman (2016, Proposition 1). Specifically, suppose that each division chooses whether to install a specific amount of capacity  $\bar{k}_i$  or not; i.e.,  $k_i \in \{0, \bar{k}_i\}$ . Suppose further that each division's revenue function  $R_i(\cdot, \tilde{\epsilon}_{it})$  is publicly known, but each divisional manager's private information is a one-dimensional parameter  $\theta_i$  which affects the probability distributions of  $\tilde{\epsilon}_{it}$ . We assume that  $\theta_i$  shifts the conditional densities  $f_i(\epsilon_{it}|\theta_i)$  in the sense of first-order stochastic dominance.

The essential simplification with binary investment choices is that the accrual accounting rules, i.e., depreciation schedule and transfer pricing rule, only need to separate the types of  $\theta_i$  for whom capacity investment is in the firm's interest from those types for whom it is not. Accordingly, we denote the threshold type where the

firm is just indifferent between investing and not investing by  $\theta_i^*$ . Thus,

$$\Gamma_i(\bar{k}_i|\theta_i^*) = \sum_{t=1}^T E_{\epsilon_i} [CM_{it}(\bar{k}_i|x_{it}, w_{it}, \tilde{\epsilon}_{it})|\theta_i^*] \cdot \gamma^t - v_i \cdot \bar{k}_i = 0. \quad (10)$$

As before,  $CM_{it}(\cdot)$  denotes the maximized value of the expected future contribution margin in period  $t$ :

$$CM_{it}(\bar{k}_i|x_{it}, w_{it}, \epsilon_{it}) \equiv x_{it} \cdot R_i(q_{it}^o(\bar{k}_i, \cdot), \epsilon_{it}) - w_{it} \cdot q_{it}^o(\bar{k}_i, \cdot).$$

Following the terminology in Baldenius, Nezlobin and Vaysman (2016), we refer to the *Relative Expected Optimized Benefit* (REOB) cost allocation rule as:

$$\bar{z}_{it} = \frac{E_{\epsilon_i} [CM_{it}(\bar{k}_i|x_{it}, w_{it}, \tilde{\epsilon}_{it})|\theta_i^*]}{\sum_{\tau=1}^T E_{\epsilon_i} [CM_{i\tau}(\bar{k}_i|x_{i\tau}, w_{i\tau}, \tilde{\epsilon}_{i\tau})|\theta_i^*] \cdot \gamma^\tau}.$$

The REOB rule is effectively the relative benefit rule for the threshold type  $\theta_i^*$ , and reduces to annuity depreciation in a stationary environment.

**Proposition 4** *Suppose the set of feasible capacity investment choices is binary and the future realizations of  $\tilde{\epsilon}_{it}$  are drawn according to conditional densities  $f(\epsilon_{it}|\theta_i)$  such that  $\theta_i$  shifts  $f(\epsilon_{it}|\theta_i)$  in the sense of first-order stochastic dominance. The full-cost transfer pricing rule*

$$TP_t(k_2, q_{2t}) = \bar{z}_{2t} \cdot v_2 \cdot k_2 + w_{2t} \cdot q_{2t}$$

*then achieves strong goal congruence, provided capacity assets are depreciated according to the REOB rule.*

Like the two-part tariff in Proposition 2, the transfer pricing rule identified in the above finding is a full-cost transfer pricing rule which satisfies the criterion of separability. To check goal congruence, it can be shown that both the expected value of the maximized contribution margin,  $E_{\epsilon_i} [CM_{i\tau}(\bar{k}_i|x_{i\tau}, w_{i\tau}, \tilde{\epsilon}_{i\tau})|\theta_i]$ , and the net present value of the capacity investment,  $\Gamma_i(\bar{k}_i|\theta_i)$ , are increasing in  $\theta_i$ . Consider

now the downstream division's incentive to invest. If that division were to focus exclusively on its profit measure in period  $t$ ,  $1 \leq t \leq T$ , it would seek to maximize:

$$E_{\epsilon_2} [\pi_2(k|\theta_2, \tilde{\epsilon}_{2t})] \equiv E_{\epsilon_2} [CM_{2t}(k_2|x_{2t}, w_{2t}, \tilde{\epsilon}_{2t}, \theta_2)] - \bar{z}_{2t} \cdot v_2 \cdot k_2.$$

By construction of the REOB rule,  $E_{\epsilon_2} [\pi_2(k|\theta_2, \tilde{\epsilon}_{2t})] > 0$  if and only if

$$E_{\epsilon_2} [CM_{2t}(\bar{k}_2|x_{2t}, w_{2t}, \tilde{\epsilon}_{2t})|\theta_2] > E_{\epsilon_2} [CM_{2t}(\bar{k}_2|x_{2t}, w_{2t}, \tilde{\epsilon}_{2t})|\theta_2^*],$$

which will be the case if and only if  $\theta_2 > \theta_2^*$ .

In concluding this section, we recall that Propositions 1-4 have identified environments where some variant of full cost transfer pricing is part of a goal congruent performance measurement system. Common to these pricing rules is that capacity related costs are charged in a lump-sum fashion against revenues so as to ensure that the charges have no effect on subsequent production decisions. Yet, the specific rules for allocating fixed costs and charging for anticipated variable costs vary with the particular setting, i.e., demand volatility, stationarity, and the investment opportunity set.

## 4 Fungible Capacity

In contrast to the scenario considered thus far, where the products or services provided by the two divisions required different production assets, we now consider the plausible alternative of fungible capacity. Accordingly, the production processes of the two divisions have enough commonalities and the demand shocks  $\epsilon_t$  are realized sufficiently early in each period, so that the initial capacity choices can be reallocated across the two divisions. The following time line illustrates the sequence of events at the initial investment date and in a generic period  $t$ .



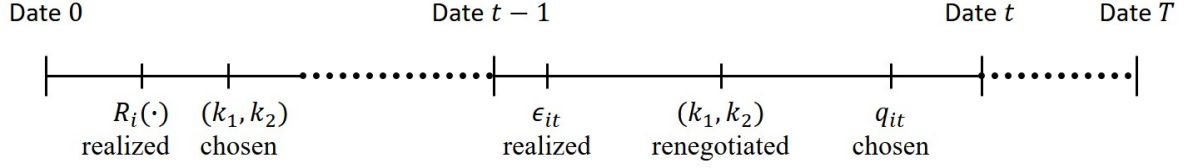


Figure 3: **Sequence of Events in the Fungible Capacity Scenario**

The analysis below focuses at first on stationary environments. With fungible capacity, the optimal investment from a firm-wide perspective is the one that maximizes total expected future cash flows:

$$\Gamma(k) = \sum_{t=1}^T E_{\epsilon_t} [CM(k|w, \tilde{\epsilon}_t)] \cdot \gamma^t - v \cdot k, \quad (11)$$

where  $w \equiv (w_1, w_2)$ , and  $\epsilon_t \equiv (\epsilon_{1t}, \epsilon_{2t})$  and  $CM(\cdot)$  denotes the maximized value of the aggregate contribution margin in period  $t$ . That is,

$$CM(k|w, \tilde{\epsilon}_t) \equiv \sum_{i=1}^2 [R_i(q_i^*(k, \cdot), \epsilon_{it}) - w \cdot q_i^*(k, \cdot)],$$

where

$$(q_1^*(k, \cdot), q_2^*(k, \cdot)) = \operatorname{argmax}_{q_1 + q_2 \leq k} \left\{ \sum_{i=1}^2 [R_i(q_i, \epsilon_{it}) - w_i \cdot q_i] \right\}.$$

As before, the notation  $q_i^*(k, \cdot)$  is short-hand for  $q_i^*(k, w, \epsilon_t)$ .

Provided the optimal quantities  $q_i^*(k, \cdot)$  are both positive, the first-order condition:

$$R'_1(q_1^*(k, \cdot), \epsilon_{1t}) - w_1 = R'_2(q_2^*(k, \cdot), \epsilon_{2t}) - w_2 \quad (12)$$

must hold. Allowing for corner solutions, we define the *shadow price* of capacity in period  $t$ , given the available capacity  $k$ , as follows:

$$S(k|w, \epsilon_t) \equiv \max\{R'_1(q_1^*(k, \cdot), \epsilon_{1t}) - w_1, R'_2(q_2^*(k, \cdot), \epsilon_{2t}) - w_2\}. \quad (13)$$

The shadow price of capacity identifies the maximal change in periodic contribution margin that the firm can obtain from an extra unit of capacity.<sup>15</sup> We note that  $S(\cdot)$  is increasing in  $\epsilon_{it}$ , but decreasing in  $w_i$  and  $k$ .

**Lemma 3** *Suppose capacity is fungible and the divisional environments are stationary. The optimal capacity level,  $k^*$ , is given by the unique solution to the equation:*

$$E_\epsilon [S(k^*|w, \tilde{\epsilon}_t)] = c, \quad (14)$$

where

$$c = \frac{v}{\sum_{t=1}^T \gamma^t}. \quad (15)$$

We next examine the divisions' capacity investment choices in the decentralized setting. Given a stationary environment, Proposition 2 suggests that the two-part full cost transfer pricing rule in (8) can induce goal congruence if the divisions are allowed to renegotiate the initial capacity rights after realization of revenue shocks  $\epsilon_t$  in each period. In this negotiation, the full cost pricing rule determines the parties' status quo payoffs.

