Production, Capacity, and Liquidity of a Self-Financed Firm

Jie Ning

Weatherhead School of Management, Case Western Reserve University, Jie.Ning@Case.edu

Matthew J. Sobel

Weatherhead School of Management, Case Western Reserve University, Matthew.Sobel@Case.edu

March 3, 2015

Abstract

We develop and analyze a dynamic stochastic model of a self-financed firm that optimizes its expected present value of dividends, profits, or a mixture of both. Each period the firm chooses how much to produce, to invest in capacity expansion, to distribute as a dividend, and to retain as liquidity subject to constraints from existing capacity and capital. Product-market prices and yields of capacity-enhancing investments are Markov-modulated. We show that linearity assumptions regarding investment yield and production cost imply that the value function has an affine structure. This leads to a complete characterization of an optimal policy and a linear program that exorcises the curse of dimensionality. We clarify the linkage between capacity and capital in a dynamic framework and provide real option interpretations. We prove that profit maximization induces a more aggressive investment attitude than does maximization of dividends.

Key words: Production, capacity, liquidity, self-financed, real option, risk

1 Introduction

1.1 A firm's assets and decisions in a risky environment

A firm's operations generally consist of production, which creates and delivers products yielding streams of revenues and costs; and investment, which depletes capital and transforms it into productive capacity. Capacity resulting from investment is either physical, such as the procurement of production equipment and the development of natural resource reserves, or intellectual, such as technology and research and development (R&D) of new products, or both physical and intellectual, such as an expansion or training of the work force.

Regardless of their specifics, different forms of capacity share three fundamental features. First, capacity is an asset that enables production which generates earnings. (We use "asset" broadly to mean a resource that enables a cash flow.) Second, the level of capacity sets the maximum production rate. In this sense, production is constrained by existing capacity. Third, capacity depreciates over time. Therefore, absent additional investment, the maximum production rate diminishes over time.

Some depreciation occurs merely because time passes, and is independent of the production rate (hereafter, the "natural" capacity depreciation). For example, technology-dependent capacity decreases as new technologies are developed and adopted in the industry. Other depreciation depends on the production rate (hereafter, the "production-induced" capacity depreciation). For example, the maximum daily output of an oil field falls as continuing extraction reduces the underground pressure which enables the oil outflow. Some depreciation is a mixture of the two types. For example, the output capacity of a machine may diminish over the years due to the normal production wear-and-tear, and technology becoming outdated.

The fact that capacity is an asset that constrains production and depreciates over time makes it necessary for a firm to restore and expands its capacity in order to develop and grow. Restoration and expansion, namely investment, usually depletes another type of asset—capital. Although capital enables investment, it also constrains it, just as capacity enables and constrains production. Further, production feeds revenue to capital but can deplete capacity, just as investment rebuilds capacity but drains capital. Therefore, the firm's two types of assets, capital and capacity, are simultaneously affected in opposite ways by its operational decisions regarding production and investment. In turn, both kinds of decisions are linked through their influences on the two types of assets.

Besides their roles in transforming the two assets, production and investment often share another important feature—the uncertainty of their outputs. Although the firm can decide at what level it will produce (subject to its capacity constraint), the amount of capital that will be generated is determined by the exogenous market. Similarly, although the firm can specify how much capital to use for investment, it may not be able to control fully the amount of capacity that will be installed. This is particularly true in the case of investment in R&D projects, which are critical to some businesses' success. Companies in oil and natural gas industries, for example, rely on exploring new reserves, and firms in pharmaceuticals and electronics engage in patent races. In general, what production and investment decisions the firm makes, and thus how it manages the two assets, is affected by these random outputs. Collectively the risky environment, production, and investment shape the firm's evolution.

Of course, there are additional factors that determine the firm's evolution, most notably its capital management decisions. These decisions dynamically adjust the capital structure by issuing and repurchasing debt and equity. A complete characterization of the firm's evolution can be achieved only by considering simultaneously both operational and financial decisions. In this paper, we consider the dynamic coordination of these decisions in a risky environment by a *self-financed* firm.

We use *self-financed* to refer to a firm that is and will remain all-equity financed, and which will not issue further equity. There are various possible reasons for such a state of affairs, the most important being financial market frictions. For example, the cost of raising additional capital, either in the form of debt or equity, may be prohibitively high. Since outside capital injection is not possible, the capital management decisions of a self-financed firm reduce to either retaining the earnings, i.e., increasing liquidity, or distributing it back to shareholders, i.e., issuing dividends. Henceforth, we refer to capital that has not been designated as a dividend or an investment as "liquidity."

The model of a *self-financed* firm is motivated by the empirical observation that companies tend to finance their innovative projects and R&D activities using internal cash reserves rather than external sources due to financial market frictions (Bates et al. 2009, Ma et al. 2013, Falato et al. 2013). Therefore, studying a self-financed firm illuminates dynamic interactions between financial and operational decisions in the presence of random investment outcomes.

1.2 The goal of the firm

The presumed goal of the firm is a key assumption in studying the dynamic coordination of operational and financial decisions. The goal that has dominated the operations literature is the maximization of the expected present value (EPV) of future profits. The finance literature, in contrast, asserts that the primary measure of a self-financed firm's value is the EPV of future dividends (Cochrane 2005), and thus its maximization should be the ultimate goal of management.

These two criteria may be consistent in an idealized perfect capital market free of "frictions" (e.g., financial distress costs, taxes, and asymmetric information) (Modigliani and Miller 1958, Allen and Michaely 1995), but they may induce different policies when market frictions are present.

In this paper and in practice, financial market frictions are manifest when a firm's risky investments are self-financed. Hence, the tractability of a construct of a self-financed firm offers the opportunity to explore the policy implications of the profit and dividend criteria. In particular, we ask and partially answer the following question: how would the optimal policy change if, instead of maximizing value, the firm were to maximize the EPV of future profits?

The answer to this question, besides illuminating implications of the two criteria, is important for its own sake. In actuality, financial market frictions abound and it is not uncommon for shareholder pressures for higher dividends to wax and wane as the economic climate fluctuates. This affects the behavior of management and, consequently, financial and operational decisions. Therefore, characterizing how these decisions depend on the emphasis placed on dividends deepens our understanding of actual business practice.

1.3 Main contributions

This paper formulates and analyzes a stochastic dynamic model of a self-financed firm in a risky environment that decides how much output to produce and how much capital to invest in capacity expansion, to distribute as dividends, and to retain as liquidity. The criterion employed to make these decisions is to maximize the EPV of the firm's profits, or dividends, or a mixture of the two. The paper makes three kinds of contributions.

Contribution 1. The model is a structured Markov decision process (MDP) and results in a dynamic program with three state variables: the current levels of capacity and capital, and an index of exogenous uncertainty. We show that under linear production costs and investment yields, the dynamic program's value function has an affine structure. Analytically, this structure yields a complete characterization of the optimal policy. Computationally, it leads to an MDP algorithm that exorcises the the curse of dimensionality.

Contribution 2. The paper provides valuable insights to the interdependence of the assets, namely capacity and capital, in a dynamic risky environment.

(i) In essence, capacity and capital are real options of the firm; the former allows but does not require production in each period, while the latter allows but does not require investment in each period. When considered dynamically, the two types of real options are closely linked because exercising the production option gives birth to the investment option, i.e., capital; and exercising the investment option gives birth to the production option, i.e., capacity. In the paper, we demonstrate that the manner in which capacity and capital are valued agrees with their option nature and reflects the linkage between them.

(ii) The paper explains why "inactive" assets, namely idle capacity and retained capital, are perceived to have positive worth in practice. We prove that the value of idle capacity derives entirely from the prospect that it will eventually be used for production. This indicates that capacity is worthless, i.e., has zero value, if and only if it will remain idle with probability one. Similarly, the value of liquidity derives entirely from the prospect that it will eventually be distributed or invested.

Contribution 3. We characterize the dependence of optimal decisions on the firm's emphasis on maximizing profit versus dividends. Given the asset levels and exogenous uncertainty, a heightened emphasis on profit causes the dividend amount either to remain the same or fall. Thus, profit maximization induces a more aggressive investment attitude than does maximization of the firm's value, because dividend distribution permanently removes the capital from the firm, whereas both immediate investment and greater liquidity retain it for future profit-generating operations.

The propositions and theorems in the paper are new.

The rest of the paper proceeds as follows. Section 2 reviews relevant literature, §3 formulates and discusses the model, and §4 simplifies it. Section 5 derives the optimal policy and provides economic and real option interpretations of the results. Section 6 delineates the interdependence of asset values, and §7 presents a linear program to solve the model and demonstrates its immunity to the curse of dimensionality. Section 8 derives the policy consequences of different optimization criteria. Section 9 summarizes the paper and discusses possible extensions. Appendices in the Electronic Companion have proofs of most formal results and some other material connected with the paper.

2 Related Literature

There does not seem to be a research literature on the triad of production, capacity investment, and finance, so we shall describe the literatures for the dyads among them. The literature on coordination of financial and investment decisions examines a firm that dynamically decides how much to invest in capacity expansion, how much debt to incur or equity to issue, and how much dividend to pay out. Usually, it considers natural capacity depreciation, but production decisions are absent and investment yields are modeled to be deterministic. For example, Titman and Tsyplakov (2007) and Bolton et al. (2011) assume that the firm always produces at capacity, while Gamba and Triantis (2008) model production as an exogenous stochastic process. An exception is Mauer and Triantis (1994), who allow production to be either "on" or "off" with fixed costs when switching from one status to the other, but they consider only a one-shot capacity investment decision. This paper differs from these works in two ways. First, the model includes production-induced capacity depreciation, which is particularly important in oil and gas field operations (Anderson et al. 2014) and other consumer product industries (Wang et al. 2013). Second, the model has random investment yields, which captures a key element of innovative investments. As demonstrated throughout the remainder of the paper, these two features generate new insights on the linkage between production and investment decisions in a risky environment.

There is a large operations management literature using dynamic models to coordinate production and capacity investment in the absence of financial considerations. Broadly speaking, production and capacity investment include national economic planning (Manne 1967), and workforce size management (Lippman et al. 1967). What is particularly relevant to us are firm-level studies which balance the risks of capacity underage and overage under random demand in a dynamic framework. See Iglehart (1965) and Sobel (1970) for different approaches, and Luss (1982) and Van Mieghem (2003) for reviews. The inclusion of capacity depreciation is conspicuous in recent research, and examples include Chao et al. (2009), which incorporates random depreciation rates, and Wang et al. (2013), which allows usage-induced depreciation. Our paper differs from these works because its motivation includes oil and gas field operations in which there are no out-of-pocket underage and overage costs. Nevertheless, the expense due to capacity overage could easily be incorporated in our model by adding a variable capacity holding cost.

The third dyad addresses the coordination of production and financial decisions in circumstances where there is random product market demand, and unsold items are inventoried from one period to the next. When inventory holding costs are important, this may lead to bankruptcy even if a firm is debt-free (see Li et al. 2013 for an overview of this literature). This paper, however, focuses on cases where the cost of capacity investment dominates the cost of inventory. Therefore, inventory is not in the model.

Some valuable insights in this paper stem from a real option perspective on the valuation and management of firm assets. This viewpoint connects our work with the large literature that studies firm decisions using real option theory (Dixit and Pindyck 1994). Studies that are particularly relevant to this paper have non-lumpy investment in capacity expansion. Pindyck (1988) is an early example devoid of financial considerations. Other works (Boyle and Guthrie 2003, Sundaresan and Wang 2007, Asvanunt et al. 2010) integrate financing and investment decisions in a real option framework, but consider only lumpy investments, with the number and sizes of investments exogenous and known. This paper, in contrast, endogenously determines the sizes of capacity investments under self-financing constraints, thus deepening understanding of the real option nature of the firm's assets and the associated optimal decisions.

Although the model in this paper is posed in a generic setting, its motivations include the exploration and extraction of natural resources such as petroleum and gas. In that context, "capacity" corresponds to maximum outflow rate from proven reserves, "risky capacity investment" to exploration, and "production" to resource extraction. Some studies in this area take the perspective of a proven reserve manager who makes decisions regarding production but not exploration (Pindyck 1981, Amit 1986, Anderson et al. 2014). Among the studies that also consider exploration investments, some investigate the dynamic exploitation and exploration decisions from an aggregate firm level, like ours, but under deterministic product market prices (Pindyck 1978, Arrow and Chang 1982, Deshmukh and Pliska 1980). Some others recognize the importance of exogenous randomness, but exploration and extraction decisions are in the context of a given project (Paddock et al. 1988, Cortazar et al. 2001). Our work contributes to this literature by providing an integrated framework to understand aggregate exploration and extraction decisions of a firm that evolves in a risky environment.

3 Model

Consider a discrete-time multiperiod model of a *self-financed* firm that faces exogenous stochastic product prices and investment yields and lives in perpetuity. In each period, the firm decides how much to produce, how much to invest in capacity expansion, and the amount of dividend to issue. The model of this process is formulated in §3.1 and discussed in §3.2.

3.1 Model formulation

At the beginning of period $t = 1, 2, ...$, the firm observes its past history and its current capacity K_t, its capital W_t, its variable production cost c_t, and the market price for its output $\tilde{\rho}_t$. It chooses the quantity of production q_t, the dollar amount of investment i_t, and the dollar amount of dividends x_t given its existing assets:

$$0 \leq q_t \leq K_t, \qquad 0 \leq x_t, \qquad \text{and} \qquad 0 \leq i_t. \tag{1}$$

For ease of exposition, assume that the firm produces on credit and pays the production cost $c_t q_t$ after the output is sold and the revenue $\tilde{\rho}_t q_t$ is received. Then the capital W_t is allocated between investment, dividend and liquidity buildup:

$$x_t + i_t \leq W_t, \tag{2}$$

where the slack between the right and left sides of (2) is the amount of capital allocated to liquidity. The dynamics of capital satisfy

$$W_{t+1} = W_t - x_t - i_t + \rho_t q_t, \tag{3}$$

in which $\rho_t \equiv \tilde{\rho}_t - c_t$ denotes the market price net of the production cost (hereafter, the "net price"), which may be negative.

Let $\epsilon_t \geq 0$ be the random amount of capacity installed by each unit of investment in period t. Then the firm's capacity level evolves as follows:

$$K_{t+1} = \theta K_t - \lambda q_t + \epsilon_t i_t, \tag{4}$$

where θ and λ satisfy $0 \leq \lambda \leq \theta \leq 1$ and reflect the depreciation of the capacity. The rate θ describes the natural capacity depreciation, and λ the production-induced depreciation. So $\theta = 1$ corresponds to zero natural depreciation, and $\lambda = 0$ to zero production-induced depreciation.

We assume that net prices ρ_t and investment yields ϵ_t are Markov-modulated. That is, there is an exogenous discrete-time Markov process $s_1, s_2, ...$ with state space Ω, and functions $p(\cdot)$ and

$e(\cdot)$ on Ω, such that

$$\rho_t = p(s_t) \quad \text{and} \quad \epsilon_t = e(s_{t+1}). \tag{5}$$

This representation of the sequences of prices and investment yields encompasses a rich array of models ranging from a constant price and yield to higher-order discrete-time Markov processes for correlated prices and yields.

The partial history up to the beginning of period t is $H_t := (K_1, W_1, s_1, q_1, x_1, i_1, ..., K_{t-1}, W_{t-1}, s_{t-1}, q_{t-1}, x_{t-1}, i_{t-1}, K_t, W_t, s_t)$. A *policy* is a decision rule based on the set of all partial histories. That is, policy π specifies (q_t, i_t, x_t), for each t and H_t, satisfying constraints (1) and (2). Let $0 \leq \beta < 1$ denote the single-period discount factor, employ geometric discounting, and assume risk neutrality. The present values of the streams of profits and dividends are then $\sum_{t=1}^{\infty} \beta^{t-1} \rho_t q_t$ and $\sum_{t=1}^{\infty} \beta^{t-1} x_t$, respectively. The mixture of the two is

$$\Pi := \mu \sum_{t=1}^{\infty} \beta^{t-1} \rho_t q_t + (1-\mu) \sum_{t=1}^{\infty} \beta^{t-1} x_t = \sum_{t=1}^{\infty} \beta^{t-1} [\mu \rho_t q_t + (1-\mu) x_t] \tag{6}$$

in which $0 \leq \mu \leq 1$ is a constant. If $\mu = 1$, the mixture is the present value of profits. If $\mu = 0$, it is the present value of dividends, and the expected value of Π is the financial value of the firm.

