Submitted to *Operations Research*
manuscript OPRE-2016-12-683

Operations Research, Available online 9 October 2018
DOI:10.1287/opre.2018.1752

Sequential capacity expansion options

Alain Bensoussan
International Center for Decision and Risk Analysis,
Jindal School of Management, University of Texas at Dallas,
800 West Campbell Road, Richardson, TX 75080-3021, United States;
College of Science & Engineering, City University of Hong Kong,
Tat Chee Avenue, Kowloon, Hong Kong SAR
alain.bensoussan@utdallas.edu

Benoît Chevalier-Roignant
King's Business School, King's College London,
Strand Campus, Bush House, 30 Aldwych, London WC2B 4BG, United Kingdom
benoit.chevalier-roignant@kcl.ac.uk

This paper considers a firm's capacity expansion decisions under uncertainty. The firm has leeway in timing investments and in choosing how much capacity to install at each investment time. We model this problem as the sequential exercising of compound capacity expansion options with embedded optimal capacity choices. We employ the impulse control methodology and obtain a quasi-variational inequality that involves two state variables: an exogenous, stochastic price process and a controlled capacity process (without a diffusion term). We provide a general verification theorem and identify—and prove the optimality of—a two-dimensional (s, S)-type policy for a specific (admittedly restrictive) choice of the model parameters and of the running profit. The firm delays investment in capacity to ensure that the perpetuity value of newly installed capacity exceeds the total opportunity cost, including the fixed cost component, by a sufficient margin. Our general model for "the option to expand" transcends a single-option exercise and yields predictions of both the optimal investment timing and the optimal scale of production.

Key words: investment under uncertainty; hysteresis; capacity investment; real options

1. Introduction

Capacity investment decisions under uncertainty are among the most challenging problems faced by firms. Investing in many industries (e.g., oil and gas) involves committing substantial financial resources in specific productive assets. Choosing to install a larger capacity lump helps the firm achieve economies of scale by lowering the average acquisition price per unit of capacity; however, it also puts the firm at greater risk of oversupplying the market if the commodity price ends up dropping significantly. Because installed capacity cannot be salvaged (or can only be salvaged at a very low price), firms will likely respond to favorable market developments by building up their capacity in stages. Such phases of "investment bursts" correspond to the *sequential* exercises of capacity expansion options. To investigate this situation, we consider a series of compound perpetual American call options for which the payoff at exercise is itself a real option value that accounts for the next stages of investment. This approach contrasts with the canonical model for the "option to expand" (see, e.g., Trigeorgis 1996) whereby a firm decides, once and for all, when to expand the scale of production by a fixed factor. Our more general setting for the option to expand offers predictions about the optimal timing of investment and about the optimal choice of production scale (or capacity).

When a firm invests, it typically faces affine investment costs, i.e., it incurs not only costs that are proportional (linear) to the capacity installed—in which case these costs correspond to the purchase price per unit of capacity—but also fixed costs—owing, e.g., to land acquisition, project management or rigidity within the firm.[1] The analysis of the interplay between the nature of these costs and optimal dynamic investment policies is at the core of the literature on investment under uncertainty. Real options analysis (see, e.g., McDonald and Siegel 1986, Dixit 1989) investigates situations subject to some form of market uncertainty where a firm trades off a contemplated switch from one "operating mode" to another against the payment of a largely nonrecoverable *fixed* cost. Real options models treat the stochastic process characterizing each mode as exogenous and "do not offer specific predictions about the *level* of investment" (Hubbard 1994, p. 1828, emphasis in original). In our paper, the underlying asset is affected by exogenous (price) developments yet is also actively managed through capacity choices; that asset is a compound option on future expansions of capacity.

The literature on capital accumulation under uncertainty (see, e.g., Abel and Eberly 1994, 1997) provides helpful benchmarks: it investigates the gradual endogenous buildup of capital (e.g., capacity) in an uncertain (continuous-time) macroeconomic environment.[2] While the firm incurs linear costs, fixed costs are assumed to be negligible (or zero). This cost assumption is reasonable if one seeks to explain gradual capital accumulation at the macroeconomic level, but it has less appeal for modeling staged capacity expansions at the microeconomic (firm) level. From the extant neoclassical literature on investment we know that, if the firm can *fully* recoup its investment cost via reselling capital, then it can invest in a myopic manner by adjusting its capital to ensure that the marginal value remains equal to the marginal cost. Otherwise, the sunk nature of investment leads to partial inertia or "hysteresis" in the sense that the firm delays its investment decision (vs. the frictionless case) until the marginal value of capital exceeds the marginal *opportunity* cost. The firm makes a lump-sum investment at the outset if its existing capital stock is not commensurate with the current market conditions substantiated, e.g., via the market-clearing price. Thereafter, the firm invests gradually—not by lump sums but instead by infinitesimal increments. We compare our results arising in a setting with affine investment costs with the literature on capital accumulation. Hysteresis arises also in our case and is exacerbated by the presence of fixed investment costs.

The mathematical expression for these decision problems depends on the cost types considered. On the one hand, scenarios where the firm incurs only fixed costs lead to *optimal stopping* problems which have been explored in the literature on real options. The dynamic programming approach used to address such timing problems requires that one solves what is known as a variational inequality (see Bensoussan and Lions 1982). On the other hand, when the firm incurs linear costs—as in the literature on capital accumulation under uncertainty—one faces a *singular control* problem. Several scholars in the field of mathematics have modeled such investment situations explicitly as singular control problems (see, e.g., Kobila 1993, Øksendal 2000, Riedel and Su 2011, Ferrari 2015). Our main technical contribution is to consider both fixed and linear costs, which leads to an *impulse control* problem. Our approach amounts to solving a quasi-variational inequality (see Bensoussan and Lions 1984); more specifically, we derive a quasi-variational inequality (QVI) involving two state variables and then prove a general (verification) theorem establishing that one specific solution of the QVI is the value function provided it is regular in a sense that we shall stipulate. Finally, we construct a two-dimensional (s, S) policy that turns out to be optimal for a specific (admittedly restrictive) choice of the firm's running profit function and restrictions on the parameter values. This policy specifies two trigger functions: one corresponds to the commodity price above which the firm should raise the capacity stock beyond its current level; the second prescribes the optimal post-investment capacity stock. We show how these trigger functions are related to benchmarks reported in extant literature. We also establish the regularity of the value function in this specific case.

In the presence of affine investment costs, hysteresis takes two forms. First, and similarly to existing models of capital accumulation (with linear costs), the firm stays put when the *marginal* opportunity cost of investing in a capacity unit exceeds the marginal perpetuity value of capital. Second, when the marginal (perpetuity) value exceeds the marginal (opportunity) cost, the firm will delay investment even further to ensure that the aggregate perpetuity value of the extra capacity lump exceeds the investment costs by a sufficient margin. This feature relates to a key result in the real options literature, according to which a firm delays a "switching" (e.g., entry) decision until the gross exogenous project value exceeds the fixed cost by a amount sufficient to account for the opportunity cost of "killing the delay option." Our model generalizes that result to situations in which the underlying asset is itself a compound option with endogenous choices of investment times and capacity scales. There are two components of the current capacity stock's value: the perpetuity value of that stock and a compound option on the values of supramarginal units of capacity.