Suppose that the downstream division has procured initial rights for  $k_2$  units of capacity, the upstream division has installed  $k_1$  units of capacity for its own use, and hence  $k = k_1 + k_2$  is the corresponding amount of firm-wide capacity. As shown in Dutta and Reichelstein (2010), if the two divisions have symmetric information about each other's revenues and costs, they can increase the firm-wide contribution margin by reallocating the available capacity  $k_1 + k_2$  at the beginning of each period after the relevant shock  $\epsilon_t$  is realized. The resulting "trading surplus" of

$$TSP \equiv CM(k|w, \epsilon_t) - \sum_{i=1}^2 CM_i(k_i|w_i, \epsilon_{it}) \quad (16)$$

---

<sup>15</sup>The assumption that  $R'_i(0, \epsilon_{it}) \geq w_i$  for all  $\epsilon_{it}$  ensures that the shadow price of capacity is always non-negative.

can then be shared by the two divisions. Let  $\delta \in [0, 1]$  denote the fraction of the total surplus that accrues to Division 1. Thus, the parameter  $\delta$  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,  $\Delta TP_t$ , that implements the above sharing rule is given by

$$R_1(q_1^*(k, \cdot), \epsilon_{1t}) - w_1 \cdot q_1^*(k, \cdot) + \Delta TP_t = CM_1(k_1|w_1, \epsilon_{1t}) + \delta \cdot TSP,$$

where we recall that  $q_1^*(k, \cdot)$  and  $q_2^*(k, \cdot)$  are the divisional production choices that maximize the aggregate contribution margin. At the same time, Division 2 obtains:

$$R_2(q_2^*(k, \cdot), \epsilon_{2t}) - w_2 \cdot q_2^*(k, \cdot) - \Delta TP_t = CM_2(k_2|w, \epsilon_{2t}) + (1 - \delta) \cdot TSP.$$

These payoffs ignore the transfer payment  $c \cdot k_2$  that Division 2 makes at the beginning of the period, since this payment is viewed as sunk at the renegotiation stage. The total transfer payment made by Division 2 in return for the ex-post efficient quantity  $q_2^*(k, \cdot)$  is then given  $c \cdot k_2 + w_2 \cdot q_2^*(k, \cdot) + \Delta TP_t$ .

After substituting for  $TSP$  from (16), the effective contribution margin to Division  $i$  can be expressed as follows:

$$CM_1^*(k_1, k_2|\epsilon_t) = (1 - \delta) \cdot CM_1(k_1|w_1, \epsilon_{1t}) + \delta \cdot [CM(k|w, \epsilon_t) - CM_2(k_2|w_2, \epsilon_{2t})]$$

and

$$CM_2^*(k_1, k_2|\epsilon_t) = \delta \cdot CM_2(k_2|w_2, \epsilon_{2t}) + (1 - \delta) \cdot [CM(k|w, \epsilon_t) - CM_1(k_1|w_1, \epsilon_{1t})].$$

We note that the expected value of the effective contribution margin,  $E_\epsilon [CM_i^*(k_i, k_j|\epsilon_t)]$ , is identical across periods for stationary environments. Combined with the annuity depreciation rule for capacity assets, this implies that division  $i$  will choose  $k_i$  to maximize:

$$E_\epsilon [CM_i^*(k_i, k_j|\tilde{\epsilon}_t)] - c \cdot k_i \tag{17}$$

taking division  $j$ 's capacity request  $k_j$  as given.

It is useful to observe that in the extreme case where Division 1 has all the bargaining power ( $\delta = 1$ ), Division 1 would fully internalize the firm's objective and choose the efficient capacity level  $k^*$ . Similarly, in the other corner case of  $\delta = 0$ , Division 2 would internalize the firm's objective and choose  $k_2$  such that Division 1 responds with the efficient capacity level  $k^*$ .

If  $(k_1, k_2)$  constitutes a Nash equilibrium of the divisional capacity choice game with  $k_i > 0$  for each  $i$ , then, by the Envelope Theorem, the following first-order conditions are met:

$$E_\epsilon \left[ (1 - \delta) \cdot CM'_1(k_1|w_1, \tilde{\epsilon}_{1t}) + \delta \cdot S(k_1 + k_2|w, \tilde{\epsilon}_t) \right] = c \quad (18)$$

and

$$E_\epsilon \left[ \delta \cdot CM'_2(k_2|w_2, \tilde{\epsilon}_{2t}) + (1 - \delta) \cdot S(k_1 + k_2|w, \tilde{\epsilon}_t) \right] = c, \quad (19)$$

where  $CM'_i(k_i|w_i, \epsilon_{it}) \equiv R'_i(q_i^o(k_i, \cdot), \epsilon_{it}) - w_i$  is the marginal contribution margin in the dedicated capacity scenario. It can be verified from the proofs of Lemma 1 and Lemma 3 that  $CM'_i(\cdot)$  and  $S(\cdot)$  are decreasing functions of  $k_i$ , and hence each division's objective function is globally concave.

Similar to the arguments in Dutta and Reichelstein (2010), the above first-order conditions show that each division's incentives to acquire capacity stem both from the unilateral "stand-alone" use of capacity as well as the prospect of trading capacity with the other division. The second term on the left-hand side of both (18) and (19) represents the firm's aggregate and *optimized* marginal contribution margin, given by the (expected) shadow price of capacity. Since the divisions individually only receive a share of the aggregate return (given by  $\delta$  and  $1 - \delta$ , respectively), this part of the investment return entails a "classical" holdup problem.<sup>16</sup> Yet, the divisions also

---

<sup>16</sup>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).

derive *direct* 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 (18) and (19), respectively.<sup>17</sup>

Equations (18) and (19) also highlight the importance of allowing both divisions to secure capacity rights. The firm would generally face an underinvestment problem if only one division were allowed to secure capacity. For instance, if only the upstream division were to acquire capacity, its marginal contribution margin at the efficient capacity level  $k^*$  would be:

$$E_\epsilon \left[ (1 - \delta) \cdot CM'_1(k^*|w_1, \tilde{\epsilon}_{1t}) + \delta \cdot S(k^*|w, \tilde{\epsilon}_t) \right].$$

This marginal revenue is, however, less than  $E_\epsilon [S(k^*|w, \tilde{\epsilon}_t)] = c$  because

$$\begin{aligned} E_\epsilon [CM'_1(k^*|w_1, \tilde{\epsilon}_{1t})] &= E_\epsilon \left[ R'_1(q^o(k^*, \cdot), \tilde{\epsilon}_{1t}) - w_1 \right] \\ &\leq E_\epsilon \left[ R'_1(q_1^*(k^*, \cdot), \tilde{\epsilon}_{1t}) - w_1 \right] \\ &\leq E_\epsilon [S(k^*|w, \tilde{\epsilon}_t)], \end{aligned}$$

where the first inequality above is a consequence of the fact that  $q_1^o(k^*, \cdot) \geq q_1^*(k^*, \cdot)$ . Thus the upstream division would have insufficient incentives to secure the firm-wide optimal capacity level on its own, since it would anticipate a classic hold-up on its investment in the subsequent negotiations.

The following result identifies a class of environments for which the two-part full cost transfer pricing rule achieves strong goal congruence provided the divisions are allowed to periodically renegotiate the initial capacity rights and capacity assets are depreciated according to the annuity depreciation rule. To that end, it will be useful to make the following assumption regarding the divisional revenue functions:

---

<sup>17</sup>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.

$$R_i(q, \theta_i, \epsilon_{it}) = \epsilon_{it} \cdot \theta_i \cdot q - h_{it} \cdot q^2. \quad (20)$$

We assume that while the quadratic functional form in (20) is commonly known, the firm's central office does not have sufficient information about the divisional revenue functions because the parameters  $(\theta_1, \theta_2)$  are known only to the two divisional managers.

**Proposition 5** *Suppose the divisional revenue functions take the quadratic form in (20) and the limited volatility condition is satisfied in the dedicated capacity setting. A system of decentralized initial capacity choices combined with the full cost transfer pricing rule*

$$TP_t(k_2, q_{2t}) = c \cdot k_2 + w_2 \cdot q_{2t}$$

*achieves strong goal congruence, provided the divisions are free to renegotiate the initial capacity rights and capacity assets are depreciated according to the annuity rule.*

The proof of Proposition 5 shows that the quadratic form of divisional revenues in (20) has the property that the resulting shadow price function  $S(k|\theta, w, \epsilon_t)$  is linear in  $\epsilon_t$ . Combined with the limited volatility condition, linearity of the shadow price  $S(\cdot)$  in  $\epsilon_t$  implies that the efficient capacity in the fungible capacity scenario is the same as in the dedicated capacity setting; i.e.,  $k^* = k_1^o + k_2^o$ . Furthermore, when the limited volatility condition holds, the stand-alone capacity levels  $(k_1^o, k_2^o)$  are the unique solution to the divisional first-order conditions in (18) and (19).