In order to specify the optimization criterion, define $\Pi_\tau = \sum_{t=\tau}^{\infty} \beta^{t-1} [\mu \rho_t q_t + (1-\mu) x_t]$ $(\tau = 1, 2, \ldots)$. A policy π^* is *optimal* if $\mathbb{E}_{\pi^*|H_t}(\Pi_t) \geq \mathbb{E}_{\pi|H_t}(\Pi_t)$ for all partial histories H_t, all $t = 1, 2, ...$, and all policies π. This paper characterizes an optimal policy and provides highly efficient algorithms to compute it and its value function.

3.2 Model discussion

A few comments about the model are in order. First, $W_t \geq 0$ for all $t = 1, 2, \ldots$ because in each period $W_t \geq x_t + i_t$, $x_t \geq 0$, and $i_t \geq 0$. This nonnegativity of capital W_t, which eliminates bankruptcy risk, and thus avoids any need to treat bankruptcy in the model, is made possible by the firm's immunity to negative exogenous shocks. By construction, the random investment yield ϵ_t is always nonnegative. So the only negative shocks to the firm come from the net price. Since the firm observes the net price ρ_t before selecting the production quantity, it can produce zero should the net price be negative, thus steering away from bankruptcy. This generates different optimal policies than if the firm were subject to both positive and negative shocks and, consequently, bankruptcy risk and cost (see e.g. Radner and Shepp 1996). Similar remarks are valid concerning risk-neutrality.

In that sense, this model provides a useful benchmark to understand how bankruptcy risk and cost, and risk sensitivity affect a firm's decisions.

Another important effect of the constraint $W_t \geq 0$ ($t = 1, 2, \ldots$) is on the set of *admissible* production decisions q_t. Combining nonnegativity with the dynamic equation (3) generates

$$W_t \geq x_t + i_t - \rho_t q_t, \tag{7}$$

which is binding only when the firm faces a negative net price given that $W_t \geq x_t + i_t$.

Nonnegativity of W_t and the dynamic equation for K_t imply that the model is a Markov decision process (MDP) in which the sequence of states is $\{(W_t,\ K_t,\ s_t) : t = 1, 2, \ldots\}$, and the state space is $\mathcal{S} := \Re_+^2 \times \Omega$ (where $\Re_+$ denotes $[0, \infty)$). The criterion in this paper is to maximize the EPV of the mixture of profits and dividends. Henceforth, "MDP" refers to the model with this criterion.

The second comment about the model concerns the important assumption that the production cost is linear in production quantity. The model does not include production smoothing costs such as costs of stopping and resuming production. This is a reasonable approximation given that the model considers production alongside capacity investments. In most industries, investment expenses are orders of magnitude higher than marginal production cost variations and production smoothing costs. For example, the cost of halting, restoring, and varying the oil flow from proven reserves is negligible compared to the cost of exploration, which is normally in billions of dollars (although the balance is in the other direction in a few industries such as ceramics manufacturing).

Third, ϵ_t in (3) denotes the random yield per unit of investment. A linear yield assumption may not be realistic for some single-project investments where the law of diminishing returns is important, but it is a reasonable approximation when i_t represents the firm's aggregate investment over a portfolio of projects.

The most critical assumptions are those on linear investment yield and linear production cost. As will become clear, these assumptions induce important features of the optimal policy and linearity in the value function of the resulting MDP.

4 Simplified Dynamic Program

This section presents a straight-forward dynamic program that corresponds to the MDP, shows that it can be simplified, and relates it to a recursion that is used in §5 to characterize the optimal policy.

Let S_j denote random variable (r.v.) s_{t+1} given $s_t = j$. That is, for each $j \in \Omega$, S_j has sample space Ω, and $s_{t+1}|_{s(t)=j} \sim S_j$. Recall that the MDP has state space $\mathcal{S} := \Re^2_+ \times \Omega$, so its corresponding dynamic program is, for all $(K, W, j) \in \mathcal{S}$,

$$V(K,W,j) = \max_{(x,i,q)} \{J(x,i,q;K,W,j) : 0 \le x,\ 0 \le i,\ i + x \le W,\ i + x - p(j)q \le W,\ 0 \le q \le K\}, \tag{8a}$$

$$J(x,i,q;K,W,j) = \mu p(j)q + (1-\mu)x + \beta\,\mathbb{E}\left[V\big(\theta K - \lambda q + e(S_j)i, W + p(j)q - x - i, S_j\big)\right]. \tag{8b}$$

The decision variables in this dynamic program, x, i, and q, correspond to the respective amounts of the dividend, the capacity-enhancing investment, and production. The state variables are the current values of capacity, K, capital, W, and the state j of the exogenous Markov process. The maximand, which is specified in (8b), consists of two parts: the immediate reward $\mu p(j)q+(1-\mu)x$; and the EPV of the value function at next period's state induced by an optimal policy used thereafter. In particular, next-period's endogenous state variables, namely capacity and capital, are governed by dynamic equations (3) and (4), and next-period's exogenous state variable is S_j. Finally, the constraints follow from (1), (2), and (7).

There are combinations of model parameters and policies that will yield unboundedly large EPVs of profits, dividends, and their mixtures as defined in (6). See §B in the Electronic Companion for a discussion of exploding models and sufficient conditions for the value function of the MDP to be bounded. For the remainder of the paper, however, we preclude the possibility of an unbounded value function with the following assumption.

Assumption 1. (Boundedness) *The model parameters are such that in dynamic program* (8) *the value function is bounded on* $\mathcal{S}$ *and the maximum is achieved.*

A finite-horizon recursion counterpart to (8) is $V_0(\cdot,\cdot,\cdot) \equiv 0$ and for $n = 1, 2, ...,$

$$V_n(K, W, j) = \max_{(x,i,q)} \{J_n(x, i, q; K, W, j) : 0 \le x,\ 0 \le i,\ i + x \le W,$$
$$i + x - p(j)q \le W,\ 0 \le q \le K\}, \tag{9a}$$
$$J_n(x, i, q; K, W, j) = \mu p(j)q + (1-\mu)x$$
$$+ \beta\, \mathbb{E}\left[V_{n-1}\big(\theta K - \lambda q + e(S_j)i, W + p(j)q - x - i, S_j\big)\right]. \tag{9b}$$

The following lemma relates the value functions of (9) to $V(\cdot,\cdot,\cdot)$ in (8). It asserts that, from any initial state, the finite-horizon value function is non-decreasing in the length of the horizon, and it converges to the value function of the infinite-horizon dynamic program. The proofs of most of the paper's formal results, including this one, are relegated to the Electronic Companion.

Lemma 1. *For all $(K, W, j) \in \mathcal{S}$ and $n = 1, 2, \ldots,$ $V_n(K, W, j) \le V_{n+1}(K, W, j)$. The limit $\bar{V}(K, W, j) = \lim_{n\to\infty} V_n(K, W, j)$ exists and solves (8). If Ω is compact, then there is a unique (bounded) solution of (8) and thus $\bar{V}(\cdot,\cdot,\cdot) = V(\cdot,\cdot,\cdot)$.*

Lemma 1 justifies approximating dynamic program (8) arbitrarily closely with the finite-horizon recursion (9), which can be simplified using the lemma below:

Lemma 2. *For $n = 1, 2, \ldots,$ the value function $V_n(\cdot,\cdot,j)$ in (9) satisfies the following recursion with $V_0(\cdot,\cdot,\cdot) \equiv 0$ and for $n = 1, 2, ...,$*

$$V_n(K, W, j) = \max_{(x,i,q)} \{J_n(x, i, q; K, W, j) : 0 \le x,\ 0 \le i,\ i + x \le W,\ 0 \le q \le K\}, \tag{10a}$$
$$J_n(x, i, q; K, W, j) = \mu p(j)q + (1-\mu)x$$
$$+ \beta\, \mathbb{E}\left[V_{n-1}\big(\theta K - \lambda q + e(S_j)i, W + p(j)q - x - i, S_j\big)\right]. \tag{10b}$$

It then follows from Lemma 1 that the value function $V(\cdot,\cdot,\cdot)$ in dynamic program (8) also is the value function of the following simpler functional equation:

$$V(K, W, j) = \max_{(x,i,q)} \{J(x, i, q; K, W, j) : 0 \le x,\ 0 \le i,\ i + x \le W,\ 0 \le q \le K\}, \tag{11a}$$
$$J(x, i, q; K, W, j) = \mu p(j)q + (1-\mu)x$$
$$+ \beta\, \mathbb{E}\left[V\big(\theta K - \lambda q + e(S_j)i, W + p(j)q - x - i, S_j\big)\right]. \tag{11b}$$

The remainder of the paper uses dynamic programs (10) and (11) to characterize the value function $V(\cdot, \cdot, \cdot)$ and the associated optimal policy.

5 Optimal Policy, Optimal Value of the Objective, and a Real Option Perspective

This section analyzes the infinite-horizon dynamic program (11). Section 5.1 characterizes its value function and optimal policy, and §5.2 discusses economic insights provided by the characterizations, the real option nature of capital and capacity assets, and the linkage between the assets' values.

5.1 Characterization of an optimal policy

The following analysis of the finite-horizon dynamic program (10) proves that the value function $V_n(K, W, j)$ is linear in the *endogenous* state variables K and W, and that this key feature is inherited by its infinite-horizon counterpart (11).

Lemma 3. *1. For each $n = 1, 2, ...,$ there are real-valued functions $f_n(\cdot)$ and $g_n(\cdot)$ on Ω such that*

$$V_n(K, W, j) = f_n(j)K + g_n(j)W, \qquad (K, W, j) \in \mathcal{S}. \tag{12}$$

2. Define the functions $\mathcal{F}_n(j) = \mathbb{E}[f_n(S_j)]$, $\mathcal{F}_n^e(j) = \mathbb{E}[f_n(S_j)e(S_j)]$, and $\mathcal{G}_n(j) = \mathbb{E}[g_n(S_j)]$ ($j \in \Omega$). The functions $f_n(\cdot)$ and $g_n(\cdot)$ satisfy $f_0(\cdot) = g_0(\cdot) \equiv 0$ and for $n = 1, 2, ...,$

$$f_n(j) = max\left\{\beta\theta\mathcal{F}_{n-1}(j),\ \beta(\theta - \lambda)\mathcal{F}_{n-1}(j) + [\mu + \beta\mathcal{G}_{n-1}(j)]p(j)\right\} \tag{13a}$$

$$g_n(j) = max\{1 - \mu, \beta\mathcal{F}_{n-1}^e(j), \beta\mathcal{G}_{n-1}(j)\} \tag{13b}$$

Proof. The inductive proof of (12) is initiated by $V_0(\cdot, \cdot, \cdot) \equiv 0$ which satisfies (12) with $n = 0$ and

$f_0(\cdot) = g_0(\cdot) \equiv 0$. In (10), making the inductive assumption that (12) is valid at n,

$$
\begin{aligned}
V_{n+1}(K,W,j) = \max_{(x,i,q)} \{ & \beta\,\mathbb{E}\left(V_n[\theta K - \lambda q + e(S_j)i, W - x - i + p(j)q, S_j]\right) + \mu p(j)q + (1-\mu)x : \\
& 0 \le x,\ 0 \le i,\ i + x \le W,\ 0 \le q \le K\} \\
= \max_{(q,x,i)} & \Big\{ \mu p(j)q + (1-\mu)x + \beta\,\mathbb{E}\Big[f_n(S_j)[\theta K - \lambda q + e(S_j)i] \\
& + g_n(S_j)(W - x - i + p(j)q)\Big] :\ 0 \le x,\ 0 \le i,\ i + x \le W,\ 0 \le q \le K \Big\} \\
= \beta\theta & \mathcal{F}_n(j)K + \max_q \{[\mu p(j) + \beta p(j)\mathcal{G}_n(j) - \lambda\beta\mathcal{F}_n(j)]q :\ 0 \le q \le K\} \qquad (14) \\
& + \max_{(x,i)} \{[1 - \mu - \beta\mathcal{G}_n(j)]x + \beta[\mathcal{F}^e_n(j) - \mathcal{G}_n(j)]i : 0 \le x, 0 \le i, i + x \le W\} \qquad (15) \\
& + \beta\mathcal{G}_n(j)W \qquad (16)
\end{aligned}
$$

The second term of (14) is a linear program, so it has an optimum at one of its extreme points which are $q = 0$ or $q = K$. Therefore, rewrite (14) as

$$
\max\Big\{\beta\theta\mathcal{F}_n(j),\ \beta\theta\mathcal{F}_n(j) + \mu p(j) + \beta p(j)\mathcal{G}_n(j) - \beta\lambda\mathcal{F}_n(j)\Big\}K
$$

which is linear in K and establishes (13a).

Optimization (15) is a linear program with a nonempty and compact feasibility set, so there is an optimum at one of the extreme points which comprise the set $\{(x,i) : (W,0), (0,W), (0,0)\}$. This permits the sum of (15) and (16) to be written as

$$
\begin{aligned}
& \max\{[1 - \mu - \beta\mathcal{G}_n(j)]W,\ \beta[\mathcal{F}^e_n(j) - \mathcal{G}_n(j)]W,\ 0\} + \beta\mathcal{G}_n(j)W \\
= \ & \max\{[1 - \mu - \beta\mathcal{G}_n(j)],\ \beta[\mathcal{F}^e_n(j) - \mathcal{G}_n(j)],\ 0\}W + \beta\mathcal{G}_n(j)W \\
= \ & \max\{1 - \mu,\ \beta\mathcal{F}^e_n(j),\ \beta\mathcal{G}_n(j)\}W \qquad (17) \\
= \ & g_{n+1}(j)W \qquad (18)
\end{aligned}
$$

with

$$
g_{n+1}(j) = \max\{1 - \mu, \beta\mathcal{F}^e_n(j), \beta\mathcal{G}_n(j)\}.
$$

□

The infinite-horizon dynamic program (11) inherits the principal property of the finite-horizon recursion (10), which is that its value function $V(K, W, j)$ has the same linear dependence on the *endogenous* state variables, K and W, as $V_n(K, W, j)$. As stated in the theorem below, this linearity leads to an "all-or-nothing" structure of the optimal policies. That is, the optimal production level is either 0 or K, and the optimal joint dividend and investment decision, (x, i), is $(W, 0)$, $(0, W)$, or $(0, 0)$. In the remainder of the paper, we refer to this as an *extremal* policy.

Let $1(\cdot)$ denote the indicator function which equals 1 if its argument is true, and 0 if it is false.

Theorem 1. *1. There are real-valued functions $f(\cdot)$ and $g(\cdot)$ on Ω such that*

$$V(K, W, j) = f(j)K + g(j)W, \qquad (W, K, j) \in \mathcal{S}. \tag{19}$$

2. Define $\mathcal{F}(j) = \mathbb{E}[f(S_j)]$, $\mathcal{F}^e(j) = \mathbb{E}[f(S_j)e(S_j)]$, and $\mathcal{G}(j) = \mathbb{E}[g(S_j)]$ $(j \in \Omega)$. The functions $f(\cdot)$ and $g(\cdot)$ satisfy

$$f(j) = \max\left\{\beta\theta\mathcal{F}(j),\ \beta(\theta - \lambda)\mathcal{F}(j) + [\mu + \beta\mathcal{G}(j)]p(j)\right\}, \tag{20a}$$

$$g(j) = \max\{1 - \mu,\ \beta\mathcal{F}^e(j),\ \beta\mathcal{G}(j)\}. \tag{20b}$$

3. Define the following subsets of Ω:

$$\Omega_+ = \{j \in \Omega : [\mu + \beta\mathcal{G}(j)]p(j) > \beta\lambda\mathcal{F}(j)\}, \tag{21a}$$

$$\Omega_- = \{j \in \Omega : [\mu + \beta\mathcal{G}(j)]p(j) \le \beta\lambda\mathcal{F}(j)\}, \tag{21b}$$

and

$$\Omega_x = \{j \in \Omega : \beta\mathcal{F}^e(j) \le 1 - \mu, \beta\mathcal{G}(j) \le 1 - \mu\}, \tag{22a}$$

$$\Omega_I = \{j \in \Omega : \beta\mathcal{F}^e(j) > 1 - \mu, \mathcal{F}^e(j) \ge \mathcal{G}(j)\}, \tag{22b}$$

$$\Omega_0 = \{j \in \Omega : \mathcal{F}^e(j) < \mathcal{G}(j), \beta\mathcal{G}(j) > 1 - \mu\}. \tag{22c}$$

Then $\{\Omega_0, \Omega_x, \Omega_I\}$ and $\{\Omega_+, \Omega_-\}$ partition Ω:

$$1(j \in \Omega_x) + 1(j \in \Omega_I) + 1(j \in \Omega_0) = 1(j \in \Omega_+) + 1(j \in \Omega_-) = 1. \tag{23}$$

There is an optimal stationary policy $\big(Q(K,W,j),\ X(K,W,j), I(K,W,j)\big)$ *that satisfies*

$$Q(K,W,j) = 1\,(j \in \Omega_+)\,K, \tag{24}$$

$$X(K,W,j) = 1\,(j \in \Omega_x)\,W, \text{ and} \tag{25}$$

$$I(K,W,j) = 1\,(j \in \Omega_I)\,W. \tag{26}$$

5.2 Insights from Theorem 1

Value function. Value function (19) in Theorem 1, $V(K,W,j) = f(j)K + g(j)W$, has four important features: (i) it is the sum of two terms, one of which pertains to the firm's capacity asset, and the other to its capital asset; (ii) it is modular with respect to the capacity and capital assets; (iii) both of its terms exhibit *linear* dependence on the associated asset; and (iv) both linear coefficients depend only on the current state of the exogenous stochastic process $\{s_t\}$.