Previous research has highlighted the need for a unifying approach to investment under uncertainty; we believe our problem involving sequential capacity expansion options is a first step in this direction. Pindyck (1988) asserted that "the assumption that firms can continuously and incrementally add capital, though common in economic models, is extreme. Most investments are lumpy, and sometimes quite so. The opposite extreme assumption is that the firm can build only a single plant, and must decide when to build it" (p. 982). Dixit and Pindyck (1994, p. 386) intuited general properties of the optimal investment strategy as follows: "if there is a stock fixed cost, that is, a one-time cost of making any positive gross addition to the stock of capital, then a finite rate of gross investment cannot be optimal... The optimal policy will allow the capital stock to jump in discrete steps at isolated moments." To the best of our knowledge, no one has since provided a satisfactory (technical) discussion of these effects. Alvarez (2011) confirmed the lack of an integrative perspective: "an interesting extension of our problem [where adjustment costs are linear] would be to introduce fixed sunk costs into the investment and divestment decisions of a firm since that case typically leads to lump-sum investment and divestment policies" (p. 1781).

Determining the (s,S) policy for the two-dimensional impulse control problem addressed in our paper is more intricate than constructing it in the one-dimensional impulse control models discussed to date in the economics and management science literature (in, e.g., Constantinides 1976, Constantinides and Richard 1978, Harrison et al. 1983, Bar-Ilan et al. 2004, Bensoussan 2011). Existing models consider a "storage system"—for example, a cash fund or an inventory stock—that evolves according to a controlled one-dimensional stochastic differential equation: the content fluctuates randomly (as a drifted Brownian motion) in the absence of control but may be adjusted at isolated moments when the firm refills its storage system. In that setup, the optimal (s,S) policy involves *scalar* triggers; s characterizes the level below which the firm refills its system, and $S > s$ is the level the system reaches after refill. In contrast, our setting involves two state equations: a stochastic differential equation drives the changes in the commodity price and a second state equation (without a diffusion term) tracks past capacity investments. Capacity expansions do not feed back into the commodity price dynamics because we assume that the firm cannot wield market power. The optimal (s,S) policy we obtain is characterized by curves, not scalars. This distinction leads to a more involved type of quasi-variational inequality as well as to other technical subtleties, which include solving for an unknown function whose argument is itself an unknown function. The contribution that is probably closest to our work is that of Guo and Wu (2009). Specifically, these authors proved the regularity of viscosity solutions of multi-dimensional QVIs and established a general verification theorem. Moreover, Guo and Wu constructed an explicit solution for the one-dimensional impulse control problem of Constantinides and Richard (1978). Our objective is different; we are interested in constructing an explicit solution (together with boundary curves) of a two-dimensional problem, introducing a generalized (s,S)-type impulse control policy.

The rest of our paper proceeds as follows. Section 2 describes the economic environment and sets the (mathematical) problem of sequential capacity expansion options with optimal capacity choices. In Section 3, we characterize the solution approach via dynamic programming, obtaining a quasi-variational inequality that involves *two* state variables. Section 4 briefly discusses existing models of reversible and irreversible investment without fixed costs; the solutions to their respective dynamic programming equations will serve as bounds for the solution(s) to the QVI of the impulse control problem. In Section 5 we establish a general verification theorem while assuming that one specific solution of the QVI is regular. Section 6 identifies conditions under which the related free boundary problem admits a regular solution. In Section 7, we construct a generalized (s, S) investment policy, which we prove is optimal for a specific form of the firm's running profit function and under (strong) restrictions on model parameter values; we also provide an explicit expression for the value function of this (degenerate) two-dimensional stochastic impulse control problem. We conclude in Section 8 with a summary of our main contributions and suggestions for further research to generalize these findings and establish their robustness.

2. Model setup

A firm faces uncertainty driven by a (standard) Brownian motion $(W_t;\, t \geq 0)$, which is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The increasing sequence of sub-σ-algebras $(\mathcal{F}_t;\, t \geq 0)$ generated by the Brownian motion—the *filtration*—captures how uncertainty unfolds over time; we assume that $\mathcal{F}_\infty \subseteq \mathcal{F}$. The firm operates in a commodity market and sells its product at a price that reflects exogenous economic forces. The price process $(Y_t;\, t \geq 0)$ follows the (one-dimensional) geometric Brownian motion

$$\mathrm{d}Y_t = \mu Y_t \,\mathrm{d}t + \sigma Y_t \,\mathrm{d}W_t, \quad Y_0 = y > 0 \text{ almost surely (a.s.)},$$

with (exponential) growth rate $\mu > 0$ and volatility parameter $\sigma > 0$. The firm decides on the times $\{\tau_n\}_n$ at which to invest and also on the lump-sum additions $\{\xi_n\}_n$ to capacity. The increasing sequence of stopping times $\{\tau_n\}_n$ is adapted to the filtration; the random variable ξ_n is $\mathcal{F}_{\tau_n}$-measurable. The firm's investment policy, denoted $\nu = \{\tau_n, \xi_n\}_n$, is termed an impulse control (in stochastic control theory). Past investments drive the firm's current capacity; we use X_t^ν to signify cumulative investments up to time t:

$$X_t^\nu = x + \sum_n \xi_n \mathbb{I}_{\{\tau_n \leq t\}},$$

where $\mathbb{I}$ is the indicator function. The firm's capacity stock would evolve in a deterministic manner if the firm followed a deterministic capacity expansion strategy ν. To simplify the presentation, we abstract from the opportunity to scrap capacity (by assuming $\xi_n > 0$ for all $n \in \mathbb{N}^\star$) and also ignore the depreciation of capacity.