Proposition 5 can be extended to non-stationary environments in which the revenue factors  $x_{it}$  and variable costs  $w_{it}$  differ across periods. Generalizing the result in Lemma 3, it can be shown that the optimal capacity  $k^*$  is given by:

$$E_\epsilon \left[ \sum_{t=1}^T \gamma^t \cdot S_t(k^* | x_t, w_t, \tilde{\epsilon}_t) \right] = v$$

where

$$S_t(k|x_t, w_t, \epsilon_t) \equiv \max\{x_{1t} \cdot R'_1(q_{1t}^*(k, \cdot), \epsilon_{1t}) - w_{1t}, x_{2t} \cdot R'_2(q_{2t}^*(k, \cdot), \epsilon_{2t}) - w_{2t}\}$$

is the shadow price of capacity in period  $t$ . With quadratic revenue functions and the limited volatility condition in place, it can again be verified that the efficient capacity in the fungible setting is the same as in the dedicated setting; i.e.,  $k^* = k_1^o + k_2^o$ . Adapting the transfer pricing rule in Proposition 3, suppose that the anticipated variable costs of production are capitalized and the divisional assets are depreciated according to the relative benefit rule. The same arguments as those in the proofs of Propositions 3 and 5 then show that the corresponding full cost transfer pricing rule:

$$TP_t(k_2) = \hat{z}_{2t} \cdot (v + \bar{w}_2) \cdot k_2,$$

with  $\hat{z}_{2t}$  and  $\bar{w}_2$  as defined in Section 3.2, will induce strong goal congruence provided the divisions are allowed to renegotiate the initial capacity rights.

The rules for choosing capacity choice and pricing transfers in Proposition 5 rely on both the limited volatility condition and the restriction that the divisional revenue functions can effectively be approximated by quadratic functions. Intuitively, the importance of the quadratic revenue functions is that the expected marginal revenue is equal to the marginal revenue at the expected value of  $\epsilon_{it}$ . With this structure, the divisional coordination problem in choosing the overall level of capacity can be solved by letting the divisions make these choices simultaneously and independently.

For more general environments, we investigate whether the coordination problem regarding capacity investments can be resolved by a sequential mechanism that gives the upstream additional supervisory authority. In effect, the upstream division can now be viewed as a “gatekeeper” whose approval is required for any capacity the downstream division wants to reserve for itself. Specifically, in order to acquire unilateral capacity rights, the downstream division needs to receive approval from the upstream division.<sup>18</sup> If the two divisions reach such an upfront agreement, it

---

<sup>18</sup>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

specifies the downstream division's unilateral capacity rights  $k_2$  and a corresponding transfer payment  $p(k_2)$  that must be made to the upstream division for granting these rights in each subsequent period. The parties report the outcome of this agreement  $(k_2, p(k_2))$  to the central office, which commits to enforce this outcome unless the parties renegotiate it.

The upstream division is free to install additional capacity for its own needs in addition to what has been secured by the downstream division. As before, capacity assets are depreciated according to the annuity depreciation rule, and thus the upstream division is charged  $c$  for each unit of capacity that it acquires. If the parties fail to reach a mutually acceptable agreement, the downstream division would have no ex-ante claim on capacity, though it may, of course, obtain capacity ex-post through negotiation with the other division. We summarize this negotiated gatekeeper transfer pricing arrangement as follows:

- The two divisions negotiate an ex-ante contract  $(k_2, p(k_2))$  which gives Division 2 unilateral rights to  $k_2$  units of capacity in return for a fixed payment of  $p(k_2)$  in each period.
- Subsequently, Division 1 installs  $k \geq k_2$  units of capacity,
- If Division 2 procures  $q_{2t}$  units of output in period  $t$ , the corresponding transfer payments is calculated as  $TP_t(k_2, q_{2t}) = p(k_2) + w_2 \cdot q_{2t}$ .
- After observing the realization of revenue shocks  $\epsilon_t$  in each period, the divisions can renegotiate the initial capacity rights.

For the result below, we assume that the optimal dedicated capacity level  $k_i^o$  is non-zero for each  $i$ . It can be readily verified from the proof of Lemma 1 that a necessary and sufficient condition for  $k_i^o$  to be positive is

$$E_{\epsilon_i} [R'_i(0, \epsilon_{it})] > c_i + w_i. \tag{21}$$

---

analysis makes clear that the role of the two divisions could be switched.



**Proposition 6** *Suppose the divisional environments are stationary and the downstream division's unilateral capacity rights are determined through negotiation. The transfer pricing rule*

$$TP_t(k_2, q_{2t}) = p(k_2) + w_2 \cdot q_{2t}$$

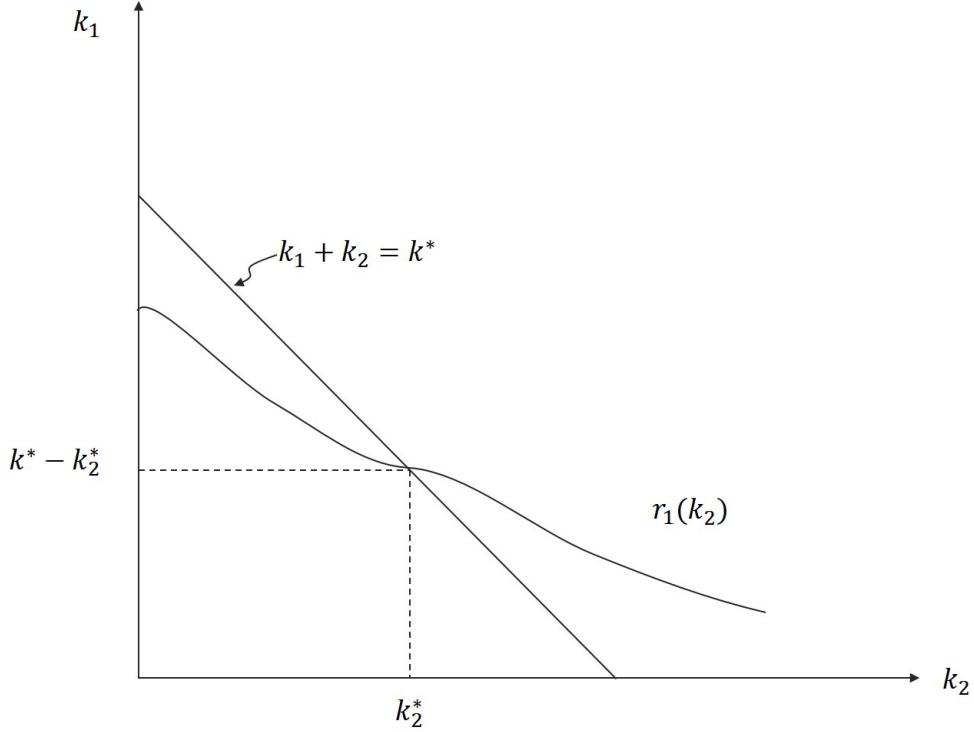
*achieves strong goal congruence, provided the divisions are free to renegotiate the initial capacity rights in each period and capacity assets are depreciated according to the annuity depreciation rule.*

A gatekeeper arrangement will attain strong goal congruence if it induces the two divisions to acquire collectively the efficient capacity level,  $k^*$ . The proof of Proposition 6 demonstrates that in order to maximize their joint expected surplus, the divisions will agree on a particular amount of capacity level  $k_2^* \in [0, k^*)$  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  $k^*$ , giving this division then an exclusive claim on  $k^* - k_2^*$  units of capacity.

To provide further intuition, suppose the two divisions have negotiated an ex-ante contract that gives the downstream division rights to  $k_2$  units of capacity in each period. In response to this choice of  $k_2$ , the upstream division chooses  $r_1(k_2)$  units of capacity for its own use, and thus installs  $r_1(k_2) + k_2$  units of aggregate capacity. The upstream division's reaction function,  $r_1(k_2)$ , will satisfy the first-order condition in (18); i.e.,

$$E_\epsilon \left[ (1 - \delta) \cdot CM_1'(r_1(k_2)|w_1, \tilde{\epsilon}_{1t}) + \delta \cdot S(r_1(k_2) + k_2|w, \tilde{\epsilon}_t) \right] = c.$$

As illustrated in Figure 4 below, the reaction function  $r_1(k_2)$  is downward-sloping because both  $CM_1'(k_1|\cdot)$  and  $S(k|\cdot)$  are decreasing functions.



**Figure 4: Division 1's Reaction Function**

Furthermore, the proof of Proposition 6 shows that  $r_1(0) \leq k^*$  and  $r_1(k^*) > 0$ . Therefore, as shown in Figure 4, there exists a  $k_2^* \in [0, k^*)$  such that the upstream division responds with  $r_1(k_2^*) = k^* - k_2^*$ , and hence installs the optimal amount of aggregate capacity  $k_1^*$  on its own.

The ex-ante agreement  $(k_2^*, p(k_2^*))$  must be such that it is preferred by both divisions to the default point of no agreement. If the two divisions fail to reach an ex-ante agreement, the upstream division will choose its capacity level unilaterally, and the downstream division will receive no initial capacity rights. By agreeing to transfer  $k_2^*$  units of capacity rights to the downstream division, the two divisions can generate additional surplus. The fixed transfer payment  $p(k_2^*)$  is chosen such that this additional surplus is split between the two divisions in proportion to their relative bargaining powers.