Feature (i) reflects the fundamental fact that a firm can be viewed simply as a collection of assets. In essence, the value function gives the financial worth of the firm evaluated under a particular optimization criterion. Thus, this feature suggests that, regardless of the criterion specifics—whether it is to maximize the value of the firm ($\mu = 0$) or the EPV of its profits ($\mu = 1$), or a mix of the two ($0 < \mu < 1$)—*the worth of the firm is always the aggregate worth of its assets.*

Feature (ii) corresponds to an important economic property. Capacity and capital are neither economic complements nor economic substitutes.

Feature (iii) reveals how the worth of both assets is determined. Specifically, the terms $f(j)K$ and $g(j)W$ embody a surprisingly simple relationship: each unit of capacity is worth the same, and each unit of capital is worth the same. There are neither increasing nor decreasing returns to scale with either type of asset. Henceforth, we refer to $f(j)$ and $g(j)$ as *unit values.* The linearity in W and K stems from the linearity of the model, which is rooted in the assumption of linear production costs and investment yields.

Feature (iv) asserts that the unit values of capacity and capital are determined solely by the current state of exogenous uncertainty j, and are thus determined exogenously, rather than endogenously by the firm. This reflects the fundamental assumption that the firm is a price taker and captures the susceptibility of real-world business to the wider economy. In particular, the dependence of these unit values on j is spelled out by the functional equations (20) for $f(\cdot)$ and $g(\cdot)$, which are discussed below.

Unit value of capacity and the optimal production decision. Recall that β is the single-period discount factor, θ and λ are the natural and production-induced depreciation rates of capacity, and $\mathcal{F}(j)$ and $\mathcal{G}(j)$ are the expected unit values of capacity and capital next period, given that the current exogenous state is j. Equation (20a) for the unit value of capacity, $f(j)$, can be parsed as follows:

$$f(j) = \max\Big\{ \underbrace{\beta\theta\mathcal{F}(j)}_{\substack{\text{EPV of gain due}\\ \text{to not producing}}} , \overbrace{\beta\theta\mathcal{F}(j) + \underbrace{\mu p(j)}_{\substack{\text{Immediate reward}\\ \text{from production}}} + \underbrace{\beta\mathcal{G}(j)p(j)}_{\substack{\text{EPV of } \textit{capital}\\ \text{generated by production}}} - \underbrace{\beta\lambda\mathcal{F}(j)}_{\substack{\text{EPV of loss due to}\\ \text{capacity depreciation}\\ \text{under production}}}}^{\text{EPV of gain resulting from production}} \Big\}.$$

Therefore, Theorem 1 essentially prescribes that given j, if the EPV of the gain resulting from using one unit of capacity for production, $\beta(\theta - \lambda)\mathcal{F}(j) + \mu p(j) + \beta p(j)\mathcal{G}(j)$, is greater than that from letting this unit be idle, $\beta\theta\mathcal{F}(j)$, then the optimal policy is to use it and produce one unit of output. Further, the value of this unit of capacity equals the EPV of the gain generated by production. Otherwise, the optimal policy is to let the unit of capacity be idle, and the value of capacity equals the EPV of the gain generated from idle capacity.

This understanding is consistent with a real option view of capacity. Since in any time period the firm can produce any amount of output up to its capacity, each unit of capacity represents an option on production. In principle, the firm can exercise this option by producing up to one unit of output. The constant unit value of capacity, however, suggests that the options embedded in various capacity units have the same worth. Consequently, the optimal action regarding any particular unit of capacity must be the same for all units. It then follows from the complete fungibility of capacity that whatever action is optimal for each unit must be either to produce to the maximum amount, or not to produce at all.

Confining attention to this class of extremal policies, one unit of capacity boils down to an option with two actions; one is the immediate production of one unit of output, and the other is to let the capacity be idle for the current period. If the former is implemented, i.e., the option is exercised, the EPV of the firm's payoff is $\beta(\theta - \lambda)\mathcal{F}(j) + [\mu + \beta\mathcal{G}(j)]p(j)$. Otherwise, the EPV of the payoff is $\beta\theta\mathcal{F}(j)$. From this perspective, Theorem 1 states that the optimal policy takes the action that generates a higher payoff, and that the value of the option, i.e., of one unit of capacity, equals this higher payoff.

Unit value of capital and the optimal tradeoff between investment and liquidity. The equation for the unit value of capital, $g(\cdot)$, can be parsed in a similar way:

$$g(j) = \max \{ \underbrace{1-\mu}_{\substack{\text{Immediate reward} \\ \text{if issued as dividend}}}, \underbrace{\beta\mathcal{F}^e(j)}_{\substack{\text{EPV of payoff} \\ \text{if invested}}}, \underbrace{\beta\mathcal{G}(j)}_{\substack{\text{EPV of payoff if} \\ \text{held as liquidity}}} \}.$$

Each term in the maximand corresponds to one action the firm can take with one unit of capital in hand. Action 1 is immediate dividend issuance, which generates an immediate reward of $1-\mu$. Action 2 is immediate investment, which installs a random number $e(S_j)$ units of capacity that will be ready for production next period, thus generating a payoff whose EPV is $\beta\,\mathbb{E}[e(S_j)f(S_j)] = \beta\mathcal{F}^e(j)$. Action 3 is to hold this unit of capital for the current period, which increases liquidity and generates a payoff with EPV $\beta\mathcal{G}(j)$.

Theorem 1's statement of the optimal investment, dividend and liquidity decisions, and the unit value of capital can be summarized as the following: the action that yields the highest payoff should be designated optimal, and the unit value of capital equals this highest payoff. This conclusion again agrees with intuition from standard option theory, with each unit of capital being viewed as an option having three possible actions: distribution as dividend, investment in additional capacity, and retention to enhance liquidity.

Linkage between production and investment decisions. A distinguishing feature of Theorem 1 is the interdependence of the unit values of capacity and capital. The interdependence reflects the production-investment linkage that is specific to our dynamic setting.

When considered dynamically, production does not simply generate earnings, but rather leads to *options* on earnings and, more to the point, *options* on future investment. These options can be exercised to generate capacity, which are options on production and can be exercised to yield more options on investment, which in turn gives birth to even more options on production. This transformative cycle between production and investment options goes on forever in a dynamic setting. Therefore, in an important way options on investment and production are embedded in each other. This gives rise to the interdependence of the unit values of capacity and capital, which is further exploited in §6.

Computational improvement. A computational improvement for the MDP, detailed in §7.2, is implicit in Theorem 1's statement that the solution of the pair of functional equations (20) is equivalent to a solution of dynamic program (11).

6 Unit Value of Assets and Optimal Policies

This section studies the pair of functional equations (20) in depth and derives the interdependence of the asset unit values, $f(\cdot)$ and $g(\cdot)$, and their implications for the optimal policy. In addition, it lays a foundation to examine in §8 the policy implications of the firm's maximization of profit versus value.

In the remainder of the paper, Ω is an arbitrarily ordered discrete set, i.e., $s_1, s_2, ...$ is a Markov chain. **Boldface** denotes vectors and matrices, and all vectors are by default column vectors unless they have the superscipt T.

Let $\mathbf{M} = (m_{jk})$ denote the generic one-step transition probability matrix where $m_{jk} = Prob(s_{t+1} = k|s_t = j)$ $(j,\ k \in \Omega)$. Instead of $f(j)$, $g(j)$, $e(j)$, $p(j)$, $\mathcal{F}(j)$, $\mathcal{G}(j)$, and $\mathcal{F}^e(j)$, we write f_j, g_j, e_j, p_j, $\mathcal{F}_j$, $\mathcal{G}_j$, and $\mathcal{F}_j^e$ in the rest of the paper. With this notation,

$$\mathbb{E}[f(S_j)] = \mathcal{F}_j = \sum_{k\in\Omega} m_{jk} f_k, \tag{27a}$$

$$\mathbb{E}[g(S_j)] = \mathcal{G}_j = \sum_{k\in\Omega} m_{jk} g_k, \tag{27b}$$

$$\mathbb{E}[f(S_j)e(S_j)] = \mathcal{F}_j^e = \sum_{k\in\Omega} m_{jk} e_k f_k. \tag{27c}$$

Define the operator diag($\mathbf{z}$), which takes a vector $\mathbf{z}$ and generates a diagonal matrix $\mathbf{Z} = \mathbf{I}\mathbf{z}$.

6.1 Unit value of capacity

Theorem 1 states that capacity should either be idle or used in full (henceforth, "active"). Analysis below compares the value of idle capacity with that of active capacity, revealing that the former derives entirely from the latter. In particular, capacity is worthless, i.e., has zero value, if and only if with probability one it will remain idle forever; and capacity achieves its highest worth when production occurs.

Recall that the capacity is idle when $j \in \Omega_-$ and active when $j \in \Omega_+$. Correspondingly, create subvectors $\mathbf{f}^+ = (f_j,\ j \in \Omega_+)$ and $\mathbf{f}^- = (f_j,\ j \in \Omega_-)$ that contain unit values of active and idle capacity, respectively. Further, create subvectors $\boldsymbol{\mathcal{F}}^+$, $\boldsymbol{\mathcal{F}}^-$, $\boldsymbol{\mathcal{G}}^+$ and $\mathbf{p}^+$, and designate submatrices of $\mathbf{M}$ corresponding to the ordering in $\mathbf{f}^+$ and $\mathbf{f}^-$ as follows: $\mathbf{M}_u^v$ contains m_{jk} such that $j \in \Omega_u$, and $k \in \Omega_v$, where $u, v \in \{+, -\}$. Then (27) implies

$$\boldsymbol{\mathcal{F}}^- = \mathbf{M}_-^-\mathbf{f}^- + \mathbf{M}_-^+\mathbf{f}^+ \quad \text{and} \quad \boldsymbol{\mathcal{F}}^+ = \mathbf{M}_+^-\mathbf{f}^- + \mathbf{M}_+^+\mathbf{f}^+. \tag{28}$$

Henceforth, $\mathbf{I}$ and $\mathbf{1}$ denote an identity matrix and a column vector of all 1's, respectively, whose dimensions are apparent from the context. For example, $\mathbf{1}_+$ and $\mathbf{I}_-$ have the same dimensions as the vector $\mathbf{f}^+$ and the matrix $\mathbf{M}_-^-$, respectively.

Lemma 4. *The unit value of idle capacity,* $\mathbf{f}^-$*, and the unit value of active capacity,* $\mathbf{f}^+$*, satisfy*

$$\mathbf{f}^- = \beta\theta\mathbf{A}\mathbf{f}^+, \tag{29a}$$

$$\mathbf{f}^+ = \beta(\theta - \lambda)\mathbf{B}\mathbf{f}^+ + \mathit{diag}(\mathbf{p}^+)\left(\mu\mathbf{1}_+ + \beta\boldsymbol{\mathcal{G}}^+\right), \tag{29b}$$

where

$$\mathbf{A} \equiv \sum_{r=0}^{\infty}(\beta\theta\mathbf{M}_-^-)^r\mathbf{M}_-^+ \quad \textit{and} \quad \mathbf{B} \equiv \beta\theta\mathbf{M}_+^-\mathbf{A} + \mathbf{M}_+^+. \tag{30}$$

Lemma 4 specifies the linear dependence of $\mathbf{f}^-$ on $\mathbf{f}^+$. In particular, (29a) asserts that the unit value of idle capacity is derived from the prospect that it will eventually be used for production, i.e., the probability is positive that the process $\{s_t\}$ whose current state is $j \in \Omega_-$ will eventually transit to Ω_+. Each period this capacity unit is idle, it depreciates at the rate θ and the delayed profit realization has to be discounted at the rate β. So $\mathbf{f}^-$ depends on $\mathbf{f}^+$ via the matrix $\mathbf{A}$ in (30), which is the discounted sum of the transition probabilities from the set in which production is nil (Ω_-) to the set in which production is maximal (Ω_+).

The unit value of active capacity $\mathbf{f}^+$, on the other hand, derives from two sources. The first source, reflected in the term $\beta(\theta - \lambda)\mathbf{B}\mathbf{f}^+$ in (29b), is the prospect that the capacity will be active again after being idle for some number of periods, analogous to the mechanism described in (29a). The second source, reflected in $\text{diag}(\mathbf{p}^+)\left(\mu\mathbf{1}_+ + \beta\boldsymbol{\mathcal{G}}^+\right)$, is the immediate reward and the value of capital generated from production, as discussed in §5.2.

Lemma 4 leads to several insights about the value of capacity and the optimal production decisions. First, capacity achieves its highest value when it is active, and achieves its lowest value, 0, when and only when it will remain idle forever.

Corollary 1. *1. For all $j \in \Omega$, $f_j \geq 0$.*

2. $f_j = 0$ if and only if $j \in \Omega_-$ and Ω_+ is inaccessible from j.

3. If $\Omega_+ \neq \emptyset$, then $\arg\max \ f_j \in \Omega_+$.

Second, the net price is positive whenever capacity is active.

Corollary 2. $\mathbf{p}^+ > 0$.

Third, the expected unit value of capacity next period, $\boldsymbol{\mathcal{F}}$, depends linearly on $\mathbf{f}^+$.

Corollary 3. *The expected unit value of capacity next period which is idle now, $\boldsymbol{\mathcal{F}}^-$, and the expected unit value of capacity next period which is active now, $\boldsymbol{\mathcal{F}}^+$, satisfy*

$$\boldsymbol{\mathcal{F}}^- = \mathbf{A}\mathbf{f}^+ \quad \textit{and} \quad \boldsymbol{\mathcal{F}}^+ = \mathbf{B}\mathbf{f}^+. \tag{31}$$

6.2 Unit value of capital

Theorem 1 asserts that an optimal policy causes capital to be either fully invested, or fully distributed, or fully retained. This subsection examines the unit values of capital under this policy and shows that capital achieves its highest value when invested and its lowest value when distributed. In particular, the value of retained capital derives entirely from the prospect that it will eventually be distributed or invested. This implies that capital is retained only when there is positive probability of eventually investing it.

Employing the same procedure as in §6.1, we use the partition $\{\Omega_x, \Omega_I, \Omega_0\}$ to create unit value subvectors $\mathbf{g}^x = (g_j, \ j \in \Omega_x)$, $\mathbf{g}^I = (g_j, \ j \in \Omega_I)$, $\mathbf{g}^0 = (g_j, \ j \in \Omega_0)$, and expected unit value subvectors $\boldsymbol{\mathcal{G}}^x$, $\boldsymbol{\mathcal{G}}^I$, $\boldsymbol{\mathcal{G}}^0$ in the next period. Also, designate submatrices of $\mathbf{M}$ as follows: $\mathbf{M}_u^v$ contains m_{jk} such that $j \in \Omega_u$, and $k \in \Omega_v$, where $u, v \in \{x, I, 0\}$. Then (27) implies

$$\boldsymbol{\mathcal{G}}^x = \mathbf{M}_x^x\mathbf{g}^x + \mathbf{M}_x^I\mathbf{g}^I + \mathbf{M}_x^0\mathbf{g}^0, \tag{32a}$$

$$\boldsymbol{\mathcal{G}}^I = \mathbf{M}_I^x\mathbf{g}^x + \mathbf{M}_I^I\mathbf{g}^I + \mathbf{M}_I^0\mathbf{g}^0, \tag{32b}$$

$$\boldsymbol{\mathcal{G}}^0 = \mathbf{M}_0^x\mathbf{g}^x + \mathbf{M}_0^I\mathbf{g}^I + \mathbf{M}_0^0\mathbf{g}^0. \tag{32c}$$

Create the following yield-adjusted one-step transition probability matrix between states in Ω_I and any state in Ω: $\mathbf{M}_I^e = (m_{jk}e_k, \ j \in \Omega_I, k \in \Omega)$, and create vector $\boldsymbol{\mathcal{F}}^{eI} = (\mathcal{F}_j^e, \ j \in \Omega_I)$ from states in Ω_I. Then by (27c), $\boldsymbol{\mathcal{F}}^{eI} = \mathbf{M}_I^e\mathbf{f}$.