The firm's output, $\delta(x)$, depends on its installed capacity x. Installing extra capacity makes it possible to produce more ($\delta(0)=0$, $\delta'>0$), but the firm faces diseconomies of scale ($\delta'(0)=\infty$, $\delta''<0$, and $\delta'(\infty)=0$). In a commodity market, one can reasonably assume that an individual firm's output, $\delta(x)$, does not materially affect the equilibrium price y; hence we can write the firm's profit as $\pi(y,x):=y\delta(x)$ in state (y,x), where we assume production costs to be negligible (alternatively, the variable y may be interpreted as a contribution margin per unit of output). In technical terms, the above implies that the stochastic differential equation driving the price process is independent of the firm's (impulse) controls. Installing capacity entails affine costs. At the investment time τ_n, the firm incurs a *fixed cost* k and also a *linear cost* c per unit of newly installed capacity. The firm is risk neutral and discounts both future profits and investment costs at a constant rate $r>0$ per unit of time t.[3] The firm maximizes its *payoff*—the present value of future profits net of capacity costs—by selecting an appropriate staged capacity investment policy ν; the *value function* v is given by

$$v(y,x):=\sup_{\nu}\mathbb{E}\Big[\int_0^\infty e^{-rt}Y_t\delta(X_t^\nu)\,\mathrm{d}t-\sum_n e^{-r\tau_n}(k+c\xi_n)\Big]; \tag{1}$$

here $\mathbb{E}$ is the expectation operator conditional on the initial state being (y,x). We confine the search to controls ν that are *admissible* in the sense that

$$\mathbb{E}\Big[\sum_n e^{-r\tau_n}(k+c\xi_n)\Big]<\infty. \tag{2}$$

Because $r>0$, we can assume that $\tau_n\to\infty$ as $n\to\infty$ without loss of generality. This is because if τ_n converged to a $\tau^\star<\infty$ then there would be no impulse subsequent to $t\geq\tau^\star$, which is equivalent to an impulse occurring at infinity. We use $\hat{\nu}$ to denote the *optimal* admissible control, if one exists.

Finally, we assume that $r>\mu$. Otherwise (i.e., if $r\leq\mu$), the firm would achieve an infinite value of $v(y,x)$ in (1) by never expanding capacity:

$$\delta(x)\mathbb{E}\Big[\int_0^\infty e^{-rt}Y_t\,\mathrm{d}t\Big]=y\delta(x)\int_0^\infty e^{-(r-\mu)t}\,\mathrm{d}t=\infty.$$

3. Solution approach via dynamic programming

We now derive an expression which is more amenable to analysis than the value function v in (1). Following the dynamic programming approach, one can simplify a dynamic optimization problem, such as (1), into pointwise maximization problems that must be satisfied almost everywhere (a.e.). In our paper, the firm must continually choose between increasing its capacity (optimally) or staying put. If a firm faces this alternative, then its value—which corresponds to the optimal choice—must be no less than the payoff from either course of action. Suppose the firm immediately raises capacity to an arbitrary stock level $z>x$ and, thereafter, pursues the optimal investment policy $\hat{\nu}$. This strategy generates a net payoff of $v(y,z)-c[z-x]-k$, whereas selecting the level z optimally yields[4]

$$\mathcal{M}v(y,x):=\sup_{z\geq x}\{v(y,z)-c[z-x]\}-k. \tag{3}$$

The value must exceed the payoff achieved by raising capacity; it follows that $v \geq \mathcal{M}v$. Consider now the firm's alternate strategy of staying put for a (short) period of length $\varepsilon > 0$ and then pursuing investment policy $\hat{\nu}$. According to Bellman's (1957) principle of optimality, the value must exceed the payoff obtained in this case:

$$v(y,x) \geq \mathbb{E}\left[\int_0^{\varepsilon} e^{-rt} Y_t \delta(x)\,\mathrm{d}t + e^{-r\varepsilon} v(Y_\varepsilon, x)\right]. \tag{4}$$

In this expression, if $\varepsilon \to 0$ then $\mathcal{L}v(y,x) \geq y\delta(x)$ where we use $\mathcal{L}$ to denote the following differential operator on the set of twice continuously differentiable functions φ:

$$\mathcal{L}\varphi(y) = \lim_{\varepsilon \downarrow 0} \frac{\varphi(y,x) - \mathbb{E}[e^{-r\varepsilon}\varphi(Y_\varepsilon, x)]}{\varepsilon} := r\varphi(y) - y\mu\varphi'(y) - \frac{1}{2} y^2 \sigma^2 \varphi''(y). \tag{5}$$

Because either alternate action is optimal, one weak inequality must be strict and the other an equality. This heuristic argument leads to a complementary slackness condition the value function must satisfy.

In short, the conditions that must be satisfied by the value function v are as follows:

$$\mathcal{L}v(y,x) - y\delta(x) \geq 0, \qquad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+, \tag{6a}$$

$$v(y,x) - \mathcal{M}v(y,x) \geq 0, \qquad \forall (y,x) \in \mathbb{R}_+^2, \tag{6b}$$

$$[\mathcal{L}v(y,x) - y\delta(x)][v(y,x) - \mathcal{M}v(y,x)] = 0, \qquad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+. \tag{6c}$$

In the terminology introduced by Bensoussan and Lions (1984), problem (6a)–(6c) is a quasi-variational inequality. This QVI can equivalently be expressed in an alternate form:

$$\min\{\mathcal{L}v(y,x) - y\delta(x), v(y,x) - \mathcal{M}v(y,x)\} = 0, \quad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+.$$

A QVI is the analogue in settings with impulse controls of the Hamilton–Jacobi–Bellman (HJB) equation obtained in settings with continuous controls. We also set two boundary conditions:

$$v(0,x) = 0, \tag{6d}$$

$$\liminf_{y \to \infty} \{v(y,x) - v_L(y,x)\} \geq 0. \tag{6e}$$

Here $v_L(y,x)$ is the perpetuity value of holding x capacity units when the commodity currently sells at price y; that value is defined as

$$v_L(y,x) := \frac{y\delta(x)}{r - \mu}. \tag{7}$$

Condition (6d) states that capacity becomes worthless when the commodity price falls to $y = 0$. According to (6e), if the commodity price is very high, then the value of capacity is driven by its perpetuity value, v_L, and the expansion option value term becomes negligible.

Because the QVI (6a)-(6e) needs to be verified for almost every $y \in \mathbb{R}_+$, we are looking for a function v that is *regular* in the sense that it belongs to the set of functions $f(y,x) \mapsto \mathbb{R}$ that are

$$f \in C^1(\mathbb{R}_+^2) \text{ everywhere and for which } f_{yy}(\cdot, x) \text{ is locally integrable,} \tag{8}$$

where $C^1(\mathbb{R}_+^2)$ denotes the space of continuously differentiable functions.

The candidate optimal boundary is the set of points (y,x) such that $v(y,x) = \mathcal{M}v(y,x)$. If we can prove optimality, the continuation set or *waiting region* $\mathcal{C}$ is then implicitly defined as the set of state values (y,x) for which the firm is better-off not raising its capacity; that is, $\mathcal{C} = \{(y,x) \in \mathbb{R}_+^2 \mid v(y,x) > \mathcal{M}v(y,x)\}$. The stopping set or *investment region* is then $\mathcal{S} = \mathbb{R}_+^2 \setminus \mathcal{C}$. In dynamic programming, the solution approach typically involves verifying that a candidate solution satisfies the dynamic programming equation (e.g., a VI, a QVI, or an HJB equation). We introduce a monotone increasing function $\bar{y}(\cdot)$ and posit a particular structure for the continuation set $\mathcal{C}$: given a stock of x capacity units, the firm supposedly delays investment if the current commodity price y is (strictly) below the cutoff value $\bar{y}(x)$ but does invest if y exceeds that value. In other words, we postulate the waiting and investment regions are, respectively, $\mathcal{C} = \{(y,x) \in \mathbb{R}_+^2 \mid y < \bar{y}(x)]\}$ and $\mathcal{S} = \{(y,x) \in \mathbb{R}_+^2 \mid y \geq \bar{y}(x)\}$. Given this conjecture, the value of threshold $\bar{y}(x)$ is determined by solving a *free boundary problem* (FBP)—namely,