We note that in comparison to the preceding setting where the two divisions have symmetric capacity rights (Proposition 5), the downstream division is worse-off under the gatekeeper arrangement. The upstream division will extract some of the expected surplus contributed by the other division. At the same time, the specification of the default outcome, in case the parties were not to reach an agreement at the initial stage, is of no particular importance for the efficiency result in Proposition 6. The same outcome, albeit with a different transfer payment, would result if the mechanism were to specify that in the absence of an agreement the downstream division could claim some share of the capacity subsequently procured by the upstream division at the transfer price:

$$TP(q_{2t}) = (c + w_2) \cdot q_{2t}.$$

The allocation mechanism in Proposition 6 can be interpreted as a hybrid between full cost and negotiated transfer pricing such that the upstream division is charged for the full cost of the entire capacity and output produced by the divisions. Those charges are split between the two divisions through a two-stage negotiation. The latter feature is also the key to the efficiency of the fixed quantity contracts in Edlin and Reichelstein (1995). In their model, a properly set default quantity of a good to be traded provides the parties with incentives to make efficient relationship-specific (unverifiable) investments. In the context of our model, an agreement on the unilateral capacity rights of the downstream division induces the investment center to acquire residual capacity rights for itself such that the overall capacity procured is efficient from a firm-wide perspective.

## 5 Conclusion

This paper has re-examined the incentive properties of full cost transfer pricing rule in multi-divisional firms. Our analysis is motivated by the fact that this form of internal pricing remains ubiquitous in practice despite the many concerns that have been

expressed about it in textbooks and the academic literature. The main ingredients in our model are that divisional managers are responsible for the initial acquisition of capacity as well as its subsequent utilization in future periods. An upstream division installs capacity and provides production services for both divisions, since it has the necessary technical expertise. In each period, the upstream division receives a transfer payment for providing capacity- and production services to the downstream division.

We identify circumstances in which a suitable variant of full cost transfer pricing induces efficient capacity acquisition and subsequent production decisions. From an ex-ante capacity planning perspective, variable cost pricing is clearly inadequate because the buying division will not internalize the relevant capacity costs, and hence this pricing rule generates incentives for the buying division to initially request an excessive amount of capacity. At the same time, a simplistic form of full cost transfer pricing that charges the buying division only for the cost of actually utilized capacity will also not achieve efficient outcomes as this rule again motivates the buying division to request an inefficiently large amount of capacity. Our results demonstrate that, depending on the characteristics of the underlying production and market environment, particular variants of two-part full cost transfer pricing can indeed lead to efficient decentralization.

When the divisions can share the same productive assets for their production needs, an efficient allocation of the available capacity can be achieved ex-post through bilateral negotiation. We find that potential hold-up problems on investments resulting from ex-post negotiation can be alleviated through an appropriate assignment of initial capacity rights in conjunction with full cost transfer prices that determine the divisions' default payoffs at the negotiation stage.

# Appendix

**Proof of Lemma 1:**

With dedicated capacity, the firm's objective function is additively separable across the two divisions. For a stationary environment, the firm seeks a capacity level  $k_i^o$  that maximize total expected cash flows

$$\Gamma_i(k_i) = \sum_{t=1}^T E_{\epsilon_i} [R_i(q_i^o(k_i, \cdot), \tilde{\epsilon}_{it}) - w_i \cdot q_i^o(\cdot)] \gamma^t - v_i \cdot k_i, \quad (22)$$

where  $q_i^o(k_i, \cdot) \equiv q_i^o(k_i, w_i, \epsilon_{it})$  is given by

$$q_i^o(\cdot) = \operatorname{argmax}_{q_i \leq k_i} \{R_i(q_i, \epsilon_{it}) - w_i \cdot q_i\}.$$

Dividing the objective function in (22) by the annuity factor  $\sum_{t=1}^T \gamma^t$ , the firm seeks a capacity level,  $k_i^o$  for Division  $i$  that maximizes:

$$E_{\epsilon_i} [CM_i(k_i | w_i, \tilde{\epsilon}_{it})] - c \cdot k_i,$$

where

$$CM_i(k_i | w_i, \epsilon_{it}) \equiv R_i(q_i^o(k_i, \cdot), \epsilon_{it}) - w_i \cdot q_i^o(k_i, \cdot)$$

is the maximized value of contribution margin in period  $t$ .

**Claim:**  $CM_i(k_i | w_i, \epsilon_{it})$  is differentiable in  $k_i$  for all  $\epsilon_{it}$  and

$$\frac{\partial}{\partial k} CM_i(k_i | w_i, \epsilon_{it}) = R'_i(q_i^o(k_i, \cdot), \epsilon_{it}) - w_i.$$

**Proof of Claim:** We first note that

$$\begin{aligned} & \frac{CM_i(k_i + \Delta | w_i, \epsilon_{it}) - CM_i(k_i | w_i, \epsilon_{it})}{\Delta} \\ & \geq \frac{R_i(q_i^o(k_i, \cdot) + \Delta, \epsilon_i) - R_i(q_i^o(k_i, \cdot), \epsilon_{it})}{\Delta} - w_i. \end{aligned} \quad (23)$$

This inequality follows directly by observing that

$$CM_i(k_i + \Delta | w_i, \epsilon_{it}) \geq R_i(q_i^o(k_i, \cdot) + \Delta, \epsilon_{it}) - w_i \cdot [q_i^o(k_i, \cdot) + \Delta].$$

At the same time, we find that

$$\begin{aligned} & \frac{CM_i(k_i + \Delta | w_i, \epsilon_{it}) - CM_i(k_i | w_i, \epsilon_{it})}{\Delta} \\ & \leq \frac{R_i(q_i^o(k_i + \Delta, \cdot), \epsilon_{it}) - R_i(q_i^o(k_i + \Delta, \cdot) - \Delta, \epsilon_{it})}{\Delta} - w_i. \end{aligned} \quad (24)$$

To see this, we note that

$$\begin{aligned} & CM_i(k_i + \Delta | w_i, \epsilon_{it}) - CM_i(k_i | w_i, \epsilon_{it}) \\ & \leq R_i(q_i^o(k_i + \Delta, w_i, \epsilon_{it}), \epsilon_{it}) - w_i \cdot q_i^o(k_i + \Delta, w_i, \epsilon_{it}) \\ & \quad - [R_i(q_i^o(k_i + \Delta, w_i, \epsilon_{it}) - \Delta, \epsilon_{it}) - w_i \cdot (q_i^o(k_i + \Delta, w, \epsilon) - \Delta)]. \end{aligned}$$

because  $q_i^o(k_i + \Delta, w_i, \epsilon_{it}) - \Delta \leq k_i$  if the division invested  $k_i + \Delta$  units of capacity. We also note that for  $\Delta$  sufficiently small,  $q_i^o(k_i + \Delta, w_i, \epsilon_{it}) - \Delta \geq 0$  because  $q_i^o(k_i, w_i, \epsilon_{it}) > 0$  by the assumption that  $R_i'(0, \epsilon_{it}) - w_i > 0$ .

By the Intermediate Value Theorem, the right-hand side of (24) is equal to

$$\frac{R_i'(\hat{q}_i(\Delta), w_i, \epsilon_i) \cdot \Delta}{\Delta} - w_i,$$

for some intermediate value  $\hat{q}_i(\Delta)$  such that  $q_i^o(k_i + \Delta, \cdot) - \Delta \leq \hat{q}_i(\Delta) \leq q_i^o(k_i + \Delta, \cdot)$ .

As  $\Delta \rightarrow 0$ , the right-hand side in both (26) and (24) converge to the following:

$$R_i'(q_i^o(k, w_i, \epsilon_{it}), \epsilon_{it}) - w_i,$$

proving the claim.

If  $k_i^o > 0$  is the optimal capacity level, then

$$\begin{aligned} \frac{\partial}{\partial k} [E_{\epsilon_i}[CM_i(k_i^o|w_i, \tilde{\epsilon}_i)] - c_i \cdot k_i^o] &= E_{\epsilon_i} \left[ \frac{\partial}{\partial k_i} CM_i(k_i^o|w_i, \tilde{\epsilon}_i) \right] - c_i \\ &= E_{\epsilon_i}[R'_i(q_i^o(k_i, \cdot), \tilde{\epsilon}_i)] - (c_i + w_i) \\ &= 0. \end{aligned}$$

Thus  $k_i^o$  satisfies equation (5) in the statement of Lemma 1.