Lemma 5. *The unit values of capital, according to whether the optimal decision is dividend issuance* $\mathbf{g}^x$*, capacity investment* $\mathbf{g}^I$*, or increased liquidity* $\mathbf{g}^0$*, are as follows:*

$$\mathbf{g}^x = (1-\mu)\mathbf{1}_x, \tag{33a}$$

$$\mathbf{g}^I = \beta \mathbf{M}_I^e \mathbf{f}, \tag{33b}$$

$$\mathbf{g}^0 = \beta \mathbf{C} \mathbf{1}_x (1-\mu) + \beta^2 \mathbf{D} \mathbf{M}_I^e \mathbf{f}, \tag{33c}$$

where

$$\mathbf{C} = \sum_{r=0}^{\infty} \left(\beta \mathbf{M}_0^0\right)^r \mathbf{M}_0^x \quad \textit{and} \quad \mathbf{D} = \sum_{r=0}^{\infty} \left(\beta \mathbf{M}_0^0\right)^r \mathbf{M}_0^I. \tag{34}$$

Lemma 5 reveals that the unit values of capital are affine functions of $\mathbf{f}^+$ with nonnegative coefficients. The unit value associated with immediate dividend issuance, $\mathbf{g}^x$ in (33a), is uniformly the weight attached to dividends in the optimization criterion, namely $1-\mu$. That is, capital diverted to dividend has no future value for the firm; it generates only an immediate reward whose unit value is $1-\mu$. In contrast, the unit value of capital that is invested or retained, $\mathbf{g}^I$ or $\mathbf{g}^0$ in (33b) or (33c), respectively, will eventually lead to production and, therefore, has a term that is proportional to $\mathbf{f}$. This reflects the interdependence between capacity and capital values discussed in §5.2.

In Lemma 5, equation (33c) regarding $\mathbf{g}^0$ reveals that liquidity derives its value entirely from the prospect of eventually being distributed as dividends or invested to enhance capacity. To see this, first observe that matrix $\mathbf{C}$ in (33c) is the discounted likelihood that the firm eventually issues all capital as dividends after retaining it for some number of periods. Therefore, the first term in (33c), $\beta \mathbf{C} \mathbf{1}_x (1-\mu)$, is the EPV of the payoff from dividend issuance. Similarly, matrix $\mathbf{D}$ is the discounted likelihood that the firm eventually invests all capital after retaining it for some number of periods, and $\beta^2 \mathbf{D} \mathbf{M}_I^e \mathbf{f}$ is the EPV of the payoff from investment. The coefficient is β^2 instead of β because there is a one period lag between capacity installation and profit generation. This leads to the following insight regarding liquidity.

Corollary 4. *If* $j \notin \Omega_I$*, then* $j \in \Omega_0$ *only if* Ω_I *is accessible from* j*.*

Corollary 4 asserts that increasing liquidity is optimal only when there is positive probability that the firm will invest in the future. The optimality of increasing liquidity is interesting because in our model it stems from the irreversibility of investment rather than an economy of scale in investment. It is well-known in the real option literature that irreversibility makes a delayed commitment a valuable tactic in a risky environment (Dixit and Pindyck 1994). Similarly, we view building liquidity to be a delaying tactic as opposed to "immediate investment". It follows that delayal is sensible only when there is a prospect for future investment, as stated in Corollary 4.

Another important implication of Lemma 5 is that capital achieves its highest value when invested and its lowest value when distributed. Formally,

Corollary 5. *1. For all $j \in \Omega$, $g_j \geq 1-\mu$,* $\arg\min\ g_j \in \Omega_x$.

2. If $\Omega_I \neq \emptyset$, then $\arg\max\ g_j \in \Omega_I$.

The combination of Lemma 5 and (32) yields

Corollary 6. *The respective expected unit values of capital next period when the current decision is dividend issuance, $\boldsymbol{\mathcal{G}}^x$, capacity investment $\boldsymbol{\mathcal{G}}^I$, and liquidity buildup $\boldsymbol{\mathcal{G}}^0$ satisfy*

$$\boldsymbol{\mathcal{G}}^x = \big(\mathbf{M}_x^x + \beta\mathbf{M}_x^0\mathbf{C}\big)\mathbf{1}_x(1-\mu) + \big(\mathbf{M}_x^I + \beta\mathbf{M}_x^0\mathbf{D}\big)\beta\mathbf{M}_I^e\mathbf{f}, \tag{35a}$$

$$\boldsymbol{\mathcal{G}}^I = \big(\mathbf{M}_I^x + \beta\mathbf{M}_I^0\mathbf{C}\big)\mathbf{1}_x(1-\mu) + \big(\mathbf{M}_I^I + \beta\mathbf{M}_I^0\mathbf{D}\big)\beta\mathbf{M}_I^e\mathbf{f}, \tag{35b}$$

$$\boldsymbol{\mathcal{G}}^0 = \mathbf{C}\ \ \mathbf{1}_x(1-\mu) + \mathbf{D}\ \ \beta\mathbf{M}_I^e\mathbf{f}. \tag{35c}$$

Corollary 6 asserts that, regardless of the current decision, the expected unit value of capital is a linear combination of the values realized under dividend issuance and under investment. In particular, the multiplier of the capital value under dividend issuance (investment) is the discounted likelihood that the firm will eventually issue a dividend (invest in capacity) after increasing liquidity for some number of periods.

7 Linear Program and Computational Improvements

This section continues to use the notation introduced in §6. Section 7.1 shows that the asset unit value vectors, $\mathbf{f}$ and $\mathbf{g}$, and the associated optimal policy can be computed with a linear program, whose major computational advantages are discussed in §7.2.

7.1 Linear Program

A necessary condition for $\{f_j, g_j : j \in \Omega\}$ to be a solution of the pair of functional equations (20) is that it satisfies the following linear inequalities:

$$f_j \geq \beta\theta\mathcal{F}_j, \qquad \forall j \in \Omega \tag{36a}$$

$$f_j \geq \beta(\theta-\lambda)\mathcal{F}_j + \mu p_j + \beta p_j \mathcal{G}_j, \qquad \forall j \in \Omega \tag{36b}$$

$$g_j \geq 1-\mu, \qquad \forall j \in \Omega \tag{36c}$$

$$g_j \geq \beta\mathcal{F}_j^e, \qquad \forall j \in \Omega \tag{36d}$$

$$g_j \geq \beta\mathcal{G}_j, \qquad \forall j \in \Omega. \tag{36e}$$

However, one can specify $\{f_j, g_j : j \in \Omega\}$ which satisfies (36) but not (20). That is, (36) is necessary but not sufficient for (20) because there may be too large a gap between the left and right sides of the constraints. For example, if for some $j \in \Omega$ both (36a) and (36b) are strict inequalities, then

$$f_j > \max\{\beta\theta\mathcal{F}_j, \beta(\theta-\lambda)\mathcal{F}_j + [\mu + \beta\mathcal{G}_j]p_j\}$$

which violates (20a).

An intuitive and effective way to eliminate this unwanted possibility is to force downward the left side of each constraint, i.e., to minimize a linear combination of $\{f_j, g_j : j \in \Omega\}$ with *arbitrary positive* coefficients $\{a_j, b_j : j \in \Omega\}$ subject to (36). This intuition leads to the following linear program in which the decision variables are $\{f_j, g_j : j \in \Omega\}$, and the constraints are obtained by defining $\delta_{jk} = 1$ if $j = k$, $\delta_{jk} = 0$ if $j \neq k$, substituting (27) from §6 in (36), and rearranging terms:

$$\min_{\{f_j,g_j:j\in\Omega\}} \sum_{j\in\Omega}(a_j f_j + b_j g_j) \tag{37a}$$

subject to, for all $j \in \Omega$,

$$\sum_{k\in\Omega}(\delta_{jk} - \beta\theta m_{jk})f_k \geq 0, \tag{37b}$$

$$\sum_{k\in\Omega}[\delta_{jk} - \beta(\theta-\lambda)m_{jk}]f_k - \beta p_j \sum_{k\in\Omega} m_{jk}g_k \geq \mu p_j, \tag{37c}$$

$$g_j \geq 1-\mu, \tag{37d}$$

$$g_j - \beta\sum_{k\in\Omega} m_{jk}e_k f_k \geq 0, \tag{37e}$$

$$\sum_{k\in\Omega}(\delta_{jk} - \beta m_{jk})g_k \geq 0. \tag{37f}$$

The following statement establishes the equivalence between the optimal solution of the linear program (37) and the solution of the pair of functional equations (20).

Proposition 1. $\{f_j, g_j : j \in \Omega\}$ *is an optimal solution of linear program (37) if and only if it is a solution of (20). This solution is unique, hence it is invariant with respect to the particular positive values of* $\{a_j, b_j\}$.

The linear program (37) not only solves the functional equations (20), but also reveals the optimal stationary policy of the dynamic program. To see this, first note that by Theorem 1, the optimal stationary policy is fully specified by sets Ω_-, Ω_+, Ω_x, Ω_I, and Ω_0 as follows: $Q(K,W,j) = 0$ if and only if (iff) $j \in \Omega_-$ and $Q(K,W,j) = K$ iff $j \in \Omega_+$; $\big(X(K,W,j), I(K,W,j)\big) = (W,0)$ iff $j \in \Omega_x$, $\big(X(K,W,j), I(K,W,j)\big) = (0,W)$ iff $j \in \Omega_I$, and $\big(X(K,W,j), I(K,W,j)\big) = (0,0)$ iff $j \in \Omega_0$. Further, the membership of any particular state j in set Ω_- or Ω_+, and in Ω_x, Ω_I, or Ω_0, is determined by the manner in which the associated $\{f_j,\ g_j\}$ solve the functional equations. For example, Theorem 1 stipulates $j \in \Omega_+$ iff $(\mu + \beta\mathcal{G}_j)p_j > \beta\lambda\mathcal{F}_j$, and $j \in \Omega_-$ otherwise. This indicates that, given j, if constraint (37b) is not binding and constraint (37c) is binding at the optimal solution, then $j \in \Omega_+$. Otherwise, $j \in \Omega_-$. This leads to the following corollary.

Corollary 7. *An optimal extreme point solution to* (37) *yields the following optimal stationary* *policy in the dynamic program* (11). *For all* $j \in \Omega$*,*

$$j \in \begin{cases} \Omega_+ & \textit{if } (37c) \textit{ is binding and } (37b) \textit{ has positive slack.} \\ \Omega_- & \textit{otherwise.} \end{cases}$$

$$j \in \begin{cases} \Omega_0 & \textit{if } (37f) \textit{ is binding, and } (37d) \textit{ and } (37e) \textit{ have positive slack.} \\ \Omega_I & \textit{if } (37e) \textit{ is binding and } (37d) \textit{ has positive slack.} \\ \Omega_x & \textit{otherwise.} \end{cases}$$

7.2 Computational improvement

A significant computational improvement is implicit in Theorem 1's statement that the solution of the pair of functional equations (20) is equivalent to a solution of dynamic program (11). This subsection compares computations based on (11) with (20) using as the metric the product of the number of states with the cumulative number of actions in all states. The curse of dimensionality rapidly makes (11) impractical but does not affect (20).

In order to solve dynamic program (11) computationally, its sets of states and actions must first be discretized and made finite. Thus far, the state space is $\mathcal{S} = \{(K, W, j)\} = \Re_+^2 \times \Omega$. After discretizing the possible capacity levels K, capital levels W, and exogenous states j, let ω denote the number of elements in the exogenous state space Ω, and let $\bar{K}$ and $\bar{W}$ denote the maximum levels of capacity and capital, respectively.

The number of elements in $\mathcal{S}$ after discretization is $(\bar{K}+1)(\bar{W}+1)\omega$, and the total number of feasible actions across all states is approximately $\omega(\bar{K}+1)(\bar{K}+2)(\bar{W}+1)(2\bar{W}^2+7\bar{W}+6)/24$ (see Appendix §C for detailed calculations). This indicates that the linear program employed to solve (11) has $(\bar{K}+1)(\bar{W}+1)\omega$ variables, about $\omega(\bar{K}+1)(\bar{K}+2)(\bar{W}+1)(2\bar{W}^2+7\bar{W}+6)/24$ constraints, and thus $(\bar{K}+1)^2(\bar{W}+1)^2\omega^2(\bar{K}+2)(2\bar{W}^2+7\bar{W}+6)/24$ elements in the coefficient matrix. When $\omega = 5$ and $\bar{K} = \bar{W} = 10$, for example, the primal problem would have 605 variables, approximately 2×10^6 constraints, and approximately 1.2×10^9 elements in the coefficient matrix. If the size of Ω remains at 5 but $\bar{K}$ and $\bar{W}$ both double, then the primal problem would have 2,205 variables, approximately 45.9×10^6 constraints, and approximately 4.4×10^9 elements in the coefficient matrix, manifesting the curse of dimensionality.

Solving (20) using linear program (37), on the other hand, is much easier. The number of elements in $\{f(j), g(j) : j \in \Omega\}$ is 2ω, compared to its counterpart of $(\bar{K}+1)(\bar{W}+1)\omega$ in (11).

The total number of "actions" is 5ω (two in (20a) and three in (20b)), compared to its counterpart of $\omega(\bar{K}+1)(\bar{K}+2)(\bar{W}+1)(2\bar{W}^2+7\bar{W}+6)/24$ in (11). This yields a linear program with 2ω variables, 5ω constraints, and thus $10\omega^2$ elements in the coefficient matrix. If $\omega = 5$, there are 10 variables and 25 constraints, *regardless of the magnitude of* $\bar{K}$ *and* $\bar{W}$. Algorithms based on (20), for example linear program (37) and successive approximations (13), are immune from the curse of dimensionality.

Further computational improvements can be achieved in linear program (37). After making the change of variable $g'_j = g_j - (1-\mu)$, all ω constraints (37d) are made implicit and become $g'_j \geq 0$. This change reduces the number of explicit constraints by twenty percent, resulting in a coefficient matrix with 2ω columns, and 4ω rows and thus in total $8\omega^2$ elements instead of $10\omega^2$. This strengthens the already overwhelming advantage of solving (20) instead of (11).

8 Asset Unit Values, the Optimal Policy, and the Optimization Criterion

Recall that the optimization criterion is $\mu\times$EPV(profits)+$(1-\mu)\times$EPV(dividends), $0 \leq \mu \leq 1$. This section examines how a change in the optimization criterion affects the asset unit values and the optimal policy by studying how the unit values and the policy depend on μ. It follows from a standard property of linear programs that when μ varies, the optimal solution, i.e., the asset unit values, also changes. How the optimal basis and thus the optimal policy change, however, is unclear. Intuitively, when μ varies slightly, an optimal basis of the linear program (37) may remain the same, thus leaving the optimal policy unaltered. A sufficiently large change in μ, on the other hand, may cause a change in the optimal basis and thus the optimal policy. The dependence of asset unit values and the optimal policy is investigated in-depth in this section by exploring the structure of the problem.

Henceforth, we write $\{\Omega_+(\mu), \Omega_-(\mu)\}$, and $\{\Omega_x(\mu), \Omega_I(\mu), \Omega_0(\mu)\}$ for the partitions, and $f_j(\mu)$, $g_j(\mu)$, $\mathcal{F}_j(\mu)$, $\mathcal{F}^e_j(\mu)$, and $\mathcal{G}_j(\mu)$ $(j \in \Omega)$ for the asset unit values to make the their dependence on μ explicit.

8.1 Dependence on μ of the asset unit values

Linear program (37) yields the following statements about $f_j(\cdot)$ and $g_j(\cdot)$ $(j \in \Omega)$.

Lemma 6. *For each $j \in \Omega$, the functions $f_j(\cdot)$ and $g_j(\cdot)$ are convex, continuous, and piece-wise linear on $[0,1]$. The breakpoints in $f_j(\cdot)$ and $g_j(\cdot)$ correspond to a change in the optimal basis in (37). The functions $\mathcal{F}_j(\cdot)$, $\mathcal{F}_j^e(\cdot)$, and $\mathcal{G}_j(\cdot)$ are convex and continuous on $[0,1]$, and they are piece-wise linear if Ω is a finite set.*

An implication of Lemma 6 is that the maxima and minima of the asset unit values, as functions of μ, are achieved either at $\mu = 0$ (corresponding to the profit criterion), or $\mu = 1$ (corresponding to the dividend criterion). Using the results developed earlier in §6, the unit value of capacity can be shown to be strictly increasing with respect to μ on $[0,1]$. Therefore, capacity has the highest value when profit generation is the sole focus, and the lowest value when dividend issuance is the sole focus. The following lemma and the subsequent theorem establish this important insight.