$$\mathcal{L}v(y,x) = y\delta(x), \qquad \text{a.e. } y < \bar{y}(x), \tag{9a}$$

$$v(y,x) = \mathcal{M}v(y,x), \qquad y \geq \bar{y}(x), \tag{9b}$$

$$v(0,x) = 0, \tag{9c}$$

$$\liminf_{y \to \infty} \{v(y,x) - v_L(y,x)\} \geq 0. \tag{9d}$$

Let function $\bar{x}(\cdot)$ be the inverse of $\bar{y}(\cdot)$. Then $\bar{x}(y)$ defines a capacity level below which the firm increases its capacity, and the continuation and stopping sets are now expressed as $\mathcal{C} = \{(y,x) \in \mathbb{R}_+^2 \mid x > \bar{x}(y)\}$ and $\mathcal{S} = \{(y,x) \in \mathbb{R}_+^2 \mid x \leq \bar{x}(y)\}$, respectively.

Because we are seeking a solution of the QVI (6a)-(6e) that satisfies the regularity (8), we look for a solution of the FBP in (9a)-(9d) that is continuously differentiable everywhere. This desired regularity gives rise to the well-known smooth-fit conditions (comprising value-matching and smooth-pasting) at threshold $\bar{y}(x)$. Note that a continuously differentiable solution of the free boundary problem in (9a)-(9d) does not necessarily solve the quasi-variational inequality in (6a)-(6e). This is because, whereas conditions (6c), (6d) and (6e) are automatically satisfied, a solution v of the FBP in (9a)-(9d) must also satisfy

$$v(y,x) \geq \mathcal{M}v(y,x), \qquad y < \bar{y}(x), \tag{10a}$$

$$\mathcal{L}v(y,x) \geq y\delta(x), \qquad \text{a.e. } y \geq \bar{y}(x) \tag{10b}$$

in order to be a solution of the QVI (6a)-(6e).

Even if one can find a solution to the quasi-variational inequality (6a)-(6e) it must still be proved that this solution does in fact coincide with the value function in (1). The last step is known as the *verification theorem*. For the reader's convenience, Table 1 summarizes our paper's frequently used symbols and notation.

Table 1 Nomenclature.

		Symbol	Interpretation	Equations
Values & solutions	Impulse control	v	Value function (impulse control problem)	(1)
			Solution of the quasi-variational inequality	(6)
			Solution of the free-boundary problem	(9)
		v_L	Lower bound on the solution of the QVI	(7)
		v_U	Upper bound on the solution of the QVI	(12)
		$\underline{v}$	Minimum solution (limit of a specific sequence)	(15)
		$\overline{v}$	Maximum solution (limit of a specific sequence)	n.a.
	Singular control	v_U	Solution of the quasi-variational inequality	(13)
			Solution of the free-boundary problem	(32)
Optimal control	(s,S) policy	$\bar{y}(\cdot)$	Investment threshold (impulse control problem)	(9)
		$\bar{x}(\cdot)$	Inverse of $\bar{y}(\cdot)$ or "s"	n.a.
		$z(\cdot)$	Capacity level reached after installment ("S")	n.a.
		$\hat{\xi}(y)$	Optimal capacity installment in state y	(17a)
	Singular control	$\bar{y}_U(\cdot)$	Investment threshold (singular control problem)	(14b)
		$\bar{x}_U(\cdot)$	Inverse of $\bar{y}_U(\cdot)$	(14c)
	Reversible case	$\bar{y}_L(\cdot)$	Investment threshold (reversible case)	(11a)
		$\bar{x}_L(\cdot)$	Inverse of $\bar{y}_L(\cdot)$	(11b)
Miscellaneous	Operators	$\mathcal{L}$	Infinitesimal operator	(5)
		$\mathcal{M}$	Intervention operator	(3)
		$\mathbb{E}$	Expectation conditional on initial state (x,y)	n.a.
	Parameters	β	Positive root of the "fundamental quadratic"	(14d)
		C	Linear opportunity cost, $C \equiv \beta c/[\beta-1]$	n.a.
		K	Fixed opportunity cost, $K \equiv \beta k/[\beta-1]$	n.a.
	Thresholds	$y^\star, x_1(\cdot)$	n.a.	(24)
	Functions	F, G	n.a.	(21)
		Y	n.a.	(23)
		X	n.a.	(24b)
		ψ, φ, ρ	Changes of variables	(28)

4. Benchmark models

Before analyzing the impulse control problem (1), we briefly discuss two known models of capital accumulation under uncertainty: (fully) reversible investment and irreversible investment with negligible fixed costs ($k=0$). These benchmarks are not well suited for explaining stages of "investment bursts" at the microeconomic firm level, where fixed costs are prevalent. However, the respective explicit solutions can serve as lower (L) and upper (U) bounds for the solution(s) of the impulse control problem's QVI (6a)-(6e).

4.1. Reversible investment

The perpetuity value of installed capacity, v_L, given in (7) is a lower bound for the value function v in (1) because the latter embeds a series of capacity expansion options.[5] Suppose the firm can resell or scrap capacity at a salvage value equal to the acquisition price and assume, moreover, that the firm faces no organizational rigidity and incurs no fixed cost ($k=0$) when either raising or scrapping capacity. Such a frictionless setting describes the case of *fully reversible* investment. Here, a myopic investment policy—with the firm continually adjusting its capacity—is optimal.

For a given stock of x capacity units,

$$\bar{y}_L(x) = \frac{c[r-\mu]}{\delta'(x)} \tag{11a}$$

is the commodity price above which the marginal (perpetuity) value of capacity exceeds the marginal cost c; the interpretation is correct because $y \geq \bar{y}_L(x)$ is equivalent to $\frac{\partial v_L}{\partial x}(y,x) \geq c$. In a frictionless environment, the firm should invest in a marginal capacity unit if the commodity price y exceeds $\bar{y}_L(x)$ [or should divest if $y \leq \bar{y}_L(x)$]. This myopic investment rule is equivalent to the net present value (NPV) investment rule in finance applied to the marginal capacity unit. The function $\bar{y}_L(\cdot)$ is increasing and concave in installed capacity x (because $\delta'' < 0$). The inverse $\bar{x}_L(\cdot)$ of $\bar{y}_L(\cdot)$, which is given by

$$\bar{x}_L(y) = (\delta')^{-1}\left(\frac{c[r-\mu]}{y}\right), \tag{11b}$$

also has an economic interpretation: it is the capacity level below which the marginal (perpetuity) value of capacity exceeds the marginal (linear) investment cost c. If the commodity price y is above $\bar{y}_L(x)$ then the firm raises capacity stock from x to $\bar{x}_L(y)$, at which point the marginal (perpetuity) value of capacity falls to equal the marginal (linear) cost. An alternative but less common description of the investment rule is in terms of capacity stock. Investment occurs when the stochastic commodity price $(Y_t;\, t \geq 0)$ exceeds the fixed, deterministic threshold $\bar{y}_L(x)$ or—which is strictly equivalent—when the deterministic capacity stock x is below the realization of the stochastic optimal capacity stock process $(\bar{x}_L(Y_t);\, t \geq 0)$. The investment and noninvestment regions are then $\mathcal{S}_L = \{(y,x) \in \mathbb{R}_+^2 \mid y \geq \bar{y}_L(x)\}$ and $\mathcal{C}_L = \{(y,x) \in \mathbb{R}_+^2 \mid y < \bar{y}_L(x)\}$, respectively.