To verify uniqueness, suppose that both  $k_i^o$  and  $k_i^o + \Delta$  satisfy equation (5). Since by definition  $q_i^o(k_i^o + \Delta, \cdot) \geq q_i^o(k_i^o, \cdot)$  for all  $\epsilon_{it}$ , it would follow that in fact

$$q_i^o(k_i^o + \Delta, \cdot) = q_i^o(k_i^o, \cdot)$$

for all  $\epsilon_{it}$ . That in turn would imply that the optimal production quantity in the absence of a capacity constraint, i.e.,  $\hat{q}_i(\epsilon_{it}, \cdot)$ , is less than  $k_i^o$ , and therefore

$$E_{\epsilon_i}[R'_i(\hat{q}_i(\tilde{\epsilon}_{it}, \cdot), \tilde{\epsilon}_{it})] = E_{\epsilon_i}[R'_i(q_i^o(k_i^o, \tilde{\epsilon}_{it}, w_i), \tilde{\epsilon}_{it})] = w_i,$$

which would contradict that  $k_i^o$  satisfies equation (5) in the first place.  $\square$

### Proof of Proposition 1:

Contingent on  $(k_1, k_2)$  and  $(q_{1t}, q_{2t}) \leq (k_1, k_2)$ , Division 1's residual income performance measure in period  $t$  is given by

$$\pi_{1t} = R_1(q_{1t}, \epsilon_{1t}) - w_1 \cdot q_{1t} - w_2 \cdot q_{2t} + TP(q_{2t}, k_2) - z_{1t} \cdot v_1 \cdot k_1 - z_{2t} \cdot v_2 \cdot k_2.$$

Regardless of the decisions made by Division 2, Division 1 will therefore choose the production quantity  $q_1^o(k_1, \cdot)$  that maximizes its contribution margin in period  $t$ . It is well known that if capacity assets are depreciated according to the annuity rate, then

$$z_{1t} \cdot v_1 = \frac{1}{\sum_t \gamma^t} \cdot v_1 = c_1.$$

In order to maximize  $E_{\epsilon_1}[\tilde{\pi}_{1t}]$  in any particular time period  $t$ , the initial capacity level  $k_1$  should be chosen so as to maximize the following objective function:

$$E_{\epsilon_1}[R_1(q_1^o(k_1, \cdot), \tilde{\epsilon}_{1t}) - w_1 \cdot q_1^o(k_1, \cdot)] - c_1 \cdot k_1.$$

This objective function is proportional to the objective function of the central office,  $\Gamma_1(k_1)$ , and the maximizing capacity level is  $k_1^o$ , as identified in Lemma 1.

For Division 2, the ex-post performance measure in period  $t$  is given by

$$\pi_{2t}(k_2, \epsilon_{2t}, w_2|\mu) = R_2(q_{2t}, \epsilon_{2t}) - TP(k_2, q_{2t}|\mu),$$

where

$$TP(k_2, q_{2t}|\mu) = (w_2 + c_2) \cdot q_{2t} + \mu \cdot (k_2 - q_{2t}).$$

We denote by  $\hat{q}_{2t}(k_2, \epsilon_{2t}, w_2|\mu)$  the maximizer of  $\pi_{2t}(k_2, \epsilon_{2t}, w_2|\mu)$ . Suppose Division 2 seeks an initial capacity level  $k_2$  so as to maximize its expected performance measure in any particular period  $t$ :

$$\pi_{2t}(k_2, \cdot|\mu) \equiv E_{\epsilon_2}[R_2(\hat{q}_{2t}(k_2, \tilde{\epsilon}_{2t}, w_2|\mu), \tilde{\epsilon}_{2t}) - TP(k_2, \hat{q}_{2t}(k_2, \tilde{\epsilon}_{2t}, w_2|\mu))].$$

Clearly,  $\pi_{2t}(k_2, \cdot|\mu) \leq \pi_{2t}(k_2, \cdot|c_2)$  for all  $k_2$ , since  $\mu \geq c_2$ . As shown in the proof of Lemma 1,

$$\pi_{2t}(k_2, \cdot|c) = \frac{1}{\sum_t \gamma^t} \cdot \Gamma_2(k_2) \equiv E_{\epsilon_2}[R_2(q_2^o(k_2, \cdot), \tilde{\epsilon}_{2t}) - w_2 \cdot q_2^o(k_2, \cdot)] - c_2 \cdot k_2.$$

By definition,  $\Gamma_2(k_2)$  is maximized at  $k_2^o$ , and by the limited volatility condition,

$$\frac{1}{\sum_t \gamma^t} \cdot \Gamma_2(k_2^o) = E_{\epsilon_2}[R_2(k_2^o, \tilde{\epsilon}_{2t})] - (w_2 + c_2) \cdot k_2^o.$$

For any  $\mu \geq c_2$ ,  $\hat{q}_{2t}(k_2, \epsilon_{2t}, w_2|\mu) \geq q_2^o(k_2, \cdot)$ . Thus,

$$\pi_{2t}(k_2, \cdot|\mu) \leq \pi_{2t}(k_2, \cdot|c_2) = \frac{1}{\sum_t \gamma^t} \cdot \Gamma_2(k_2) \leq \frac{1}{\sum_t \gamma^t} \cdot \Gamma_2(k_2^o) = \pi_{2t}(k_2^o, \cdot|\mu),$$



proving that for any  $\mu \geq c_2$ , Division 2 will choose  $k_2 = k_2^o$  regardless of the weights it attaches to its performance measure in different periods.  $\square$

**Proof of Proposition 2:**

Given a two-part full cost transfer pricing mechanism of the form  $TP_t(k_2, q_{2t}) = c_2 \cdot k_2 + w_2 \cdot q_{2t}$ , the claim follows directly from the arguments given in the proof of Proposition 1. In particular, the arguments for goal congruence for Division 2 now exactly parallels the one given for Division 1 in the previous result.  $\square$

**Proof of Lemma 2:** The proof proceeds along the lines of the proof of Lemma 1. In particular

$$CM_{it}(k_i|w_{it}, \epsilon_{it}) = \max_{q_{it} \leq k_i} \{x_{it} \cdot R_i(q_{it}, \epsilon_{it}) - w_{it} \cdot q_{it}\}$$

is differentiable in  $k_i$  and

$$CM'_{it}(k_i|w_{it}, \epsilon_{it}) = x_{it} \cdot R'_i(q_{it}^o(k_i, \cdot), \epsilon_{it}) - w_{it}.$$

We can interchange the order of differentiation and integration to conclude that the firm's objective function  $\Gamma_i(k_i)$  is differentiable with derivative:

$$\Gamma'_i(k_i) = \sum_{t=1}^T E_{\epsilon_{it}} [x_{it} \cdot R'_i(q_{it}^o(k_i, \cdot), \tilde{\epsilon}_{it}) - w_{it}] \cdot \gamma^t - v_i.$$

Given the definition of  $\bar{w}_i \equiv \sum_{t=1}^T w_{it} \cdot \gamma^t$  in the statement of Lemma 2, the first order condition in the statement of Lemma 2 now follows immediately.

**Proof of Corollary to Proposition 2:**

The firm's objective function for Division 2 is to maximize

$$\sum_{t=1}^T E_{\epsilon_2} [CM_2(k_2|\tilde{\epsilon}_{2t}, w_{2t})] \cdot \gamma^t - v_2 \cdot k_2,$$

where

$$CM_2(k_2|\epsilon_{2t}, w_{2t}) = \max_{q_{2t} \leq k_2} \{R_2(q_{2t}, \epsilon_{2t}) - w_{2t} \cdot q_{2t}\} = R_2(q_{2t}^o(k_2, \cdot), \epsilon_{2t}) - w_{2t} \cdot q_{2t}^o(k_2)$$

If  $w_{2t} = x_{2t} \cdot w_2$ , the above objective function reduces to

$$\sum_{t=1}^T E_{\epsilon_2} [x_{2t} \cdot CM_2(k_2|\tilde{\epsilon}_{2t}, w_{2t})] \cdot \gamma^t - v_2 \cdot k_2$$

Given the transfer pricing rule

$$TP_t(k_2, q_{2t}) = \hat{z}_{2t} \cdot v_2 \cdot k_2 + w_{2t} \cdot q_{2t},$$

the expected profit for Division 2 in period  $t$  is given by

$$E_{\epsilon_2} [x_{2t} \cdot R_2(q_{2t}^o(k_2, \cdot), \tilde{\epsilon}_{2t}) - x_{2t} \cdot w_2 \cdot q_{2t}^o(k_2, \cdot)] - \hat{z}_{2t} \cdot v_2 \cdot k_2.$$

Since  $\hat{z}_{2t} = \frac{x_{2t}}{\sum_{\tau=1}^T x_{2\tau} \cdot \gamma^\tau}$ , Division 2's objective function in period  $t$  is proportional to the firm's overall objective function,  $\Gamma_2(k_2)$ , and thus Division 2 will choose the optimal capacity level  $k_2^o$  at the initial stage.

### Proof of Proposition 3:

Given the limited volatility condition, the optimal  $k_i^o$  is such that

$$\begin{aligned} \Gamma_i(k_i^o) &= E_{\epsilon_i} \left[ \sum_{\tau=1}^T \gamma^\tau \cdot x_{i\tau} \cdot R_i(k_i^o, \tilde{\epsilon}_{i\tau}) \right] - (v_i + \bar{w}_i) \cdot k_i^o \\ &\geq \Gamma_i(k_i) \\ &= E_{\epsilon_i} \left[ \sum_{\tau=1}^T \gamma^\tau \cdot x_{i\tau} \cdot R_i(k_i, \tilde{\epsilon}_{i\tau}) \right] - (v_i + \bar{w}_i) \cdot k_i \end{aligned} \tag{25}$$

for all  $k_i$ .