Lemma 7. *On each subinterval of $[0,1]$ where the optimal basis of the linear program (37) remains unchanged as μ varies, for all $j \in \Omega$ the functions $f_j(\mu)$ and $\mathcal{F}_j(\cdot)$ increase.*

Theorem 2. *For each $j \in \Omega$, $f_j(\cdot)$ and $\mathcal{F}_j(\cdot)$ are convex, continuous, and (strictly) increasing functions on $[0,1]$. The function $f_j(\cdot)$ is piece-wise linear and, if Ω is finite, so is $\mathcal{F}_j(\cdot)$.*

8.2 Dependence on μ of optimal dividend, investment, and liquidity decisions

The following proposition confirms the intuition that the optimal dividend is a weakly increasing function of the amount of emphasis placed on dividends in the criterion. That is, given capacity K, capital W, and the exogenous state j, when a firm heightens the emphasis on maximizing profit, i.e., μ increases in the optimization criterion, then the amount of dividend either remains the same or falls. Therefore, a firm under-distributes dividends if it aspires to maximize its value ($\mu = 0$) but actually optimizes profits ($\mu = 1$).

Since the firm's optimal dividend is either 0 or the entire capital W, weak monotonicity of the dividend decision implies that there is a threshold weight on profit such that all capital is distributed as dividend if μ is less than the threshold. Formally,

Proposition 2. *For all $(K, W, j) \in \mathcal{S}$, there is a threshold $\mu^x(j)$ such that $X(\mu; K, W, j) = W$ if $\mu \le \mu^x(j)$, and $X(\mu; K, W, j) = 0$ if $\mu > \mu^x(j)$.*

Proposition 2 asserts that the threshold weight depends only on the exogenous state and not at all on existing asset levels. The invariance of the threshold with respect to asset levels may seem counterintuitive because one expects a firm with abundant assets to behave differently from one with few assets. However, when the optimal dividend is positive, its amount is W, which is completely dependent on the amount of capital.

Proposition 2 also explains how the firm's investment *attitude* varies with μ. Note that as opposed to dividend issuance which permanently removes capital from the firm, both immediate investment and liquidity buildup retain capital, thus enabling it to be fed into operations that eventually generate profit. (Recall that by Corollary 4, capital is used to build liquidity only if there is positive probability of investing it in the future.) In this sense, Proposition 2 implies that *the total capital reserved for investment,* $W - X(\mu; K, W, j)$*, is nondecreasing with respect to* $\mu \in [0, 1]$. That is, *a profit criterion (*$\mu = 1$*) induces a (weakly) more aggressive investment attitude than does a dividend criterion (*$\mu = 0$*).*

We emphasize the word *attitude* because the investment amount, $I(\mu; K, W, j)$, is not necessarily monotone in μ. To see why this might be true, consider a state j in which the optimal decision is immediate investment when the emphasis on profit, μ, is low. As μ increases, it may become worthwhile for the firm to build liquidity instead of investing immediately, so that more capital is available in the future to exploit an exogenous state reflecting a more favorable combination of market and investment conditions, thus generating a higher EPV of overall profit. As μ increases even higher, however, the optimal decision in state j may return to investment. See Appendix D for a more detailed analysis.

9 Summary and Extensions

Firms in diverse industries have capacity constraints on the quantity they can produce, and capital constraints on the amount they can invest to enhance capacity, retain to increase liquidity, and issue as dividends. Selling the production output augments capital, the amount of which depends partly on the exogenous market price. Investing the capital creates capacity, the amount of which depends partly on random investment yields. Modeling the market price and investment yields as Markov-modulated, we develop and analyze a discrete-time infinite-horizon Markov decision process in which a generic firm decides each period how much capacity to use for production, and how much capital to invest, distribute, and retain, while trying to maximize the EPV of profits, or dividends, or a mix of both.

Under linearity assumptions regarding investment yield and production cost, we show that the optimal policy is *extremal*. That is, given existing capacity and capital, the firm should either produce nil or maximally, and should invest, distribute and retain either nil or maximally. Furthermore, whether the production, investment, dividend, and liquidity amount is zero or positive is determined solely by the exogenous uncertainty underlying market prices and investment yields. This leads to a linear program for the Markov decision process that exorcises the curse of dimensionality.

Our results highlight the real option nature of capital and capacity, and explain why their values are interdependent. We show that, given capacity, capital, and exogenous uncertainty, the value of capacity rises as the firm increasingly emphases profit maximization. Thus, maximizing profit induces a more aggressive investment attitude than maximizing dividend because, all else equal, a firm which aims to maximize profit perceives the investment outcome—the capacity—to be more valuable.

Extensions of the model that would preserve major results include interest rates on liquidity, and depreciation rates of capacity, both driven by the exogenous Markov process. The model has a mathematical structure that we generalize and apply elsewhere (Ning and Sobel 2015).

References

Allen, F., R. Michaely. 1995. Dividend policy. R. A. Jarrow, V. Maksimovic, W. T. Ziemba, eds., *Finance. Handbooks in Operation Research and Management Science, Vol. 9*. North-Holland, Amsterdam, 793–838.

Amit, Raphael. 1986. Petroleum reservoir exploitation: Switching from primary to secondary recover. *Operations Research* **34(4)** 534–549.

Anderson, Soren T, Ryan Kellogg, Stephen W Salant. 2014. Hotelling under pressure. Tech. rep., mimeo, University of Michigan.

Arrow, Kenneth J., Sheldon Chang. 1982. Optimal pricing, use, and exploration of uncertain natural resource stocks. *Journal of Environmental Economics and Management* **9(1)** 1–10.

Asvanunt, Attakrit, Mark Broadie, Suresh Sundaresan. 2010. Growth options and optimal default under liquidity constraints: The role of corporate cash balances. Tech. rep., Columbia Business School Research Paper.

Bates, Thomas, Kathleen Kahle, Rene Stulz. 2009. Why do u.s. firms hold so much more cash than they used to? *Journal of Finance* **64** 1985–2021.

Bolton, Patrick, Hui Chen, Neng Wang. 2011. A unified theory of tobin's q, corporate investment, financing, and risk management. *Journal of Finance* **66(5)** 1545–1578.

Boyle, Glenn W., Graeme A. Guthrie. 2003. Investment, uncertainty, and liquidity. *Journal of Finance* **58(5)** 2143–2166.

Chao, Xiuli, Hong Chen, Shaohui Zheng. 2009. Dynamic capacity expansion for a service firm with capacity deterioration and supply uncertainty. *Operations Research* **57(1)** 82–93.

Cochrane, John H. 2005. *Asset Pricing*. Princeton University Press, Princeton, N.J.

Cortazar, Gonzalo, Eduardo S. Schwartz, Jaime Casassus. 2001. Optimal exploration investments under price and geological-technical uncertainty: a real option model. *R&D management* **31(2)** 181–189.

Dantzig, George B. 1963. *Linear Programming and Extensions*. Princeton University Press, Princeton, N.J.

Denardo, Eric V. 1967. Contraction mapping in the theory underlying dynamic programming. *SIAM Review* **9** 165–177.

Deshmukh, Sudhakar D., Stanley R. Pliska. 1980. Optimal consumption and exploration of nonrenewable resources under uncertainty. *Econometrica* **48(1)** 177–200.

Dixit, Avinash K., Robert S. Pindyck. 1994. *Investment under uncertainty*. Princeton University Press, Princeton, N.J.

Falato, Antonio, Dalida Kadyrzhanova, Jae Sim. 2013. Rising intangible capital, shrinking debt capacity, and the us corporate savings glut. Tech. rep., FEDS Working Paper.

Gamba, Andrea, Alexander J. Triantis. 2008. The value of financial flexibility. *Journal of Finance* **63(5)** 2263–2296.

Heyman, Daniel P., Matthew J. Sobel. 1984. *Stochastic Models in Operations Research, Vol. II: Stochastic Optimization*. McGraw Hill, New York, N.Y.

Iglehart, Donald L. 1965. Capital accumulation and production for the firm: Optimal dynamic policies. *Management Science* **14** 193–205.

Li, Lode, Martin Shubik, Matthew J. Sobel. 2013. Control of dividends, capital subscriptions, and physical inventories. *Management Science* **59(5)** 1107–1124.

Lippman, S. A., A. J. Rolfe, H. M. Wagner, J. S. C. Yuan. 1967. Optimal production scheduling and employment smoothing with deterministic demands. *Management Science* **14** 127–158.

Luss, Hannan. 1982. Operations research and capacity expansion problems: A survey. *Operations Research* **30** 907–947.

Ma, Liang, Antonio S. Mello, Youchang Wu. 2013. Industry competition, winner's advantage, and cash holdings. Working paper.

Manne, Alan S. 1967. *Investments for Capacity Expansion: Size, Location and Time-Phasing*. MIT Press.

Mauer, David C., Alexander J. Triantis. 1994. Interactions of corporate financing and investment decisions: A dynamic framework. *Journal of Finance* **49(4)** 1253–1277.

Modigliani, Franco, Merton H. Miller. 1958. The cost of capital, corporation finance, and the theory of investment. *American Economic Review* **48(3)** 261–297.

Ning, Jie, Matthew J. Sobel. 2015. Affine value mdps and their applications. Working paper, Dept. of Operations, Case Western Reserve University.

Paddock, James L., Daniel R. Siegel, James L. Smith. 1988. Option valuation of claims on real assets: The case of offshore petroleum leases. *The Quarterly Journal of Economics* **103(3)** 479–503.

Pindyck, R. S. 1981. The optimal production of an exhaustible resource when price is exogenous and stochastic. *The Scandinavian Journal of Economics* 277–288.

Pindyck, R. S. 1988. Irreversible investment, capacity choice, and the value of the firm. *The American Economic Review* **78(5)** 969–985.

Pindyck, Robert S. 1978. The optimal exploration and production of nonrenewable resources. *Journal of Political Economy* **86(5)** 841–861.

Radner, Roy, Larry Shepp. 1996. Risk vs. profit potential: A model for corporate strategy. *Journal of economic dynamics and Control* **20(8)** 1373–1393.

Sobel, Matthew J. 1970. Employment smoothing (capital accumulation) with production for stochastic demand. *Management Science* **16(5)** 340–349.

Sundaresan, Suresh, Neng Wang. 2007. Investment under uncertainty with strategic debt service. *The American Economic Review* **97(2)** 256–261.

Titman, Sheridan, Sergey Tsyplakov. 2007. A dynamic model of optimal capital structure. *Review of Finance* **11** 401–451.

Van Mieghem, J. A. 2003. Capacity management, investment and hedging: Review and recent developments. *Manufacturing & Service Operations Management* **5(4)** 269–302.

Wang, Wenbin, Mark E. Ferguson, Shanshan Hu, Gilvan C. Souza. 2013. Dynamic capacity investment with two competing technologies. *Manufacturing & Service Operations Management* **15(4)** 616–629.

Proofs of Statements and Additional Details

A Proofs

A.1 Proofs of statements in §4

Lemma 1. *For all* $(K, W, j) \in \mathcal{S}$ *and* $n = 1, 2, \ldots,$ $V_n(K, W, j) \leq V_{n+1}(K, W, j)$. *The limit* $\bar{V}(K, W, j) = \lim_{n\to\infty} V_n(K, W, j)$ *exists and solves* (8). *If* Ω *is compact, then there is a unique (bounded) solution of* (8) *and thus* $\bar{V}(\cdot, \cdot, \cdot) = V(\cdot, \cdot, \cdot)$.

Proof. The first step proves $V_n(K, W, j) \leq V_{n+1}(K, W, j)$ for all $(K, W, j) \in \mathcal{S}$ and $n = 1, 2, \ldots$ by induction starting with $n = 0$ and $V_0(\cdot, \cdot, \cdot) \equiv 0$. Setting $n = 1$ in (9), $0 = V_0(K, W, j) \leq V_1(K, W, j)$ for all $(K, W, j) \in \mathcal{S}$. Make the inductive assumption for any $n \geq 1$ that $V_{n-1}(K, W, j) \leq V_n(K, W, j)$ for all $(K, W, j) \in \mathcal{S}$. Replacing n with $n + 1$ in (9b),

$$
\begin{aligned}
J_{n+1}(x, i, q; K, W, j) &= \mu p(j)q + (1 - \mu)x \\
&\qquad + \beta\, \mathbb{E}\left[V_n\big(\theta K - \lambda q + e(S_j)i, W + p(j)q - x - i, S_j\big)\right] \\
&\geq \mu p(j)q + (1 - \mu)x \\
&\qquad + \beta\, \mathbb{E}\left[V_{n-1}\big(\theta K - \lambda q + e(S_j)i, W + p(j)q - x - i, S_j\big)\right] \\
&= J_n(x, i, q; K, W, j).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
V_{n+1}(K, W, j) &= \max_{(x,i,q)} \{J_{n+1}(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\} \\
&\geq \max_{(x,i,q)} \{J_n(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\} \\
&= V_n(K, W, j)
\end{aligned}
$$

which completes the induction.

The second step is the observation that monotonicity in n, the Boundedness Assumption, and the monotone convergence theorem imply the existence of the limit $\bar{V}(K, W, j) = \lim_{n\to\infty} V_n(K, W, j)$.

The third step shows that $\bar{V}(\cdot, \cdot, \cdot)$ solves (8). Since $V_n(K, W, j)$ converges to $\bar{V}(K, W, j)$ and

$\bar{V}(K, W, j) \geq V_n(K, W, j)$ for all n, (9) implies

$$\bar{V}(K, W, j) \geq \lim_{n\to\infty} \max_{(x,i,q)} \{J_n(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\}.$$

Given (K, W, j), the feasible decision vectors (x, i, q) lie in a compact set, and it can be shown that convergence in n is uniform on every compact subset of $\mathcal{S}$. Therefore, the limit and maximum can be interchanged, so

$$\begin{aligned}\bar{V}(K, W, j) &\geq \max_{(x,i,q)} \{\lim_{n\to\infty} J_n(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\} \\ &= \max_{(x,i,q)} \{\lim_{n\to\infty} J_n(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\} \\ &= \max_{(x,i,q)} \{\bar{J}(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\}\end{aligned}$$

where $\bar{J}$ and J are the same except that $\beta\bar{V}$ replaces βV in (8c). In order to establish the reverse inequality, hence equality, $V_n(K, W, j) \leq \bar{V}(K, W, j)$ and (9) imply

$$V_n(K, W, j) \leq \max_{(x,i,q)} \{\bar{J}(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\}.$$

Take the limit as $n \to \infty$ to obtain

$$\bar{V}(K, W, j) \leq \max_{(x,i,q)} \{\bar{J}(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ i + x - p(j)q \leq W,\ 0 \leq q \leq K\}.$$

Therefore, equality prevails and the limit function $\bar{V}(\cdot, \cdot, \cdot)$ satisfies (8).

The proof is completed by establishing the uniqueness of the solution to (8) when Ω is compact. The basic idea, only sketched here, is to compactify the state space $\mathcal{S} = \Re^2_+ \times \Omega$, invoke the Boundedness Assumption, and then apply the contraction mapping argument in Denardo (1967). First, modify the dynamics with

$$K_{t+1} = \min\{\bar{K},\ \theta K_t - \lambda q_t + e(s_t)i_t\} \qquad \text{and} \qquad W_{t+1} = \min\{\bar{W},\ W_t - x_t - i_t + p(s_t)q_t\}$$

where $\bar{K} < \infty$ and $\bar{W} < \infty$. The assumption that Ω is compact implies that the comparably modified (8) has a compact state space and a compact set of feasible actions. Therefore, it satisfies the boundedness, continuity, and compactness assumptions in Denardo (1967) and, thus, has a unique value function *which depends on $\bar{K}$ and $\bar{W}$*. However, the Boundedness Assumption implies that, as $\bar{K}$ and $\bar{W} \to \infty$, at each argument the value function converges to a finite limit. $\square$

Lemma 2. *For $n = 1, 2, \ldots,$ the value function $V_n(\cdot, \cdot, j)$ in (9) satisfies the following recursion with $V_0(\cdot, \cdot, \cdot) \equiv 0$ and for $n = 1, 2, ...,$*

$$V_n(K, W, j) = \max_{(x,i,q)} \{J_n(x, i, q; K, W, j) : 0 \leq x,\ 0 \leq i,\ i + x \leq W,\ 0 \leq q \leq K\}, \tag{38a}$$

$$J_n(x, i, q; K, W, j) = \mu p(j) q + (1 - \mu) x + \beta\, \mathbb{E}\left[V_{n-1}\big(\theta K - \lambda q + e(S_j) i, W + p(j) q - x - i, S_j\big)\right]. \tag{38b}$$

Proof. Dynamic program (38) differs from (9) only in that it lacks the constraint $i + x - p(j)q \leq W$. The proof shows that $i + x - p(j)q \leq W$ can be deleted from (9a) without affecting optimality by discussing two scenarios: (i) $(K, W) \in \Re^2_+$ and $j \in \{j \in \Omega : p(j) \geq 0\}$, and (ii) $(K, W) \in \Re^2_+$ and $j \in \{j \in \Omega : p(j) < 0\}$.