4.2. Irreversible investment with $k = 0$

We now consider the case of *full investment irreversibility* under uncertainty—that is, when capacity cannot easily be resold as would be the case, for instance, if the capacity is firm-specific. Since incurring a fixed cost k depresses the value of the capacity expansion options, it follows that setting the fixed cost component to zero ($k=0$) yields an upper bound for the value function v in (1). Then that problem degenerates into a singular control problem. A *singular control* is an adapted increasing (càdlàg) process (ξ_t; $t \geq 0$) that represents the cumulative investments up to time t; it is admissible if $\xi_{0-}=0$ and $\mathbb{E}\big[\int_0^\infty e^{-rt}\,\mathrm{d}\xi_t\big] < \infty$. In this degenerate case, the capacity stock at time t is simply $X_t^\xi = x + \xi_t$. The value function of the singular control problem is[6]

$$v_U(y,x) := \sup_\xi \mathbb{E}\left[\int_0^\infty e^{-rt} Y_t \delta(X_t^\xi)\,\mathrm{d}t - c\int_0^\infty e^{-rt}\,\mathrm{d}\xi_t\right]. \tag{12}$$

The dynamic programming equation for the singular control problem in (12) is a degenerate version of the QVI (6a)-(6e), namely[7]

$$\mathcal{L}v_U(y,x) - y\delta(x) \geq 0, \qquad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+, \tag{13a}$$

$$\frac{\partial v_U}{\partial x}(y,x) \leq c, \qquad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+, \tag{13b}$$

$$[\mathcal{L}v_U(y,x) - y\delta(x)]\left[\frac{\partial v_U}{\partial x}(y,x) - c\right] = 0, \qquad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+, \tag{13c}$$

$$v_U(0,x) = 0, \tag{13d}$$

$$\liminf_{y\to\infty}\{v_U(y,x) - v_L(y,x)\} \geq 0. \tag{13e}$$

A function must satisfy the regularity (8) to solve the QVI (13a)-(13e) (in the classical sense).

As we shall see, it is easier to solve the QVI of the singular control problem in (13a)-(13e) than to solve the QVI of the impulse control problem in (6a)-(6e). Proposition 1 identifies an explicit solution to the QVI (13a)-(13e) together with the free boundary curve $\bar{y}_U(\cdot)$. The sets $\mathcal{C}_U = \{(y,x) \in \mathbb{R}_+^2 \mid y < \bar{y}_U(x)\}$ and $\mathcal{S}_U = \{(y,x) \in \mathbb{R}_+^2 \mid y \geq \bar{y}_U(x)\}$ would correspond, respectively, to the continuation and stopping sets if we prove the optimality of the threshold policy. The proof of Proposition 1 as well as all other proofs are provided in the appendices.

Proposition 1. *The function v_U given by*

$$v_U(y,x) = \begin{cases} v_L(y,x) + A_U(x)y^\beta, & y \leq \bar{y}_U(x), \\ v_L(y,\bar{x}_U(y)) - c[\bar{x}_U(y) - x] + A_U(\bar{x}_U(y))y^\beta, & y \geq \bar{y}_U(x), \end{cases} \tag{14a}$$

is a solution of (13a)-(13e) *that satisfies the regularity* (8)*. The threshold functions $\bar{y}_U(\cdot)$ and $\bar{x}_U(\cdot)$ are, respectively,*

$$\bar{y}_U(x) = \frac{C[r-\mu]}{\delta'(x)}, \tag{14b}$$

$$\bar{x}_U(y) = (\delta')^{-1}\left(\frac{C[r-\mu]}{y}\right), \tag{14c}$$

where $C := \beta c/[\beta - 1]$ and

$$\beta := -\frac{\mu - \sigma^2/2}{\sigma^2} + \sqrt{\left(\frac{\mu - \sigma^2/2}{\sigma^2}\right)^2 + \frac{2r}{\sigma^2}} \in (1, r/\mu). \tag{14d}$$

The value of the capacity expansion options (when $k = 0$) is

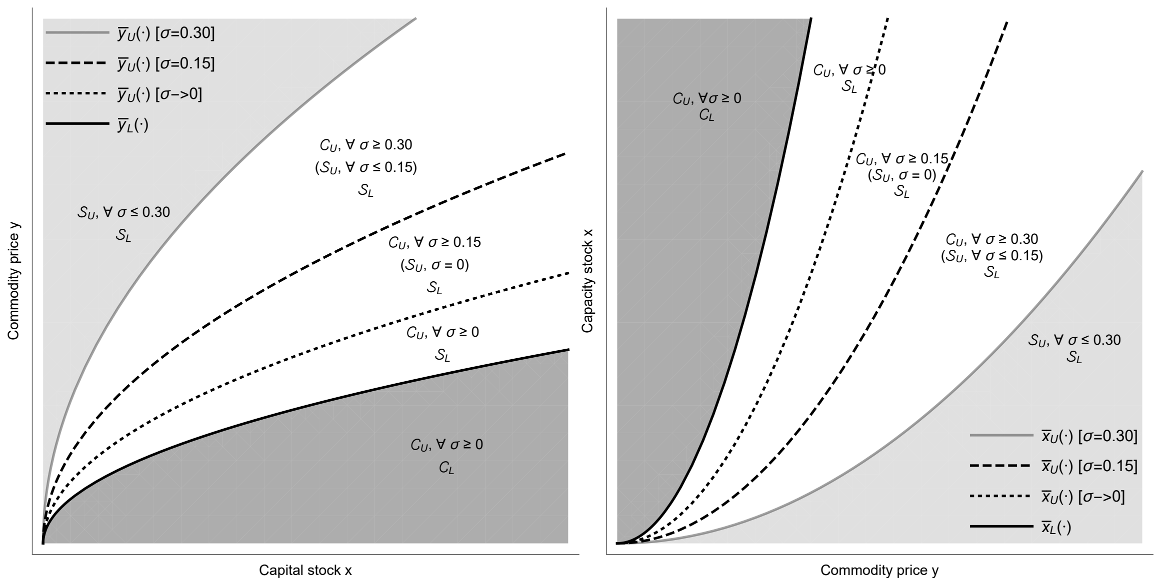
$$\begin{aligned} A_U(x) y^\beta &= \int_x^\infty \left[\frac{\partial v_L}{\partial x}(\bar{y}_U(\xi), \xi) - c\right] \left(\frac{y}{\bar{y}_U(\xi)}\right)^\beta \mathrm{d}\xi, \\ &= \frac{C^{\beta-1}}{\beta} \left(\frac{y}{r-\mu}\right)^\beta \int_x^\infty \delta'(\xi)^\beta \,\mathrm{d}\xi. \end{aligned} \tag{14e}$$