We next show that in order to maximize its expected profit in period  $t$ , the downstream division would choose  $k_2^o$ . To see this, we recall that because  $TP(k_2) = \hat{z}_{2t} \cdot (v_2 + \bar{w}_2) \cdot k_2$ , the expected profit for Division 2 in period  $t$  bis given by

$$E_{\epsilon_2}[\tilde{\pi}_{2t}(k_2)] = E_{\epsilon_2}[x_{2t} \cdot R_2(q_{2t}^o(k_2, \cdot), \tilde{\epsilon}_{2t})] - \hat{z}_{2t} \cdot (v_2 + \bar{w}_2) \cdot k_2.$$

Clearly  $q_{2t}^o(k_2, \cdot) = k_2$  as Division 2 is not charged any variable costs. We recall that  $\hat{z}_{2t} = \frac{x_{2t}}{\sum_{\tau=1}^T x_{2\tau} \cdot \gamma^\tau}$ , and therefore

$$E_{\epsilon_2}[\tilde{\pi}_{2t}(k_2)] = z_{2t} \cdot \left[ \sum_{t=1}^T x_{2\tau} \cdot \gamma^\tau \cdot E_{\epsilon_2}[R_2(k_2, \tilde{\epsilon}_{2t})] - (v_2 + \bar{w}_2) \cdot k_2 \right],$$

which, according to (25), is maximized at  $k_2^o$ .

The argument for Division 1 is the same since under relative benefit depreciation rule the capital charge to Division 1 in period  $t$  is  $\hat{z}_{1t}(v_1 + \bar{w}_1) \cdot k_1$  if Division 1 invested in  $k_1$  units of capacity.  $\square$

#### Proof of Proposition 4:

As defined in the main text, the threshold type  $\theta_i^*$  is the one achieving a zero NPV for the capacity investment  $\bar{k}_i$ ; that is,

$$\sum_{\tau=1}^T E_{\epsilon_i} [CM_{i\tau}(\bar{k}_i | x_{i\tau}, w_{i\tau}, \tilde{\epsilon}_{i\tau}) | \theta_i^*] \cdot \gamma^\tau = v_i \cdot \bar{k}_i.$$

We note that the net present value,  $\Gamma(\bar{k}_i | \theta_i)$ , is increasing in  $\theta_i$ . This follows directly from Theorem 6D1 in Mas-Colell et al. (1995) because  $\theta_i$  shifts the densities  $f_i(\cdot | \theta_i)$  in the sense of first-order stochastic dominance and  $CM_{it}(\bar{k}_i | x_{it}, w_{it}, \epsilon_{it})$  is increasing in  $\epsilon_{it}$ . If the downstream division were to focus exclusively on its profit measure in period  $t$ ,  $1 \leq t \leq T$ , it would be seek to maximize the following:

$$E_{\epsilon_2} [\pi_2(k | \theta_2, \tilde{\epsilon}_{2t})] \equiv E_{\epsilon_2} [CM_{2t}(k_2 | x_{2t}, w_{2t}, \tilde{\epsilon}_{2t}) | \theta_2] - \bar{z}_{2t} \cdot v_2 \cdot k_2.$$

Direct substitution for  $\bar{z}_{2t}$  according to the REOB rule shows that

$$E_{\epsilon_2} [\pi_2(k | \theta_2, \tilde{\epsilon}_{2t})] = E_{\epsilon_2} [CM_{2t}(\bar{k}_2 | x_{2t}, w_{2t}, \tilde{\epsilon}_{2t}) | \theta_2] - E_{\epsilon_2} [CM_{2t}(\bar{k}_2 | x_{2t}, w_{2t}, \tilde{\epsilon}_{2t}) | \theta_2^*],$$

which will be greater than zero if and only if  $\theta_2 > \theta_2^*$ .  $\square$

**Proof of Lemma 3:**

The maximized contribution margin is given by

$$CM(k|w, \epsilon_t) = \sum_{i=1}^2 [R_i(q_i^*(k, \cdot), \epsilon_{it}) - w_i \cdot q_i^*(k, \cdot)]$$

where  $q_i^*(k, \cdot) \equiv q_i^*(k, w, \epsilon_t)$

**Claim:**  $CM(k|w, \epsilon_t)$  is differentiable in  $k$  for any  $\epsilon_t$  and  $w$ , such that

$$\frac{\partial}{\partial k} CM(k|w, \epsilon_t) = S(k|w, \epsilon_t),$$

where

$$S(k|w, \epsilon_t) = \max\{R'_1(q_1^*(k, \cdot), \epsilon_{1t}) - w_1, R'_2(q_2^*(k, \cdot), \epsilon_{2t}) - w_2\}.$$

We distinguish three cases:

Case 1:  $0 < q_1^*(k, \cdot) < k$

It follows that  $q_2^*(k, \cdot) > 0$  and

$$S(k|w, \epsilon_t) = R'_1(q_1^*(k, \cdot), \epsilon_{1t}) - w_1 = R'_2(q_2^*(k, \cdot), \epsilon_{2t}) - w_2.$$

We then claim that for  $\Delta \geq 0$  sufficiently small,

$$\frac{CM(k + \Delta|w, \epsilon_t) - CM(k|w, \epsilon_t)}{\Delta} \geq S(k|w, \epsilon_t). \quad (26)$$

Like in the proof of Lemma 1, this inequality is derived from observing that

$$CM(k + \Delta|w, \epsilon_t) \geq R_1(q_1^*(k, \cdot) + \Delta, \epsilon_{1t}) - w_1(q_1^*(k, \cdot) + \Delta) + R_2(q_2^*(k, \cdot), \epsilon_{2t}) - w_2 \cdot q_2^*(k, \cdot).$$

Therefore, the left-hand side of (26) is at least as large as the following expression:

$$\frac{R_1(q_1^*(k, \cdot) + \Delta, \epsilon_{1t}) - R_1(q_1^*(k, \cdot), \epsilon_{1t})}{\Delta} - w.$$

As  $\Delta \rightarrow 0$ , this expression converges to

$$R_1'(q_1^*(k, \cdot), \epsilon_{1t}) - w_1 = S(k|w, \epsilon_t).$$

Following the same line of arguments as in the proof of Lemma 1, we also find that

$$\frac{CM(k + \Delta|w, \epsilon_t) - CM(k|w, \epsilon_t)}{\Delta} \leq S(k|w, \epsilon_t)$$

for  $\Delta$  sufficiently small and thus

$$\frac{\partial}{\partial k} CM(k|w, \epsilon_t) = S(k|w, \epsilon_t).$$

Case 2:  $q_1^*(k, \cdot) = k$ .

In this case  $q_2^*(k, \cdot) = 0$  and  $S(k|w, \epsilon_t) = R_1'(q_1^*(k, \cdot), \epsilon_{1t}) - w_1$ . Using the same arguments as in Case 1, it can then be shown that  $\frac{\partial}{\partial k} CM(k|w, \epsilon_t) = S(k|w, \epsilon_t)$ .

Case 3:  $q_2^*(k, \cdot) = k$ .

In this case,  $q_1^*(k, \cdot) = 0$  and  $S(k|w, \epsilon_t) = R_2'(q_2^*(k, \cdot), \epsilon_{2t}) - w_2$ . For  $\Delta \geq 0$  sufficiently small, it can again be shown that

$$\frac{CM(k + \Delta|w, \epsilon) - CM(k|w, \epsilon)}{\Delta} \geq S(k|w, \epsilon_t).$$

To see this, note that

$$CM(k + \Delta|w, \epsilon) \geq R_1(q_1^*(k, \cdot) + \Delta, \epsilon_{1t}) - w_1 \cdot q_1^*(k, \cdot) + R_2(q_2^*(k, \cdot) + \Delta, \epsilon_{2t}) - w_2 \cdot [q_2^*(k, \cdot) + \Delta].$$

Therefore,  $\frac{CM(k + \Delta|w, \epsilon) - CM(k|w, \epsilon)}{\Delta}$  is at least as large as the following expression:

$$\frac{R_2(q_2^*(k, \cdot) + \Delta, \epsilon_{2t}) - R_2(q_2^*(k, \cdot), \epsilon_{2t})}{\Delta} - w.$$

As  $\Delta \rightarrow 0$ , this expression converges to

$$R_2'(q_2^*(k, \cdot), \epsilon_{2t}) - w_2 = S(k|w, \epsilon_t).$$

Following the same line of arguments as used in the proof of Lemma 1, it can also be verified that

$$\frac{CM(k + \Delta|w, \epsilon_t) - CM(k|w, \epsilon_t)}{\Delta} \leq S(k|w, \epsilon_t)$$

for  $\Delta$  sufficiently small, and thus

$$\frac{\partial}{\partial k} CM(k|w, \epsilon_t) = S(k|w, \epsilon_t).$$

The expected value of the maximized contribution margin,  $E_\epsilon [CM(k|w, \tilde{\epsilon}_t)]$ , is identical across periods in the stationary setting. Hence, the firm will choose the optimal capacity level  $k^*$  to maximize the following objective function:

$$E_\epsilon [CM(k|w, \tilde{\epsilon}_t)] - c \cdot k.$$

Equation (14) in the statement of Lemma 3 then follows from the first-order condition of the above optimization problem. The uniqueness of  $k^*$  follows from a similar argument as used in the proof of Lemma 1.  $\square$