When $(K, W) \in \Re^2_+$ and $j \in \{j \in \Omega : p(j) \geq 0\}$, $i + x \leq W$ implies that $i + x - p(j)q \leq W$ is redundant in (9a) and its deletion does not affect optimality.

The remainder of the proof shows that $(K, W) \in \Re^2_+$ and $j \in \{j \in \Omega : p(j) < 0\}$ implies that $q = 0$ is an optimal solution to (9a), i.e., $J_n(x, i, q; K, W, j) \leq J_n(x, i, 0; K, W, j)$. Therefore, $i + x - p(j)q \leq W$ reduces to $i + x \leq W$ when $J_n(x, i, q; K, W, j)$ achieves its maximum, so $i + x - p(j)q \leq W$ is necessarily satisfied when $i + x \leq W$.

The proof of $J_n(x, i, q; K, W, j) \leq J_n(x, i, 0; K, W, j)$ has two steps. First, the following induction proves that $V_n(\cdot, \cdot, j)$ is nondecreasing on its domain (for each $j \in \Omega$ and $n = 1, 2, ...$). Initiate the induction with $V_0(\cdot, \cdot, j) \equiv 0$ which is nondecreasing. If $V_{n-1}(\cdot, \cdot, j)$ is nondecreasing for an $n \geq 0$, then in (9b) $J_n(\cdot, \cdot, j)$ is nondecreasing. Furthermore, increasing K and W loosens the constraints in (9a), which cannot damage the value of the objective $V_n(\cdot, \cdot, j)$. Thus, $V_n(\cdot, \cdot, j)$ is nondecreasing on its domain, which completes the induction.

Second, by (9b), if $p(j) < 0$, then the immediate reward from $q = 0$ is greater than that from $q > 0$. Monotonicity of $V_{n-1}(\cdot, \cdot, j)$ implies that the EPV of the future reward is also at least as great. Therefore, $J_n(x, i, q; K, W, j) \leq J_n(x, i, 0; K, W, j)$. $\square$

A.2 Proofs of statements in §5

Theorem 1. *1. There are real-valued functions $f(\cdot)$ and $g(\cdot)$ on Ω such that*

$$V(K,W,j) = f(j)K + g(j)W, \qquad (W,K,j) \in \mathcal{S}. \tag{39}$$

2. Define $\mathcal{F}(j) = \mathbb{E}[f(S_j)]$, $\mathcal{F}^e(j) = \mathbb{E}[f(S_j)e(S_j)]$, and $\mathcal{G}(j) = \mathbb{E}[g(S_j)]$ $(j \in \Omega)$. The functions $f(\cdot)$ and $g(\cdot)$ satisfy

$$f(j) = \max\left\{\beta\theta\mathcal{F}(j),\ \beta(\theta-\lambda)\mathcal{F}(j) + [\mu + \beta\mathcal{G}(j)]p(j)\right\}, \tag{40a}$$

$$g(j) = \max\{1-\mu,\ \beta\mathcal{F}^e(j),\ \beta\mathcal{G}(j)\}. \tag{40b}$$

3. Define the following subsets of Ω:

$$\Omega_+ = \{j \in \Omega : [\mu + \beta\mathcal{G}(j)]p(j) > \beta\lambda\mathcal{F}(j)\}, \tag{41a}$$

$$\Omega_- = \{j \in \Omega : [\mu + \beta\mathcal{G}(j)]p(j) \le \beta\lambda\mathcal{F}(j)\}, \tag{41b}$$

and

$$\Omega_x = \{j \in \Omega : \beta\mathcal{F}^e(j) \le 1-\mu, \beta\mathcal{G}(j) \le 1-\mu\}, \tag{42a}$$

$$\Omega_I = \{j \in \Omega : \beta\mathcal{F}^e(j) > 1-\mu, \mathcal{F}^e(j) \ge \mathcal{G}(j)\}, \tag{42b}$$

$$\Omega_0 = \{j \in \Omega : \mathcal{F}^e(j) < \mathcal{G}(j), \beta\mathcal{G}(j) > 1-\mu\}. \tag{42c}$$

Then $\{\Omega_0, \Omega_x, \Omega_I\}$ and $\{\Omega_+, \Omega_-\}$ partition Ω:

$$1(j \in \Omega_x) + 1(j \in \Omega_I) + 1(j \in \Omega_0) = 1(j \in \Omega_+) + 1(j \in \Omega_-) = 1. \tag{43}$$

There is an optimal stationary policy $\big(Q(K,W,j),\ X(K,W,j), I(K,W,j)\big)$ that satisfies

$$Q(K,W,j) = 1\,(j \in \Omega_+)\,K, \tag{44}$$

$$X(K,W,j) = 1\,(j \in \Omega_x)\,W, \ \text{and} \tag{45}$$

$$I(K,W,j) = 1\,(j \in \Omega_I)\,W. \tag{46}$$

Proof. 1. From Lemma 1, $V_n(K,W,j) \to V(K,W,j)$ as $n \to \infty$, and Lemma 3 states $V_n(K,W,j) = f_n(j)K + g_n(j)W$. So there are functions $f(\cdot)$ and $g(\cdot)$ on Ω such that $V(K,W,j) = f(j)K + g(j)W$ ($(K,W,j) \in \mathcal{S}$), which confirms (19).

2. The proof of the second part, namely (40), is similar to that of the first. An induction starting with $n = 0$ and $f_0(\cdot) = g_0(\cdot) \equiv 0$ and using (13) establishes $f_n(j) \le f_{n+1}(j)$ and $g_n(j) \le g_{n+1}(j)$ for each n and j. The Boundedness Assumption implies that $f(j) < \infty$ and $g(j) < \infty$. Therefore, $f_n(j) \to f(j)$ and $g_n(j) \to g(j)$ as $n \to \infty$ ($j \in \Omega$), and thus $f(\cdot)$ and $g(\cdot)$ satisfy (40).

3. The third part is established with an infinite-horizon analogue of the series of expressions that culminate in (14) and (15):

$$\begin{aligned} V(K,W,j) = \max_{(x,i,q)} \{ & \mu p(j)q + (1-\mu)x + \beta\,\mathbb{E}\left[V\big(\theta K - \lambda q + e(S_j)i, W - x - i + p(j)q, S_j\big)\right] : \\ & 0 \le x,\ 0 \le i,\ i + x \le W,\ 0 \le q \le K\} \\ = {} & \beta\theta\mathcal{F}(j)K + \max_q\{[[\mu + \beta\mathcal{G}(j)]p(j) - \beta\lambda\mathcal{F}(j)]q : 0 \le q \le K\} \qquad (47) \\ & + \beta\mathcal{G}(j)W + \max_{(x,i)}\{[1 - \mu - \beta\mathcal{G}(j)]x + \beta[\mathcal{F}^e(j) - \mathcal{G}(j)]i : x, i \ge 0,\ i + x \le W\}. \qquad (48) \end{aligned}$$

Optimization (47) is a linear program with the decision variable q whose optimal value is

$$Q(K,W,j) = \begin{cases} K & \text{if } \beta\theta\mathcal{F}(j) > \beta(\theta - \lambda)\mathcal{F}(j) + [\mu + \mathcal{G}(j)]p(j) \\ 0 & \text{if } \beta\theta\mathcal{F}(j) \le \beta(\theta - \lambda)\mathcal{F}(j) + [\mu + \mathcal{G}(j)]p(j) \end{cases}$$

After simplification, this yields sets Ω_+ and Ω_- defined in (41), thus establishing the optimal production decision in (44). Clearly, $\{\Omega_+,\ \Omega_-\}$ partitions the state space Ω.

Optimization (48) is a linear program with optimal solution

$$\Big(X(K,W,j), I(K,W,j)\Big) = \begin{cases} (W,0) & \text{if } \beta\mathcal{F}^e(j) \le 1 - \mu,\ \beta\mathcal{G}(j) \le 1 - \mu \\ (0,W) & \text{if } \beta\mathcal{F}^e(j) > 1 - \mu,\ \mathcal{F}^e(j) \ge \mathcal{G}(j) \\ (0,0) & \text{if } \mathcal{F}^e(j) < \mathcal{G}(j),\ \beta\mathcal{G}(j) > 1 - \mu \end{cases}$$

This yields sets Ω_x, Ω_I and Ω_0 defined in (42), thus establishing the optimal dividend and investment decisions in (45) and (46). Clearly, $\{\Omega_x,\ \Omega_I,\ \Omega_0\}$ partitions Ω. □

A.3 Proofs of statements in §6

A.3.1 Proofs of statements in §6.1

Lemma 4. *The unit value of the capacity when it is idle,* $\mathbf{f}^-$*, and the unit value when it is used for production,* $\mathbf{f}^+$*, satisfy*

$$\mathbf{f}^- = \beta\theta\mathbf{A}\mathbf{f}^+, \tag{49a}$$

$$\mathbf{f}^+ = \beta(\theta - \lambda)\mathbf{B}\mathbf{f}^+ + diag(\mathbf{p}^+)\left(\mu\mathbf{1}_+ + \beta\boldsymbol{\mathcal{G}}^+\right), \tag{49b}$$

where

$$\mathbf{A} \equiv \sum_{r=0}^{\infty}(\beta\theta\mathbf{M}_-^-)^r\mathbf{M}_-^+, \tag{50}$$

$$\mathbf{B} \equiv \beta\theta\mathbf{M}_+^-\mathbf{A} + \mathbf{M}_+^+. \tag{51}$$

Proof. The functional equation of f_j and the partition Ω_+ yield

$$\mathbf{f}^- = \beta\theta\left(\mathbf{M}_-^-\mathbf{f}^- + \mathbf{M}_-^+\mathbf{f}^+\right), \tag{52a}$$

$$\mathbf{f}^+ = \beta(\theta - \lambda)\left(\mathbf{M}_+^-\mathbf{f}^- + \mathbf{M}_+^+\mathbf{f}^+\right) + \mathrm{diag}(\mathbf{p}^+)\left(\mu\mathbf{1}_+ + \beta\boldsymbol{\mathcal{G}}^+\right). \tag{52b}$$

Simplifying (52a),

$$(\mathbf{I}_- - \beta\theta\mathbf{M}_-^-)\mathbf{f}^- = \beta\theta\mathbf{M}_-^+\mathbf{f}^+. \tag{53}$$

Because $\mathbf{M}_-^-$ is a submatrix of the stochastic matrix $\mathbf{M}$, $\mathbf{I}_- - \beta\mathbf{M}_-^-$ is nonsingular and satisfy,

$$\left(\mathbf{I}_- - \beta\theta\mathbf{M}_-^-\right)^{-1} = \sum_{r=0}^{\infty}(\beta\theta\mathbf{M}_-^-)^r,$$

Combining this equation with (52a) establishes (49a).

A similar simplification yields (49b) after simplifying (52b) using (49a). □

Corollary 1. *1. For all* $j \in \Omega$*,* $f_j \geq 0$*.*

2. $f_j = 0$ *if and only if* $j \in \Omega_-$ *and* Ω_+ *is inaccessible from* j*.*

3. If $\Omega_+ \neq \emptyset$*, then* $\arg\max\ f_j \in \Omega_+$*.*

Proof. 1. In (13), a straightforward induction on n beginning with $f_0 \equiv 0$ yields, for all $j \in \Omega$ and $n = 1, 2, ...$, (i) $f_n(j) \geq 0$, and (ii) $f_n(j) \leq f_{n+1}(j)$. Therefore, the Boundedness Assumption and the monotone convergence theorem imply that for each $j \in \Omega$, $f_n(j) \geq 0$ converges to $f(j) \geq 0$ as $n \to \infty$.

2. First prove that $f_j > 0$ if $j \in \Omega_+$. From (40),

$$f_j = \beta(\theta - \lambda)\mathcal{F}_j + (\mu + \beta\mathcal{G}_j)p_j > \beta\theta\mathcal{F}_j \geq 0,$$

where the first inequality follows from the definition of Ω_+ and the second from part 1.

Hence, $f_j = 0$ only when $j \in \Omega_-$. By definition, j cannot access Ω_+ if and only if its row in matrix $\mathbf{A}$ is all 0's. Then it follows from (52a) that $f_j = 0$.

3. The row sums of $\beta\theta\mathbf{A}$ (defined in (50)) are nonnegative and strictly less than 1, so $\max_{j\in\Omega_-} f_j < \max_{j\in\Omega_+} f_j$ by (52a), establishing the statement. □

Corollary 2. $\mathbf{p}^+ > 0$.

Proof. From the proof of Lemma 2, the optimal production decision $Q(K, W, j) = 0$ if $p_j < 0$. Thus, $\mathbf{p}^+$ must have nonnegative entries. By (21a) in Theorem 1, $j \in \Omega_+$ if and only if

$$[\mu + \beta\mathcal{G}_j]p_j > \beta\lambda\mathcal{F}_j.$$

On the left side, $\mu + \beta\mathcal{G}_j \geq \mu + \beta(1 - \mu) \geq \beta > 0$, and on the right side, $f_j \geq 0$ for all $j \in \Omega$ according to Corollary 1 which implies $\mathcal{F}_j \geq 0$. So this inequality would be violated if $p_j = 0$. Therefore, $\mathbf{p}^+$ has only positive entries. □

Corollary 3. *The expected unit value of capacity next period which is idle now,* $\boldsymbol{\mathcal{F}}^-$, *and the expected unit value of capacity next period which produces now,* $\boldsymbol{\mathcal{F}}^+$, *satisfy*

$$\boldsymbol{\mathcal{F}}^- = \mathbf{A}\mathbf{f}^+, \tag{54a}$$

$$\boldsymbol{\mathcal{F}}^+ = \mathbf{B}\mathbf{f}^+. \tag{54b}$$

Proof. This is a direct result of the previous corollaries. □

A.3.2 Proof of statement in §6.2

Lemma 5. *The unit values of capital, according to whether the optimal decision is dividend issuance* $\mathbf{g}^x$*, capacity investment* $\mathbf{g}^I$*, or increased liquidity* $\mathbf{g}^0$*, are as follows:*

$$\mathbf{g}^x = (1-\mu)\mathbf{1}_x, \tag{55a}$$

$$\mathbf{g}^I = \beta\mathbf{M}_I^e\mathbf{f}, \tag{55b}$$

$$\mathbf{g}^0 = \beta\mathbf{C}\mathbf{1}_x(1-\mu) + \beta^2\mathbf{D}\mathbf{M}_I^e\mathbf{f}, \tag{55c}$$

where

$$\mathbf{C} = \sum_{r=0}^{\infty}\left(\beta\mathbf{M}_0^0\right)^r\mathbf{M}_0^x, \tag{56a}$$

$$\mathbf{D} = \sum_{r=0}^{\infty}\left(\beta\mathbf{M}_0^0\right)^r\mathbf{M}_0^I. \tag{56b}$$

Proof. From (20) in Theorem 1,

$$\mathbf{g}^x = (1-\mu)\mathbf{1}_x, \tag{57a}$$

$$\mathbf{g}^I = \beta\boldsymbol{\mathcal{F}}^{eI}, \tag{57b}$$

$$\mathbf{g}^0 = \beta\boldsymbol{\mathcal{G}}^0 \tag{57c}$$

which confirms (55a). Equation (55b) results from combining $\boldsymbol{\mathcal{F}}^{eI} = \mathbf{M}_I^e\mathbf{f}$ and (57b).