The function v_U exceeds the perpetuity value of capacity, v_L, given in (7)*.*

Because investment is irreversible (ξ is an increasing process), installing capacity is a commitment—in that the firm "kills the option to defer"—and so entails an opportunity cost; that cost is captured not by the linear term cx but rather by Cx, where $C > c$. Under irreversibility, the firm stays put as long as the marginal opportunity cost C exceeds the marginal perpetuity value (because $y \leq \bar{y}_U(x) \iff \frac{\partial v_L}{\partial x}(y, x) \leq C$); but under negligible fixed costs, the firm invests in the marginal capacity unit when the current commodity price is higher ($y \geq \bar{y}_U(x)$). Since the perpetuity value of a marginal capacity unit ($\frac{\partial v_L}{\partial x}$) is monotone increasing in the commodity price y and we have $C > c$, it follows that the price threshold $\bar{y}_U(x)$ in the case of irreversible investment ($k = 0$) exceeds the full-reversibility threshold $\bar{y}_L(x)$ in (11a). The range of state values at which the firm would invest if the investment were perfectly reversible but would not invest if doing so were irreversible (viz., $\mathcal{S}_L \cap \mathcal{C}_U$) is called the *hysteresis* or inertial set. Investing in this set is ill advised when the firm cannot recoup the investment cost ($c\xi$) because doing so entails opportunity costs in excess of the net investment proceeds. It is not necessary to incur a fixed cost ($k > 0$) for hysteresis to arise. For irreversible investment with negligible fixed costs ($k = 0$), the investment (timing) rule resembles the optimal exercise policy in the case of an option to defer (McDonald and Siegel 1986) but here applied to the marginal capacity unit. It is intuitive that the presence of a fixed cost k requires some revision of the investment rule just described; this is because the optimal policy should accommodate the *fixed* (opportunity) cost.

In (14c) we defined $\bar{x}_U(\cdot)$ as the inverse of $\bar{y}_U(\cdot)$. If the firm invests at the outset (i.e., if $y \geq \bar{y}_U(x)$), it will install capacity up to the level $\bar{x}_U(y)$ above which the value of additional units falls below the unit opportunity cost C. Figure 1 compares the two benchmark cases of reversible investment and singular control, distinguishing their respective investment regions ($\mathcal{S}_L$ and $\mathcal{S}_U$) and noninvestment regions ($\mathcal{C}_L$ and $\mathcal{C}_U$). Panel (a) plots the boundary curves $\bar{y}_L(\cdot)$ and $\bar{y}_U(\cdot)$ as functions of the capital stock x; panel (b) shows the boundary curves $\bar{x}_L(\cdot)$ and $\bar{x}_U(\cdot)$ as functions of the commodity price y. (Readers who are acquainted with the literature on real options may find panel (a)'s depiction in terms of "trigger prices" more appealing.) The two graphs are clearly symmetric and equivalent because $\bar{x}_L(\cdot)$ and $\bar{x}_U(\cdot)$ are the respective inverses of $\bar{y}_L(\cdot)$ and $\bar{y}_U(\cdot)$.

Figure 1 **Comparison of investment and noninvestment/waiting regions: reversible investment (L) versus singular control (U).**



(a) Curves $\bar{y}_U(\cdot)$ vs. $\bar{y}_L(\cdot)$ as functions of stock x (b) Curves $\bar{x}_U(\cdot)$ vs. $\bar{x}_L(\cdot)$ as functions of price y

Note. Assuming $\delta(x) = \sqrt{x}$, $r = 0.05$, $\mu = 0.014$, and $c = 10$.

In both panels of this figure, the dark gray region $\mathcal{C}_L$ corresponds to the noninvestment region in the reversible case, while the light gray region $\mathcal{S}_U$ is the investment region in the singular control case (when volatility $\sigma = 0.30$). In fact, the light gray region is a subset of the stopping sets $\mathcal{S}_U$ for any $\sigma < 0.30$ because the threshold $\bar{y}_U(\cdot)$ is monotone increasing in σ: greater uncertainty (in the form of a higher volatility parameter σ) leads to greater caution in terms of investment timing. The intermediate, white region corresponds to the hysteresis region ($\mathcal{C}_U \cap \mathcal{S}_L$) for a price volatility of $\sigma = 0.30$. In this region, the dashed curve represents the thresholds $\bar{y}_U(\cdot)$ and $\bar{x}_U(\cdot)$ for $\sigma = 0.15$. The (white) gap between the light gray boundary and the dashed curves captures the widening of the hysteresis region when price volatility rises (here, from $\sigma = 0.15$ to $\sigma = 0.30$). Irreversibility creates inertia and depresses capacity investment (since $\bar{x}_U(y) \le \bar{x}_L(y)$), an effect that is exacerbated under uncertainty. Uncertainty does not create hysteresis (irreversibility does), but uncertainty does amplify the inertial effect. The curves $\bar{y}_L(\cdot)$ and $\bar{x}_L(\cdot)$—as given in (11a) and (11b)—are not the limits of $\bar{y}_U(\cdot)$ and $\bar{x}_U(\cdot)$ when uncertainty σ vanishes; these limits are plotted as dotted lines in Figure 1.

By construction, v_U—which is given in closed form in (14a)—ignores fixed costs. If the commodity price is low ($y \le \bar{y}_U(x)$) then the firm stays put with x units of installed capacity; should the price rise significantly, it would install additional capacity. The firm's wealth, $v_U(y, x)$, corresponds to the perpetuity value of installed capacity, $v_L(y, x)$, *plus* the term $A_U(x)y^\beta$ (given in (14e)) that corresponds to the real option value of acquiring the supramarginal capacity units; the firm will purchase these units in the future over several stages of *incremental* capacity expansions. The net proceeds obtained when the firm invests at the commodity price $\bar{y}_U(\xi)$ is $\frac{\partial v_L}{\partial x}(\bar{y}_U(\xi),\xi)-c$, an amount that is discounted back to time 0 by using $\mathbb{E}\left[e^{-r\tau_U(y,\xi)}\right]=(y/\bar{y}_U(\xi))^\beta$ with $\tau_U(y,x):=\inf\{t\geq 0 \mid Y_t\geq \bar{y}_U(x)\}$. We consider a stream of such proceeds (hence the improper integral from x to ∞) as the firm will acquire all supramarginal units in stages. If the initial commodity price is higher, viz. $y\geq \bar{y}_U(x)$, then the firm immediately installs $\bar{x}_U(y)-x$ more units of capacity; it thereby resets the net perpetuity value to $v_L(y,\bar{x}_U(y))-c[\bar{x}_U(y)-x]$ and adjusts the value of the capacity expansion options to $A_U(\bar{x}_U(y))y^\beta$.