**Proof of Proposition 5:**

We first show that with quadratic revenue functions of the form  $R_i(q, \theta_i, \epsilon_{it}) = \theta_i \cdot \epsilon_{it} \cdot q - h_i \cdot q^2$  and limited volatility, the efficient capacity level in the fungible scenario is equal to the sum of the efficient capacity levels in the dedicated capacity scenario; that is

$$k^* = k_1^o + k_2^o.$$

From Lemma 1, we know that in the dedicated capacity setting the efficient capacity levels satisfy:

$$E_{\epsilon_i} [R'_i(q_1^o(k_i^o, \tilde{\epsilon}_{it}), \theta_i, \tilde{\epsilon}_{it})] - w_i = c.$$

The limited volatility condition implies that for all  $\tilde{\epsilon}_{it}$

$$q_i^o(k_i^o, \tilde{\epsilon}_{it}) = k_i^o.$$

Furthermore, with quadratic revenue functions, we find that

$$\begin{aligned}
E_{\epsilon_i}[R'_i(q_i^o(k_i^o, \tilde{\epsilon}_{it}), \theta_i, \tilde{\epsilon}_{it})] - w_i &= E_{\epsilon_i}[R'_i(k_i^o, \theta_i, \tilde{\epsilon}_{it})] - w_i \\
&= R'_i(k_i^o, \theta_i, \hat{\epsilon}_{it}) - w_i \\
&= c,
\end{aligned} \tag{27}$$

where  $\hat{\epsilon}_{it} \equiv E(\tilde{\epsilon}_{it})$ . In the fungible capacity scenarios, Lemma 3 has shown that at the efficient  $k^*$

$$E_{\epsilon}[R'_i(q_i^*(k^*, \cdot), \theta_i, \tilde{\epsilon}_{it})] - w_i = c.$$

It is readily seen that in the quadratic revenue scenario,  $q_i^*(k^*, \theta, w, \epsilon_t)$  is linear in  $\epsilon_t$  provided that  $q_i^*(\cdot) > 0$  for all  $\epsilon_t$ . Thus,

$$\begin{aligned}
E_{\epsilon}[S(k^*|\theta, w, \tilde{\epsilon}_t)] &= E_{\epsilon}[R'_i(q_i^*(k^*, \cdot), \theta_i, \tilde{\epsilon}_{it})] - w_i \\
&= E_{\epsilon}[\theta_i \cdot \tilde{\epsilon}_{it} - 2h_i \cdot q_i^*(k^*, \theta, w, \tilde{\epsilon}_t)] - w_i \\
&= \theta_i \cdot \hat{\epsilon}_{it} - 2h_i \cdot q_i^*(k^*, \theta, w, \hat{\epsilon}_t) - w_i \\
&= R'_i(q_i^*(k^*, \theta, w, \hat{\epsilon}_t), \theta_i, \hat{\epsilon}_{it}) - w_i \\
&= c,
\end{aligned} \tag{28}$$

where  $\hat{\epsilon}_t \equiv E(\epsilon_t)$ . It follows from (27) and (28) that

$$q_i^*(k^*, \theta, w_i, \hat{\epsilon}_t) = k_i^o,$$

and thus  $k^* = k_1^o + k_2^o$ .

It remains to show that  $(k_1^o, k_2^o)$  is a Nash equilibrium at the initial date. Given the full cost transfer pricing rule  $TP_t(k_2, q_{2t}) = c \cdot k_2 + w_2 \cdot q_{2t}$ , the divisional profit of Division 2 in period  $t$ , contingent on  $\epsilon_t$  and  $k_1^o$  is:

$$\pi_{2t}(k_2, \epsilon_t | k_1^o) = \delta \cdot CM_2(k_2 | \theta_2, w_2, \epsilon_{2t}) + (1 - \delta)[CM(k_1^o + k_2 | \theta, w, \epsilon_t) - CM_1(k_1^o | \theta_1, w_1, \epsilon_t)] - c \cdot k_2,$$

where, as before,

$$CM_i(k_i | \theta_i, w_i, \epsilon_{it}) = \max_{q_i \leq k_i} \{R_i(q_i, \theta_i, \epsilon_{it}) - w_i \cdot q_i\}.$$

We note that in a stationary environment, the expected value of Division 2's profit,  $E_\epsilon[\pi_{2t}(k_2, \tilde{\epsilon}_t|k_1^o)]$ , is the same in each period. By definition,  $k_2^o$  is the unique maximizer of  $E_{\epsilon_2}[\delta \cdot CM(k_2|\theta_2, w_2, \tilde{\epsilon}_{2t})] - \delta \cdot c \cdot k_2$ . By Lemma 3,  $k_2^o$  maximizes  $E_\epsilon[(1 - \delta) \cdot CM(k_1^o + k_2|\theta, w, \tilde{\epsilon}_t)] - (1 - \delta) \cdot c \cdot k_2$ . It thus follows that  $k_2^o$  is also a maximizer of Division 2's expected profit in each period,  $E_\epsilon[\pi_{2t}(k_2, \tilde{\epsilon}_t|k_1^o)]$ .

A symmetric argument can be used to show that in order to maximize its expected residual income in any period, Division 1 will choose  $k_1^o$  if it conjectures that the downstream division chooses  $k_2^o$ .  $\square$

### Proof of Proposition 6:

Suppose that the two divisions has agreed to an ex-ante contract under which the downstream division has initial rights for  $k_2$  units of capacity for a transfer payment of  $p(k_2) + w_2 \cdot q_{2t}$  in each period. Further, suppose that Division 1 has installed a capacity of  $k_1$  units over which it has unilateral rights.

After observing  $\epsilon_t$ , the two divisions will renegotiate the initial capacity rights to maximize the joint surplus in each period. Following the same arguments as used in deriving (17), it can be checked that Division 1's effective contribution margin after reallocation of capacity rights is given by

$$CM_1^*(k_1+k_2|w, \epsilon_t) = (1-\delta) \cdot CM_1(k_1|w_1, \epsilon_{1t}) + \delta \cdot [CM(k_1 + k_2|w, \epsilon_t) - CM_2(k_2|w_2, \epsilon_{2t})].$$

For stationary environments, the expected value of effective contribution margin,  $E_\epsilon [CM_1^*(k|w, \epsilon_t)]$ , is the same in each period. Since capacity assets are depreciated according to the annuity rule, this implies that taking  $k_2$  as given, Division 1 will choose  $k_1$  to maximize

$$E_\epsilon [CM_1^*(k_1 + k_2|w, \tilde{\epsilon}_t)] - c \cdot k.$$

As a function of  $k_2$ , let  $r_1(k_2)$  denote Division 1's optimal response; i.e.,  $k_1 = r(k_2)$  maximizes the above objective function. Let  $r(k_2) \equiv r_1(k_2) + k_2$  denote the corresponding aggregate amount of capacity. Division 1's reaction function,  $r_1(k_2)$ ,



will satisfy the following first-order condition:

$$E_\epsilon \left[ (1 - \delta) \cdot CM'_1(r_1(k_2)|w_1, \tilde{\epsilon}_{1t}) + \delta \cdot S(r(k_2)|w, \tilde{\epsilon}_t) \right] \leq c, \quad (29)$$

which must hold as an equality whenever  $r_1(k_2) > 0$ . We note from (29) that  $r_1(k_2)$  is downward sloping because  $CM'_1(k_1|\cdot)$  and  $S(k|\cdot)$  are decreasing functions of  $k_1$  and  $k$ , respectively.

We now investigate the values of  $r_1(k_2)$  at  $k_2 = 0$  and  $k_2 = k^*$ . We first claim that  $r_1(0) \leq k^*$ . Suppose to the contrary,  $r_1(0) > k^*$ . This implies that  $r(0) > k^*$ , and hence

$$E_\epsilon[S(r(0)|w, \tilde{\epsilon}_t)] < E_\epsilon[S(k^*|w, \tilde{\epsilon}_t)] = c. \quad (30)$$

Furthermore,

$$\begin{aligned} E_{\epsilon_1} [CM'_1(r_1(0)|w_1, \tilde{\epsilon}_{1t})] &< E_{\epsilon_1} [CM'_1(k^*|w_1, \tilde{\epsilon}_{1t})] \\ &= E_{\epsilon_1} [R'_1(q_1^o(k^*, \tilde{\epsilon}_{1t}, \cdot), \tilde{\epsilon}_{1t}) - w_1] \\ &\leq E_\epsilon [R'_1(q_1^*(k^*, \tilde{\epsilon}_t, \cdot), \tilde{\epsilon}_{1t}) - w_1] \\ &\leq E_\epsilon[S(k^*|w, \tilde{\epsilon}_t)] \\ &= c, \end{aligned} \quad (31)$$

where we have used the result that  $q_1^o(k^*, \epsilon_{1t}, \cdot) \geq q_1^*(k^*, \epsilon_t, \cdot)$  for all  $\epsilon_t$  to derive the second inequality above. Inequalities in (30) and (31) imply that the first-order condition in (29) cannot hold as an equality, which contradicts the assumption that  $r_1(0) > k^*$  is optimal.