In order to prove $\mathbf{g}^0 = \beta\boldsymbol{\mathcal{G}}^0$, combine (32c), (55a), and (55b) to obtain

$$\mathbf{g}^0 = \beta\mathbf{M}_x^0(1-\mu)\mathbf{1}_x + \beta^2\mathbf{M}_I^I\mathbf{M}_I^e\mathbf{f} + \beta\mathbf{M}_0^0\mathbf{g}^0.$$

That is

$$\left(\mathbf{I}_0 - \beta\mathbf{M}_0^0\right)\mathbf{g}^0 = \beta\mathbf{M}_x^0(1-\mu)\mathbf{1}_x + \beta^2\mathbf{M}_I^I\mathbf{M}_I^e\mathbf{f}.$$

Since $\mathbf{M}_0^0$ is a submatrix of the stochastic matrix $\mathbf{M}$, $\mathbf{I}_0 - \beta\mathbf{M}_0^0$ is nonsingular and satisfy,

$$\left(\mathbf{I}_0 - \beta\mathbf{M}_0^0\right)^{-1} = \sum_{r=0}^{\infty}\left(\beta\mathbf{M}_0^0\right)^r,$$

confirming the specifications of $\mathbf{C}$ and $\mathbf{D}$ in (56) and in turn (55c). □

Corollary 4. *If $j \notin \Omega_I$, then $j \in \Omega_0$ only if Ω_I is accessible from j.*

Proof. A proof by contradiction begins with the supposition that $j \in \Omega_0$ but j cannot access Ω_I. Then g_j satisfies (33c) and its corresponding row in $\mathbf{D}$ consists entirely of zeros. Let $\mathbf{C}_j$ be the row in matrix $\mathbf{C}$ corresponding to state j. Then $g_j = \beta \mathbf{C}_j \mathbf{1}_x(1-\mu) < 1-\mu$ which contradicts $g_j \geq 1-\mu$ for all $j \in \Omega$ and establishes the statement. □

Corollary 5. *1. For all $j \in \Omega$, $g_j \geq 1-\mu$,* $\arg\min\ g_j \in \Omega_x$.

2. If $\Omega_I \neq \emptyset$, then $\arg\max\ g_j \in \Omega_I$.

Proof. 1. In (13), a straightforward induction on n beginning with $g_0 \equiv 0$ proves that for all $j \in \Omega$ and $n = 1, 2, ...$: (i) $g_n(j) \geq 1-\mu$, and (ii) $g_n(j) \leq g_{n+1}(j)$. Therefore, the Boundedness Assumption and the monotone convergence theorem imply that for each $j \in \Omega$, $g_n(j) \geq 1-\mu$ converges to $g(j) \geq 1-\mu$ as $n \to \infty$. The definition of Ω_x in Theorem 1 implies $\arg\min\ g_j \in \Omega_x$.

2. First, $\arg\max\ g_j \notin \Omega_x$ because $\Omega_I \neq \emptyset$, and second, $\arg\max\ g_j \notin \Omega_0$ because, if $j_0 = \arg\max\ g_j \in \Omega_0$, then $g_{j0} = \beta \mathcal{G}_{j0} \leq \beta g_{j0} < g_{j0}$ which is a contradiction. Hence, $\arg\max\ g_j \in \Omega_I$. □

Corollary 6. *The expected unit values of capital $\boldsymbol{\mathcal{G}}^k$ ($k \in \{x, I, 0\}$) satisfy*

$$\boldsymbol{\mathcal{G}}^x = \left(\mathbf{M}_x^x + \beta \mathbf{M}_x^0 \mathbf{C}\right) \mathbf{1}_x (1-\mu) + \left(\mathbf{M}_x^I + \beta \mathbf{M}_x^0 \mathbf{D}\right) \beta \mathbf{M}_I^e \mathbf{f}, \tag{58a}$$

$$\boldsymbol{\mathcal{G}}^I = \left(\mathbf{M}_I^x + \beta \mathbf{M}_I^0 \mathbf{C}\right) \mathbf{1}_x (1-\mu) + \left(\mathbf{M}_I^I + \beta \mathbf{M}_I^0 \mathbf{D}\right) \beta \mathbf{M}_I^e \mathbf{f}, \tag{58b}$$

$$\boldsymbol{\mathcal{G}}^0 = \mathbf{C} \quad \mathbf{1}_x (1-\mu) + \mathbf{D} \quad \beta \mathbf{M}_I^e \mathbf{f}. \tag{58c}$$

Proof. Combine Lemma 5 and (32). □

A.4 Proofs of statements in §7

Proposition 1. *$\{f_j, g_j : j \in \Omega\}$ is an optimal solution of linear program (37) if and only if it is a solution of (20). This solution is unique, hence it is invariant with respect to the particular positive values of $\{a_j, b_j\}$.*

Proof. The equivalence of (20) and (37) is similar to the equivalence between the functional equation for infinite-horizon discounted Markov decision processes and the dual linear program for that model. Therefore, paraphrasing the proof of Theorem 4-11 in Heyman and Sobel (1984) establishes the result.

The uniqueness of the solution (hence, invariance with respect to positive $\{a_j, b_j\}$) follows from $0 \leq \beta < 1$ and the contraction mapping approach to discounted dynamic programming (Denardo 1967). □

Corollary 7. *An optimal extreme point solution to* (37) *yields the following optimal stationary policy in the dynamic program* (11)*. For all $j \in \Omega$,*

$$j \in \begin{cases} \Omega_+ & \textit{if } (37c) \textit{ is binding and } (37b) \textit{ has positive slack.} \\ \Omega_- & \textit{otherwise.} \end{cases}$$

$$j \in \begin{cases} \Omega_0 & \textit{if } (37f) \textit{ is binding, and } (37d) \textit{ and } (37e) \textit{ have positive slack.} \\ \Omega_I & \textit{if } (37e) \textit{ is binding and } (37d) \textit{ has positive slack.} \\ \Omega_x & \textit{otherwise.} \end{cases}$$

Proof. The corollary follows from the discussion in the main body of the paper immediately prior to the statement of the corollary. □

A.5 Proofs of statements in §8

A.5.1 Proofs of statements in §8.1

Lemma 6. *For each $j \in \Omega$, the functions $f_j(\cdot)$ and $g_j(\cdot)$ are convex, continuous, and piece-wise linear on* $[0, 1]$*. The breakpoints in $f_j(\cdot)$ and $g_j(\cdot)$ correspond to a change in the optimal basis in* (37)*. The functions $\mathcal{F}_j(\cdot)$, $\mathcal{F}^e_j(\cdot)$, and $\mathcal{G}_j(\cdot)$ are convex and continuous on* $[0, 1]$*, and they are piece-wise linear if Ω is a finite set.*

Proof. Nonnegative linear combinations of functions that are convex and continuous inherit those properties, and nonnegative linear combinations of finitely many functions that are piece-wise linear inherit that property. Hence, (27a,b) implies that $\mathcal{F}_j(\cdot)$, $\mathcal{F}^e_j(\cdot)$, and $\mathcal{G}_j(\cdot)$ are convex and continuous and, if Ω is a finite set, piece-wise linear, *if* $\{f_j(\cdot), g_j(\cdot)\}$ have these properties. The remainder of the proof shows that this is indeed true.

The proof of convexity and continuity of $\{f_j(\cdot), g_j(\cdot)\}$ is inductive and uses the recursion (13) in which we write $f_{j,n}(\mu)$ and $g_{j,n}(\mu)$ for $f_n(j)$ and $g_n(j)$, respectively. Initiate the proof with $f_{j,0}(\cdot) = g_{j,0}(\cdot) \equiv 0$ for all $j \in \Omega$ and $\mu \in [0,1]$, both of which are continuous and convex as functions of μ. Make the inductive assumption that $f_{j,n-1}(\cdot)$ and $g_{j,n-1}(\cdot)$ are convex and continuous. This implies that $\mathcal{F}_{j,n-1}(\cdot), \mathcal{F}^e_{j,n-1}(\cdot)$, and $\mathcal{G}_{j,n-1}(\cdot)$ are convex and continuous too. Then by (13), for each $j \in \Omega$, $f_{j,n}(\mu)$ and $g_{j,n}(\mu)$ are maxima of two and three continuous convex functions, respectively. So each of them is continuous and convex, completing the induction.

Recall from the proof of Theorem 1 that $f_{j,n}(\mu) \leq f_{j,n+1}(\mu)$ and $g_{j,n}(j) \leq g_{j,n+1}(\mu)$. Thus, $\mathcal{F}_{j,n}(\mu) \leq \mathcal{F}_{j,n+1}(\mu)$, $\mathcal{G}_{j,n}(\mu) \leq \mathcal{G}_{j,n+1}(\mu)$, and $\mathcal{F}^e_{j,n}(\mu) \leq \mathcal{F}^e_{j,n+1}(\mu)$. Letting $n \to \infty$, the Boundedness Assumption implies that $f_j(\cdot)$ and $g_j(\cdot)$ inherit convexity and continuity.

The piece-wise linearity of $f_j(\cdot)$ and $g_j(\cdot)$ and the correspondence between their breakpoints and the change in the optimal basis is a consequence of linear program (37), Proposition 1, and post-optimality sensitivity analysis of linear programs (Dantzig 1963, chapter 12). □

Lemma 7. *On each subinterval of* $[0,1]$ *where the optimal basis of the linear program* (37) *remains unchanged as* μ *varies, for all* $j \in \Omega$ *the functions* $f_j(\mu)$ *and* $\mathcal{F}_j(\cdot)$ *increase.*

Proof. The proof relies on the results in §6. Since the optimal basis remains fixed on the subinterval, it follows from Corollary 7 that the partitions $\{\Omega_+(\cdot), \Omega_-(\cdot)\}$ and $\{\Omega_x(\cdot), \Omega_I(\cdot), \Omega_0(\cdot)\}$ are similarly constant. Therefore, the matrices $\mathbf{A}$ and $\mathbf{B}$ introduced in Lemma 4, and $\mathbf{M}^e_I$, $\mathbf{C}$ and $\mathbf{D}$ used in Lemma 5, are invariant and are thus treated as constants in the proof.

First, note that by Lemmas 4 and 7 and Corollary 3, if each component of the vector function $\mathbf{f}^+(\cdot)$ is increasing on the interval, then each component of the vector functions $\mathbf{f}^-(\cdot)$ and $\boldsymbol{\mathcal{F}}(\cdot)$ is increasing, thus establishing the result. Hence, proving the monotonicity of $\mathbf{f}^+(\cdot)$ is the focus of the remainder of the proof.

Lemma 4 states that, given the partitions, $\mathbf{f}^-(\mu)$ depends linearly on $\mathbf{f}^+(\mu)$. Hence, one can construct a matrix $\mathbf{E}$ that gives

$$\mathbf{f}(\mu) = \mathbf{E}\mathbf{f}^+(\mu). \tag{59}$$

Combining Lemma 4 and Corollary 6 and using (59),

$$\left[\mathbf{I}_+ - \beta(\theta-\lambda)\mathbf{B} - \beta^2 \text{diag}(\mathbf{p}^+)\tilde{\mathbf{D}}\mathbf{M}^e_I\mathbf{E}\right]\mathbf{f}^+(\mu) = \text{diag}(\mathbf{p}^+)\left[\mu(\mathbf{1}_+ - \beta\tilde{\mathbf{C}}\mathbf{1}_x) + \beta\tilde{\mathbf{C}}\mathbf{1}_x\right]. \tag{60}$$

where $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{D}}$ satisfy

$$\boldsymbol{\mathcal{G}}^{+}(\mu) = \tilde{\mathbf{C}}\mathbf{1}_x(1-\mu) + \tilde{\mathbf{D}}\beta\mathbf{M}_I^e\mathbf{f}(\mu). \tag{61}$$

By Corollary 6, for all $j \in \Omega$, $\mathcal{G}_j(\mu)$ is a linear combination of $\mathbf{1}_x(1-\mu)$ and $\beta\mathbf{M}_I^e\mathbf{f}(\mu)$. Therefore, $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{D}}$ exist and can be constructed as follows.

Matrix $\tilde{\mathbf{C}}$ consists of selected rows from $(\mathbf{M}_x^x + \beta\mathbf{M}_x^0\mathbf{C})$, $(\mathbf{M}_I^x + \beta\mathbf{M}_I^0\mathbf{C})$, and $\mathbf{C}$ with states $j \in \Omega_+ \bigcap \Omega_x$, $j \in \Omega_+ \bigcap \Omega_I$, and $j \in \Omega_+ \bigcap \Omega_0$, respectively. Let N_+ and N_x denote the number of states in Ω_+ and Ω_x, respectively, so matrix $\tilde{\mathbf{C}}$ is $N_+ \times N_x$.

Similarly, $\tilde{\mathbf{D}}$ is a collection of rows in $(\mathbf{M}_I^x + \beta\mathbf{M}_x^0\mathbf{D})$, $(\mathbf{M}_I^I + \beta\mathbf{M}_I^0\mathbf{D})$, and $\mathbf{D}$ with states $j \in \Omega_+ \bigcap \Omega_x$, $j \in \Omega_+ \bigcap \Omega_I$, and $j \in \Omega_+ \bigcap \Omega_0$, respectively. Let N_I denote the number of states in Ω_I. Then matrix $\tilde{\mathbf{D}}$ is $N_+ \times N_I$. Since $\mathbf{M}_I^e\mathbf{f}(\mu)$ is $N_I \times 1$, the second term in (61) is $N_+ \times 1$.

For any two vectors $\mathbf{a} = (a_k)$ and $\mathbf{b} = (b_k)$ with the same number of components, we write $\mathbf{a} > \mathbf{b}$ if $a_k > b_k$ for every k. First, note that the right side of (60)

$$\mathbf{1}_+ - \beta\tilde{\mathbf{C}}\mathbf{1}_x > 0,$$

because each component of the column vector $\tilde{\mathbf{C}}\mathbf{1}_x$ is a row sum of matrix $\tilde{\mathbf{C}}$. By definition, each row of $\tilde{\mathbf{C}}$ consists of the *discounted* probabilities that each state in Ω_+ will transit to Ω_x in 0, 1, ... periods. Therefore, the row sums of $\tilde{\mathbf{C}}$ are less than 1, and the right side of (60) increases linearly with μ (recall that $\mathbf{p}^+ > 0$ by Corollary 2).

On the left side of (60), $\mathbf{B}$ consists of the *discounted* probabilities that each state in Ω_+ transits to Ω_+ after staying in Ω_- for 0, 1, ... periods. Matrix $\tilde{\mathbf{D}}$ consists of the *discounted* probabilities that each state in Ω_+ transits to Ω_I in 0, 1, ... periods. Matrix $\mathbf{M}_I^e$ consists of the yield-adjusted one-step transition probabilities that each state in Ω_I transits to any state in Ω. Matrix $\mathbf{E}$ consists of the *discounted* probabilities that each state in Ω transits to Ω_+ in 0, 1, ... periods.

Therefore, when the vectors of market prices $\mathbf{p}^+$ and investment yields $\mathbf{e}$ are low (recall that $\mathbf{e}$ is nonnegative by construction), the row sums of the square matrix

$$\mathbf{I}_+ - \beta(\theta-\lambda)\mathbf{B} - \beta^2\text{diag}(\mathbf{p}^+)\tilde{\mathbf{D}}\mathbf{M}_I^e\mathbf{E} \tag{62}$$

are nonnegative and strictly less than one and thus the matrix is nonsingular.

The Boundedness Assumption implies that $\mathbf{p}^+$ and $\mathbf{e}$ are not too high, i.e., (62) is nonsingular.

Further, because all the elements in both $\beta(\theta-\lambda)\mathbf{B}$ and $\beta^2\text{diag}(\mathbf{p}^+)\tilde{\mathbf{D}}\mathbf{M}_I^e\mathbf{E}$ are nonnegative, the inverse of (62) is nonnegative. Therefore,

$$\mathbf{f}^+(\mu) = \left[\mathbf{I}_+ - \beta(\theta-\lambda)\mathbf{B} - \beta^2\text{diag}(\mathbf{p}^+)\tilde{\mathbf{D}}\mathbf{M}_I^e\mathbf{E}\right]^{-1}\text{diag}(\mathbf{p}^+)\left[\mu(\mathbf{1}_+ - \beta\tilde{\mathbf{C}}\mathbf{1}_x) + \beta\tilde{\mathbf{C}}\mathbf{1}_x\right] \tag{63}$$

which increases linearly with μ. □

A.5.2 Proof of statement in §8.2

Proposition 2. *For all $(K, W, j) \in \mathcal{S}$, there is a threshold $\mu^x(j)$ such that $X(\mu; K, W, j) = W$ if $\mu \leq \mu^x(j)$, and $X(\mu; K, W, j) = 0$ if $\mu > \mu^x(j)$.*

Proof. The proof proceeds in two steps. Step 1 proves the existence of a threshold as described in the proposition, and step 2 proves that such a threshold, which is in general a function of all state variables (K, W, j), depends only on j.

Proof of existence. By Theorem 1, for given $(K, W, j) \in \mathcal{S}$, $X(\mu; K, W, j) = W$ iff $j \in \Omega_x(\mu)$, or equivalently, iff

$$\beta\mathcal{F}_j^e(\mu) \leq 1-\mu, \text{ and} \tag{64}$$

$$\beta\mathcal{G}_j(\mu) \leq 1-\mu. \tag{65}$$

Therefore, proving the existence of a threshold as described in the proposition reduces to proving the following result: if (64) and (65) hold at an arbitrary $\mu_0 \in [0, 1]$, then they are satisfied at all $\mu \in [0, \mu_0)$. This task is fulfilled by examining these two conditions in turn.