5. General properties of the quasi-variational inequality

We now consider our more general problem of sequential capacity expansions with fixed costs given in (1). Solving the QVI (6a)-(6e) is more intricate than solving its degenerate version in (13a)-(13e). If we can find solutions to the QVI (6a)-(6e) that are regular in the sense of (8) then we can prove, by way of a verification theorem, that one of them—the "minimum" one—corresponds to the value function in (1).

5.1. Existence of solutions to the quasi-variational inequality

We start by narrowing down, based on the discussion in Section 4, the range of functions viewed as candidate solutions. Capacity expansion options are not worthless, so the value v in (1) must exceed the perpetuity value of capacity, v_L, in (7). Besides, a fixed cost ($k>0$) most likely reduces value, leading to the inequality $v\leq v_U$. Hence solutions for v of the QVI (6a)-(6e) such that $v_L\leq v\leq v_U$ are natural because the value function in (1) should be in this interval.

Our objective in this section is to elaborate on the following statement:

Statement. *The quasi-variational inequality in* (6a)-(6e) *admits solutions satisfying the regularity* (8) *in the interval* $[v_L,v_U]$. *The set of solutions has a minimum and a maximum, which are denoted (respectively)* $\underline{v}$ *and* $\bar{v}$.

We do not provide a full proof of this statement because the regularity (8) cannot be readily established using general arguments. At this stage, regularity is posited; it will be established later when constructing an explicit solution in Section 7.

The following constructive method is a classical approach to obtaining candidate solutions. It amounts to approximating a solution of the QVI as the limit of a (converging) sequence of solutions to recursive variational inequalities.[8] One sequence, starting at v_L, turns out to be increasing and converges to a *minimum solution* $\underline{v}$; another sequence, starting at v_U, is decreasing and converges to a *maximum solution* $\bar{v}$. Specifically, consider the sequence $\{v_n\}_{n\geq 0}$ defined as follows: the first term is $v_0 = v_L$, and all other terms are defined recursively as *regular* solutions of

$$\mathcal{L}v_{n+1}(y,x) - y\delta(x) \geq 0, \qquad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+, \tag{15a}$$

$$v_{n+1}(y,x) - \mathcal{M}v_n(y,x) \geq 0, \qquad \forall (y,x) \in \mathbb{R}_+^2, \tag{15b}$$

$$[\mathcal{L}v_{n+1}(y,x) - y\delta(x)][v_{n+1}(y,x) - \mathcal{M}v_n(y,x)] = 0, \qquad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+, \tag{15c}$$

$$v_{n+1}(0,x) = 0, \qquad \forall x \in \mathbb{R}_+, \tag{15d}$$

$$\liminf_{y\to\infty}[v_{n+1}(y,x) - v_L(y,x)] \geq 0, \qquad \forall x \in \mathbb{R}_+. \tag{15e}$$

The recursive relation in (15a)-(15e) defining v_{n+1} for a given v_n is a variational inequality, not a quasi-variational inequality. A noted difference is that the "obstacle" of this VI, $\mathcal{M}v_n$, involves the given term v_n, while the "implicit obstacle" $\mathcal{M}v$ of the QVI (6a)-(6e) depends on the sought solution v. We establish the following lemma:

Lemma 1. *Assume that each term of the sequence $\{v_n\}_{n\geq 0}$ defined in* (15a)-(15e) *satisfies the regularity* (8). *Then, $\{v_n\}_{n\geq 0}$ is an increasing sequence, satisfying*

$$v_L \leq \cdots \leq v_n \leq v_{n+1} \leq \cdots \leq v_U. \tag{16}$$

It converges pointwise to a function $\underline{v}$, which we call the minimum solution.

We consider another sequence $\{v^n\}_{n\geq 0}$ of regular solutions of recursive variational inequalities—one that starts with $v^0 = v_U$ instead of with v_L. By similar techniques, we can prove that $\{v^n\}_{n\geq 0}$ is a decreasing sequence, with

$$v_U \geq \cdots \geq v^n \geq v^{n+1} \geq \cdots \geq v_L,$$

and that it converges pointwise to a function $\bar{v}$, called here the maximum solution. This constructive procedure helps prove that the set of solutions of the QVI (13a)-(13e) in the interval $[v_L, v_U]$ admits a minimum $\underline{v}$ and a maximum $\bar{v}$, where $\underline{v} \leq \bar{v}$—provided of course that the regularity can be established.

5.2. Verification theorem

Before characterizing the value function in (1) in the set of solutions of the QVI (6a)-(6e), we shall conjecture that the optimal impulse control $\hat{\nu} = \{\hat{\tau}_n, \hat{\xi}_n\}_{n\geq 1}$ can be characterized using a generalized (s,S)-type policy. Here both s and S are defined as functions; this is in contrast to traditional settings, where s and S are scalars. The impulse control must be admissible in that (2) is satisfied. The candidate optimal boundary is the set of points (y,x) such that $v(y,x) = \mathcal{M}v(y,x)$. Consider a monotone increasing function $\bar{y}(\cdot)$, and suppose the continuation and stopping sets can be written as (respectively) $\mathcal{C} = \{(y,x) \in \mathbb{R}_+^2 \mid y < \bar{y}(x)\}$ and $\mathcal{S} = \{(y,x) \in \mathbb{R}_+^2 \mid y \geq \bar{y}(x)\}$. This particular structure implies that, given an initial stock of x installed units, capacity expansion occurs when the commodity price Y_t exceeds the cutoff value $\bar{y}(x)$. The inverse of $\bar{y}(\cdot)$—that is, $\bar{x}(\cdot)$, which is defined on $(\bar{y}_0, \infty)$ with $\bar{y}_0 := \bar{y}(0)$—corresponds to the s in the conjectured (s,S)-type policy; the stock reached after the capacity installment, here denoted $z(\cdot)$, corresponds to the S. Given the economic context, we define the stock level reached after investment at the boundary $\bar{x}(y)$ as $z(y) := \bar{x}(y) + \hat{\xi}(y)$; here $\hat{\xi}(y) > 0$, given by

$$\hat{\xi}(y) = \arg\max_{\xi > 0} \{v(y, \bar{x}(y) + \xi) - c\xi\}, \tag{17a}$$

represents the optimal (lump-sum) capacity installment. We conjecture that the function $z(\cdot)$ is monotone increasing on $(\bar{y}_0, \infty)$: a higher commodity price y incentivizes the firm to install more capacity. By construction,

$$z(y) > \bar{x}(y) \quad \forall y > 0. \tag{17b}$$

Consider now a sequence of stopping times $\{\hat{\tau}_n\}_{n \geq 1}$. The first term is

$$\hat{\tau}_1 = \inf\left\{t \geq 0 \mid Y_t = \max\{y, \bar{y}(x)\}\right\},$$

and subsequent terms are given by the recursive relation

$$\hat{\tau}_{n+1} = \inf\{t > \hat{\tau}_n \mid Y_t = \bar{y}(X_{\hat{\tau}_n})\}, \quad n \geq 1.$$