We next claim that  $r_1(k^*) > 0$ , and hence  $r(k^*) > k^*$ . Suppose to the contrary  $r_1(k^*) = 0$ . This implies that  $r(k^*) = k^*$ , and hence

$$E_\epsilon[S(r(k^*)|w, \tilde{\epsilon}_t)] = c.$$

Furthermore,

$$E_{\epsilon_1}[CM'_1(r_1(k^*)|w_1, \tilde{\epsilon}_{1t})] = E_{\epsilon_1}[CM'_1(0|w_1, \tilde{\epsilon}_{1t})] > c,$$

because of the assumption in (21). It thus follows that the left hand side of (29) is strictly greater than  $c$ , which contradicts the assumption that  $r(k_2) = 0$  is the optimal response to  $k_2 = k^*$ .

We have thus proven that  $r(0) \leq k^*$  and  $r(k^*) > k^*$ . The Intermediate Value Theorem then implies that there exists a  $k_2^* \in [0, k^*)$  such that  $r(k_2^*) = k^*$ . We have thus shown that if the two divisions sign an ex-ante contract that provides the downstream division with initial capacity rights of  $k_2^*$  units, Division 1 will choose the efficient amount of aggregate capacity  $k^*$ .

To complete the proof, we need to show that there exists a fixed transfer payment  $p(k_2^*)$  such that the ex-ante contract  $(k_2^*, p(k_2^*))$  will be preferred by both divisions to the default point of no agreement. If the two divisions fail to reach an ex-ante agreement, Division 1 will choose its capacity level unilaterally, and Division 2 will receive no capacity rights (i.e.,  $k_2 = 0$ ). Let  $\hat{k}$  denote Division 1's optimal choice of capacity under the "default" scenario. Division 1's expected periodic payoff under the default scenario is then given by

$$\hat{\pi}_1 = E_\epsilon \left[ (1 - \delta) \cdot CM_1(\hat{k}|w_1, \tilde{\epsilon}_{1t}) + \delta \cdot CM(\hat{k}|w, \tilde{\epsilon}_t) \right] - c \cdot \hat{k}$$

while Division 2's default payoff is

$$\hat{\pi}_2 = (1 - \delta) \cdot E_\epsilon \left[ CM(\hat{k}|w, \tilde{\epsilon}_t) - CM_1(\hat{k}|w_1, \tilde{\epsilon}_{1t}) \right].$$

By agreeing to transfer  $k_2^*$  units of capacity rights to Division 2, the two divisions can increase their periodic joint surplus by

$$\Delta\pi \equiv E_\epsilon [CM(k^*|w, \tilde{\epsilon}_t) - c \cdot k^*] - E_\epsilon [CM(\hat{k}|w, \tilde{\epsilon}_t) - c \cdot \hat{k}].$$

The two divisions can then split this additional surplus between them in proportion to their relative bargaining power. The periodic transfer price  $p(k_2^*)$  that implements this allocation is given by

$$E_\epsilon [(1 - \delta) \cdot CM_1(k^* - k_2^*|w_1, \tilde{\epsilon}_{1t}) + \delta \cdot CM(k^*|w, \tilde{\epsilon}_t)] - c \cdot k^* + p(k_2^*) = \hat{\pi}_1 + \delta \cdot \Delta\pi.$$

Division 2's expected periodic payoff with this choice of transfer payment will be equal to  $\hat{\pi}_2 + (1 - \delta) \cdot \Delta\pi$ . Therefore, both divisions will prefer the ex-ante contract  $(k_2^*, p(k_2^*))$  to the default scenario of no contract  $(0, 0)$ .

## References

- Anctil, R. and S. Dutta (1999). Negotiated Transfer Pricing and Divisional vs. Firm-Wide Performance Evaluation. *The Accounting Review* 74(1), 87-104.
- Arrow, K. (1964). Optimal Capital Policy, The Cost of Capital and Myopic Decision Rules. *Annals of the Institute of Statistical Mathematics* 1-2, 21-30.
- Balachandran, B., Balakrishnan, R. and S. Sivaramakrishnan (1997). On the Efficiency of Cost-Based Decision Rules for Capacity Planning. *The Accounting Research* 72(4), 599-619.
- Balakrishnan, R. and S. Sivaramakrishnan (2002). A Critical Overview of Full-Cost Data for Planning and Pricing. *Journal of Management Accounting Research* 3-31.
- Baldenius, T., Dutta, S. and S. Reichelstein (2007). Cost Allocations for Capital Budgeting Decisions. *The Accounting Review* 82(4), 837-867.
- Baldenius, T., Nezlobin, A., and I. Vaysman (2016). Managerial Performance Evaluation and Real Options. *The Accounting Review* 91(3), 741-766.
- Baldenius, T. and S. Reichelstein, and S. Sahay (1999). Negotiated versus Cost-Based Transfer Pricing. *The Review of Accounting Studies* 4(2), 67-91.
- Banker, R. and J. Hughes (1994). Product Costing and Pricing. *The Accounting Review* 69(3), 479-494.
- Bastian, N. and S. Reichelstein (2004). Transfer Pricing at Timken. Stanford GSB, Case # A-190.
- Bouwens, J and B. Steens (2016). Full-Cost Transfer Pricing and Cost Management. *Journal of Management of Accounting Research* 28(3), 63-81.
- Carlton, S., & Perloff, J. (2005). *Advanced Industrial Organization*. 4th ed. New York, NY: Pearson/Addison Wesley.
- Datar, S. and M. Rajan. (2014). *Managerial Accounting: Making Decisions and Motivating Performance*. Pearson, NY: Pearson/Addison Wesley.
- Dutta, S. (2008). Dynamic Performance Measurement. *Foundations and Trends in Accounting*, 2(3), 175-240.

- Dutta, S., and S. Reichelstein (2002). Controlling Investment Decisions: Depreciation and Capital Charges. *Review of Accounting Studies* 7, 253-281.
- Dutta, S. and S. Reichelstein (2010). Decentralized Capacity Management and Internal Pricing. *Review of Accounting Studies*, 15(3), 442-478.
- Eccles, R., and H. White (1988). Price and Authority in Inter-Profit Center Transactions. *American Journal of Sociology* (94) (Supplement), 17-51.
- Edlin, A. and S. Reichelstein (1995). Negotiated Transfer Pricing: An Efficiency Result. *The Accounting Review* 69(3), 479-494.
- Ernst & Young (2003), *Global Transfer Pricing Survey*. www.ey.com.
- Feinschreiber, R. and M. Kent. (2012). *Transfer Pricing Handbook*. Wiley Publishers.
- Goex, R. (2002), Capacity Planning and Pricing under Uncertainty. *Journal of Management Accounting Research* 59-79.
- Goex, R. and U. Schiller (2007). An Economic Perspective on Transfer Pricing. *Handbook of Management Accounting Research*. C.S. Chapman, A. Hopwood and M. Shields (eds).
- Gramlich, J. and K. Ray (2016). Reconciling full-Cost and Marginal-Cost Transfer Pricing. *Journal of Management Accounting Research* 28(1), 27-37.
- Hotelling, H. (1925). A General Mathematical Theory of Depreciation. *Journal of the American Statistical Association* 20, 340-353.
- Kaplan, R. (2006). Activity-based Costing and Capacity. *Harvard Business School*, Case # 9-105-059
- Martinez-Jerez, A. (2007). Understanding Customer Profitability at Charles Schwab. *Harvard Business School*, Case # 106-002.
- Mas-Colell, A., Whinston, M. and J. Green (1995), *Microeconomic Theory*, Oxford University Press, Oxford, New York.
- Nezlobin, A., Rajan, M and S. Reichelstein (2012). Dynamics of Rate of Return Regulation. *Management Science* 58(5), 980-995.
- Pfeiffer, T., Lengsfeld, S., Schiller, U. and J. Wagner (2009). Cost-based Transfer Pricing. *Review of Accounting Studies* 7, 253-281.

- Rajan, M. and S. Reichelstein (2009). Depreciation Rules and The Relation Between Marginal and Historical Cost. *Journal of Accounting Research* 47(3),1-43.
- Reichelstein, S. and A. Rohlfsing-Bastian (2015). Levelized Product Cost: Concept and Decision Relevance. *The Accounting Review* 90 (4), 1653-1682.
- Reichelstein, S. and A. Sahoo (2018). Relating Long-Run Marginal Cost to Product Prices: Evidence from Solar Photovoltaic Modules. *Contemporary Accounting Research* forthcoming.
- 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.
- Rogerson, W. (2011). On the relationship between historic cost, forward looking cost and long run marginal cost. *Review of Network Economics* 10(2), 1-31.
- Sahay, S. (2002). Transfer Pricing Based on Actual Cost. *Journal of Management Accounting Research* 15, 177-192.
- Simons, R. (2000). Polysar Limited. *Harvard Business School Case # 9-187-098*.
- Solomons, D. (1964), *Divisional Performance Measurement and Control*. Irvin. Homewood, Illinois.
- Tang, R. (2002). *Current Trends and Corporate Cases in Transfer Pricing*. Westport, CT: Quorum Books.
- Wielenberg, S. (2000). Negotiated Transfer Pricing, Specific Investment and Optimal Capacity Choice. *Review of Accounting Studies* 5, 197-216.
- Zimmerman, J. (2016), *Accounting for Decision Making and Control*. 8th ed., McGraw Hill, New York.