Start by considering (64). Because $\mathcal{F}_j^e(\cdot)$ is nondecreasing by Theorem 2,

$$\beta\mathcal{F}_j^e(\mu) \leq \beta\mathcal{F}_j^e(\mu_0) \leq 1-\mu_0 < 1-\mu, \quad \forall\mu \in [0, \mu_0). \tag{66}$$

Therefore, if (64) holds at $\mu_0 \in [0, 1]$, it holds for the entire interval $[0, \mu_0)$.

Next consider (65). Recall that by Corollary 6, $\boldsymbol{\mathcal{G}}(\mu)$ is a linear combination of $1-\mu$ and $\mathbf{f}(\mu)$.

Therefore, there exist vectors $\tilde{\mathbf{c}}(\mu)$ and $\tilde{\mathbf{d}}(\mu)$ such that

$$\mathcal{G}_j(\mu) = (1-\mu)\tilde{\mathbf{c}}^T(\mu)\mathbf{1}_x(\mu) + \beta\tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\mathbf{f}(\mu), \tag{67a}$$

$$= (1-\mu)\tilde{\mathbf{c}}^T(\mu)\mathbf{1}_x(\mu) + \beta\tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\mathbf{E}(\mu)\mathbf{f}^+(\mu), \tag{67b}$$

where the second line uses the matrix $\mathbf{E}(\mu)$ defined in (59), and $\tilde{\mathbf{c}}^T(\mu)$ is a row of $\mathbf{M}_x^x(\mu) + \beta\mathbf{M}_x^0(\mu)\mathbf{C}(\mu)$ in (35a) if $j \in \Omega_x(\mu)$, a row of $\mathbf{M}_I^x(\mu) + \beta\mathbf{M}_I^0(\mu)\mathbf{C}(\mu)$ in (35b) if $j \in \Omega_I(\mu)$, and a row of $\mathbf{C}(\mu)$ in (35c) otherwise. Similarly, $\tilde{\mathbf{d}}^T(\mu)$ is a row of $\mathbf{M}_x^I(\mu) + \beta\mathbf{M}_x^0(\mu)\mathbf{D}(\mu)$ if $j \in \Omega_x(\mu)$, a row of $\mathbf{M}_I^I(\mu) + \beta\mathbf{M}_I^0(\mu)\mathbf{D}(\mu)$ if $j \in \Omega_I(\mu)$, and a row of $\mathbf{D}(\mu)$ otherwise. Note that $\tilde{\mathbf{c}}^T(\mu)$, $\mathbf{1}_x(\mu)$, $\tilde{\mathbf{d}}^T(\mu)$, $\mathbf{M}_I^e(\mu)$, and $\mathbf{E}(\mu)$ depend μ because the partitions may change as μ varies on $[0, \mu_0)$.

Combining (65) and (67b) yields

$$\left[1 - \beta\tilde{\mathbf{c}}^T(\mu)\mathbf{1}_x(\mu)\right](1-\mu) \geq \beta^2\tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\mathbf{E}(\mu)\mathbf{f}^+(\mu). \tag{68}$$

The following shows that if (68) holds at μ_0, then it is true for all $\mu \in [0, \mu_0)$ by discussing two scenarios: (i) when the partitions do not change on $[0, \mu_0]$, and (ii) when the partitions change on $[0, \mu_0]$.

Scenario (i): When the partitions do not change on $[0, \mu_0]$, $\tilde{\mathbf{c}}^T(\mu)$, $\mathbf{1}_x(\mu)$, $\tilde{\mathbf{d}}^T(\mu)$, $\mathbf{M}_I^e(\mu)$, and $\mathbf{D}(\mu)$ are fixed. By construction, $\tilde{\mathbf{c}}^T(\mu)\mathbf{1}_x(\mu) \leq 1$ and $\tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\mathbf{E}(\mu) \geq 0$ for all $\mu \in [0,\ 1]$. Therefore,

$$\left[1 - \beta\tilde{\mathbf{c}}^T\mathbf{1}_x\right](1-\mu) > \left[1 - \beta\tilde{\mathbf{c}}^T\mathbf{1}_x\right](1-\mu_0) \tag{69a}$$

$$\geq \beta^2\tilde{\mathbf{d}}^T\mathbf{M}_I^e\mathbf{E}\mathbf{f}^+(\mu_0) \tag{69b}$$

$$> \beta^2\tilde{\mathbf{d}}^T\mathbf{M}_I^e\mathbf{E}\mathbf{f}^+(\mu), \tag{69c}$$

where the arguments of $\tilde{\mathbf{c}}^T(\mu)$, $\mathbf{1}_x(\mu)$, $\tilde{\mathbf{d}}^T(\mu)$, $\mathbf{M}_I^e(\mu)$, and $\mathbf{D}(\mu)$ are ignored for ease of exposition, and the last line follows from Theorem 2.

Scenario (ii): Let $\mu_1 \in (0, \mu_0)$ be the value at which the first change occurs when μ decreases from μ_0. Then μ_1 is an indifference point where at least two different bases are optimal in the linear program (37). Therefore, μ_1 is the lowest value in the interval that consists of all μ's whose corresponding optimal basis is the same as that of μ_0. This means that the reasoning in Scenario (i) applies and thus (69) holds on $\mu \in [\mu_1, \mu_0)$.

On the other hand, μ_1 is also the highest value in the interval that consists of all μ's that are associated with the new optimal basis, and repeating the reasoning above yields that (69) holds on the entire interval with the new optimal basis. This argument can be iterated many times until, at the final iteration, $\mu_1 = 0$.

Therefore, if (68), or equivalently (65), holds at μ_0, then it is true for all $\mu \in [0, \mu_0)$. Combining this result with the earlier analysis on (64) establishes that, given $(K, W, j) \in \mathcal{S}$, there is a threshold $\bar{\mu} \in [0, 1]$ such that $X(\mu; K, W, j) = W$ if $\mu \in [0, \bar{\mu}]$, and $X(\mu; K, W, j) = 0$ if $\mu \in (\bar{\mu}, 1]$.

Proof of dependence. Because neither (64) nor (65) depends on (K, W), it immediately follows that for any state $(K, W, j) \in \mathcal{S}$, this threshold $\bar{\mu}$ only depends on j. This completes the proof. □

B Boundedness assumption

Section B.1 presents a simple example of unboundedness, thus confirming that the Boundedness Assumption is non-trivial. Section B.2 specifies sufficient conditions on a model's parameters to preclude unboundedness.

B.1 Example of unboundedness

This is a simple deterministic example in which unboundedness occurs. Let Ω have a single state at which $p = 2$ and $e = 3$, and let $\theta = 1$, $\lambda = 0$, and $W_1 = K_1 = 1$. The policy that assigns $x_t = i_t = 0.5W_t$ and $q_t = K_t$ in every period t, causes $K_2 = 2.5$ and $W_2 = 1.5$, yields profits 2 and 5 in periods 1 and 2, and dividends 0.5 and 0.75 in periods 1 and 2. Thus,

$$\beta[\mu q_2 p + (1 - \mu)x_2] - [\mu q_1 p + (1 - \mu)x_1] = \beta(0.25 + 2.75\mu).$$

Every period, there is an analogous increase in the *discounted* contribution to the maximization criterion. Therefore, if $\mu > (4 - \beta)/(4\beta)$, then $\beta^{t-1}[\mu q_t p + (1 - \mu)x_t]$ diverges to ∞ as $t \to \infty$, so the value function of this policy is unboundedly large.

B.2 Sufficient condition for boundedness

This subsection presents sufficient conditions for the value function to be bounded. Equation (60) (derived in the proof of Lemma 7) is

$$\left[\mathbf{I}_+ - \beta(\theta-\lambda)\mathbf{B} - \beta^2 \text{diag}(\mathbf{p}^+)\tilde{\mathbf{D}}\mathbf{M}_I^e\mathbf{E}\right]\mathbf{f}^+ = \text{diag}(\mathbf{p}^+)\left[\mu(\mathbf{1}_+ - \beta\tilde{\mathbf{C}}\mathbf{1}_x) + \beta\tilde{\mathbf{C}}\mathbf{1}_x\right]$$

which embeds sufficient conditions on the MDP's parameters for the value function to be bounded when the exogenous process $\{s_t\}$ is a Markov chain. Given a pair of partitions, $\{\Omega_+, \Omega_-\}$ and $\{\Omega_x, \Omega_I, \Omega_0\}$, if the matrix

$$I_+ - \beta(\theta-\lambda)\mathbf{B} - \beta^2 \text{diag}(\mathbf{p}^+)\tilde{\mathbf{D}}\mathbf{M}_I^e\mathbf{E} \tag{70}$$

is non-singular, the vector $\mathbf{f}^+$ has finite components. Then it follows from Lemmas 4 and 5 that vectors $\mathbf{f}$ and $\mathbf{g}$ have finite components. Consequently, the value function $V(K,W,j) = f_j K + g_j W$ is finite for all $j \in \Omega$.

If the model parameters cause (70) to be singular, however, the elements of $\mathbf{f}^+$ become infinite and so do those of vectors $\mathbf{f}$ and $\mathbf{g}$, which then leads to an unbounded MDP. A possible reason for (70) to be singular is that either $\mathbf{p}$ or $\mathbf{e}$ has a component that is too large. Intuitively, sufficiently high investment yields or net prices cause the firm to keep producing, installing capacity, and thus generating profits, which eventually cause the model to explode.

Because matrix (70) is partition pair-specific, a sufficient condition for the MDP to have a bounded value function is that the model parameters cause (70) to be non-singular for *all possible partition pairs*. When Ω is finite with ω elements, the total number of partition pairs is 6^ω.

C Computational improvement calculations in §7.2

After discretizing the state space and truncating it so that it has finitely many states, it becomes $\mathcal{S} = \{(K,W,j)\} = \{0,1,2,\ldots,\bar{K}\} \times \{0,1,\ldots,\bar{W}\} \times \{1,2,\ldots,\omega\}$ which has $(\bar{K}+1)(\bar{W}+1)\omega$ elements. The feasible actions in state (K,W,j) is the set of discrete values of (q,x,i) satisfying $0 \le q \le K,\ x+i \le W,\ 0 \le x$, and $0 \le i$. In state (K,W,j) there are $K+1$ feasible values of q, approximately $(W+1)^2/2$ of (x,i), and, therefore, approximately $(K+1)(W+1)^2/2$ feasible actions, namely (q, x, i). The cumulative number of feasible actions in all states is approximately

$$\begin{aligned}
& \sum_{K=0}^{\bar{K}} \sum_{W=0}^{\bar{W}} \sum_{z=1}^{\omega} (K+1)(W+1)^2/2 \\
= \;& \omega \sum_{K=0}^{\bar{K}} (K+1) \sum_{W=0}^{\bar{W}} (W+1)^2/2 \\
= \;& \omega(\bar{K}+1)(\bar{K}+2)[2(\bar{W}+1)^3 + 3(\bar{W}+1)^2 + \bar{W} + 1]/24 \\
= \;& \omega(\bar{K}+1)(\bar{K}+2)(\bar{W}+1)(2\bar{W}^2 + 7\bar{W} + 6)/24
\end{aligned}$$

which uses $1 + 2 + \cdots + n = n(n+1)/2$ and $1^2 + 2^2 + \cdots + n^2 = n^3/3 + n^2/2 + n/6$. The product of the number of states and the cumulative number of actions is approximately

$$\begin{aligned}
& (\bar{K}+1)(\bar{W}+1)\omega \times \omega(\bar{K}+1)(\bar{K}+2)(\bar{W}+1)[2(\bar{W}+1)^2 + 3\bar{W} + 4]/24 \\
= \;& (\bar{K}+1)^2(\bar{W}+1)^2\omega^2(\bar{K}+2)(2\bar{W}^2 + 7\bar{W} + 6)/24.
\end{aligned}$$

D Non-monotonicity of investment and liquidity

The following analysis shows that investment, hence liquidity, is not necessarily monotone when μ varies between 0 and 1. Theorem 1 implies that investment decision $I(\mu; K, W, j) \in \{0,\ W\}$, and the resulting liquidity $W - X(\mu; K, W, j) - I(\mu; K, W, j) \in \{0,\ W\}$, so we show that, as μ traverses $[0, 1]$, the sets $\Omega_I(\cdot)$ and $\Omega_0(\cdot)$ may vary in such a way that the membership of a particular $j \in \Omega$ alternates between the two sets. The endogenous state variables W and K are irrelevant here for the same reasons discussed in Proposition 2.

By Theorem 1, $j \in \Omega_I(\mu)$ if and only if

$$\beta \mathcal{F}_j^e(\mu) \geq 1 - \mu, \tag{71}$$

$$\mathcal{F}_j^e(\mu) \geq \mathcal{G}_j(\mu), \tag{72}$$

and $j \in \Omega_0(\mu)$ if and only if

$$\beta \mathcal{G}_j(\mu) > 1 - \mu, \tag{73}$$

$$\mathcal{G}_j(\mu) > \mathcal{F}_j^e(\mu). \tag{74}$$

The remainder of the analysis shows that,

1. if (71) and (72) hold at an arbitrary $\mu_0 \in [0,1)$, then (71) is satisfied at all $\mu \in (\mu_0, 1]$ but (72) may not hold.

2. if (73) and (74) hold at an arbitrary $\mu_0 \in [0,1)$, then (73) is satisfied at all $\mu \in (\mu_0, 1]$ but (74) may not hold.

This implies that, for a particular $j \in \Omega$, as μ increases from its dividend issuance threshold $\mu^x(j)$, the membership of j may switch between the sets $\Omega_I(\cdot)$ and $\Omega_0(\cdot)$.

First consider (71) and (72), and suppose that both are satisfied at $\mu_0 \in [0,1)$. By Theorem 2 and (27c), when μ increases from μ_0, the left side of (71) is non-decreasing while the right side is strictly decreasing. Therefore, (71) holds for all $\mu > \mu_0$.

Let $\mathbf{M}^e = (m_{jk}e_k,\ j,k \in \Omega)$ be the yield-adjusted transition probability matrix and let $\mathbf{m}_e^T$ be the j-th row of $\mathbf{M}^e$, which is independent of μ. Then

$$\mathcal{F}_j^e(\mu) = \mathbf{m}_e^T \mathbf{f}(\mu). \tag{75}$$

Combining (75) with the expression for $\mathcal{G}_j(\mu)$ in (67a), (72) becomes

$$\left[\mathbf{m}_e^T - \beta \tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\right] \mathbf{f}(\mu) \geq \tilde{\mathbf{c}}^T(\mu)\mathbf{1}_x(\mu)(1-\mu). \tag{76}$$

Consider the case when the partitions are fixed as μ increases from μ_0. Then $\tilde{\mathbf{d}}^T(\mu)$, $\tilde{\mathbf{c}}^T(\mu)$, $\mathbf{M}_I^e(\mu)$ and $\mathbf{1}_x(\mu)$ are constants in (76) but as μ increases, the right side of (76) decreases. The change on the left side, however, is unclear. Although every component of the vector $\mathbf{f}(\mu)$ increases with μ, the constant vector $\left[\mathbf{m}_e^T - \beta \tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\right]$ may contain both positive and negative components. Because the unit values of capacity at different states $f_j(\mu)$ $(j \in \Omega)$ may have different rates of increase in μ, the left side of (76) could either increase or decrease. Hence, (72) may be violated as μ increases from μ_0 and thus state j switches from set $\Omega_I(\cdot)$ to $\Omega_0(\cdot)$, even when no other states switch sets in the process.

Next consider (73) and (74) and suppose that they are both satisfied at $\mu_0 \in [0,1)$. Using (67a), (73) reduces to

$$\beta^2 \tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\mathbf{f}(\mu) > \left[1 - \beta \tilde{\mathbf{c}}^T(\mu)\mathbf{1}_x(\mu)\right](1-\mu_0). \tag{77}$$

Because $\tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)$ is nonnegative and $\mathbf{f}(\cdot)$ increases on $[0,1]$, the left side of (77) rises as μ increases from μ_0. The right side of (77), on the other hand, decreases in μ because $\left[1-\beta\tilde{\mathbf{c}}^T(\mu)\mathbf{1}_x(\mu)\right] > 0$ for all $\mu \in [0,1]$. Hence, (77) always holds as μ increases.

Although condition (74) does not have the same property, it is the opposite of (72), so the prior discussion of (72) applies. That is, due to the indeterminate sign of the vector $\left[\mathbf{m}_e^T - \beta\tilde{\mathbf{d}}^T(\mu)\mathbf{M}_I^e(\mu)\right]$ and the heterogeneous increasing rates among the components of $\mathbf{f}(\cdot)$, (74) may be violated as μ increases from μ_0 even when no other state switches sets.