Because $\{\hat{\tau}_n\}_n$ is strictly increasing, we necessarily have $\hat{\tau}_n \to \infty$ almost surely as $n \to \infty$. The first time the firm expands is $\hat{\tau}_1$; at that time, its capacity jumps from x to $X_{\hat{\tau}_1} = z(\max\{y, \bar{y}(x)\})$ as the firm installs $\hat{\xi}_1 = X_{\hat{\tau}_1} - x$ new units of capacity. If $y > \bar{y}(x)$ at the outset then the firm invests straight away, with $\tau_1 = 0$. However, if $y < \bar{y}(x)$ then the first investment time, $\hat{\tau}_1$, is random, while both $Y_{\hat{\tau}_1} = \max\{y, \bar{y}(x)\}$ and $X_{\hat{\tau}_1}$ are deterministic values. When the firm subsequently expands at time $\hat{\tau}_n$, capacity jumps to $X_{\hat{\tau}_n} = z(Y_{\hat{\tau}_n})$ with the firm making a lump-sum investment of $\hat{\xi}_n = X_{\hat{\tau}_n} - X_{\hat{\tau}_{n-1}}$. If there exists an optimal strategy $\hat{\nu}$, then the level of capacity at time t is given by $\hat{X}_t = X_t^{\hat{\nu}}$. The capacity stock process $(\hat{X}_t;\, t \geq 0)$ is stochastic, depends on past price developments, and increases in stages. We assume that the *transversality condition*,

$$\lim_{T \to \infty} \mathbb{E}[v_U(Y_T, \hat{X}_T) e^{-rT}] = 0, \tag{18}$$

is satisfied.

Proposition 2 provides a verification theorem, which identifies the value function v among the set of solutions of the QVI (6a)-(6e) as the minimum solution $\underline{v}$.

Proposition 2. *Assume that* (i) *there exists a pair of monotone increasing functions* $(\bar{x}(\cdot), z(\cdot))$ *satisfying* (17a)-(17b)*,* (ii) *the corresponding impulse control* $\hat{\nu}$ *is admissible* (*cf.* (2))*,* (iii) *the transversality condition* (18) *is satisfied, and* (iv) *the minimum solution* $\underline{v} \in [v_L, v_U]$ *of the QVI* (6a)-(6e) *satisfies the regularity* (8)*. Then the minimum solution* $\underline{v}$ *coincides with the value function in* (1)*.*

6. Free boundary problem

We wish to construct a generalized (s,S)-type policy and prove that it is optimal. Toward that end, we examine the FBP in (9a)-(9d) to identify the (s,S) curves. If this (s,S)-type policy turns out to be optimal, then $(y,\bar{x}(y))$ corresponds to the boundary of the continuation set; an equivalent expression for this boundary is $(\bar{y}(x),x)$. We determine the (s,S) curves as implicit solutions in Section 6.1 and then as explicit solutions in Section 6.2.

6.1. Implicit expressions for $\bar{x}(\cdot)$ and $z(\cdot)$

Our next proposition establishes general conditions on the boundary curves $\bar{x}(\cdot)$ and $z(\cdot)$ that are both necessary and sufficient for the existence of a continuously differentiable solution of the free boundary problem in (9a)-(9d).

PROPOSITION 3. *Suppose we can find two functions $\bar{x}(\cdot)$ and $z(\cdot)$ that are monotone increasing on $[\bar{y}_0,\infty)$ with $z(\cdot)>\bar{x}(\cdot)$ and that satisfy*

$$v_L(y,z(y)) - v_L(y,\bar{x}(y)) = C[z(y)-\bar{x}(y)] + K \quad \textit{and} \tag{19a}$$

$$\frac{\partial v_L}{\partial x}(y,z(y)) - c = \left[\frac{\partial v_L}{\partial x}(\bar{y}(z(y)),z(y)) - c\right]\left(\frac{y}{\bar{y}(z(y))}\right)^{\beta}, \tag{19b}$$

where $C := \beta c/[\beta-1]$ and $K := \beta k/[\beta-1]$. Then the function v given by

$$v(y,x) = \begin{cases} v_L(y,x) + A(x)y^{\beta}, & y \le \bar{y}(x),\\ v_L(y,z(y)) - c[z(y)-x] - k + A(z(y))y^{\beta}, & y \ge \bar{y}(x), \end{cases} \tag{19c}$$

where

$$A(x)y^{\beta} = \int_x^{\infty}\left[\frac{\partial v_L}{\partial x}(\bar{y}(\xi),\xi) - c\right]\left(\frac{y}{\bar{y}(\xi)}\right)^{\beta} \mathrm{d}\xi, \tag{19d}$$

is a solution of the FBP (9a)-(9d) *that satisfies the regularity* (8).

Our "selection" of the optimal barriers $\bar{x}(\cdot)$ and $z(\cdot)$ can be interpreted as follows. First, note that condition (19a) can be written as

$$\int_{\bar{x}(y)}^{z(y)}\left[\frac{\partial v_L}{\partial x}(y,\xi) - C\right]\mathrm{d}\xi = K. \tag{20}$$

Therefore, if the firm invests at the boundary $(y,\bar{x}(y))$ of the continuation set, then it invests up to the capacity level $z(y)$ at which the aggregate value of the newly installed (inframarginal) capacity units *in excess* of their linear opportunity acquisition costs (on the left-hand side, LHS) is equal to the fixed opportunity cost K (on the right-hand side, RHS). At the boundary of the continuation set, $K>k>0$ entails that the capacity stock will jump from $\bar{x}(y)$ to $z(y)>\bar{x}(y)$. By (20), if the fixed cost is negligible ($K=k=0$) then the boundaries $\bar{x}(\cdot)$ and $z(\cdot)$ coincide.

Condition (19b) states that, if the capacity stock reaches level $z(y)$ and if $k>0$, then the firm is indifferent between immediately investing in a marginal unit when the commodity price is y—and thereby obtaining the net proceeds on the LHS of (19b)—or delaying investment in the marginal unit until the commodity price Y_t reaches the threshold $\bar{y}(z(y)) > \bar{y}(\bar{x}(y)) = y$. So when the firm invests when the commodity price reaches $\bar{y}(z(y))$, the net incremental future value of the marginal unit is $\frac{\partial v_L}{\partial x}(\bar{y}(z(y)), z(y)) - c$ and its present value is obtained by multiplying that amount by the discount factor $[y/\bar{y}(z(y))]^{\beta}$.

The payoff expression in (19c) has the following interpretation. If installing capacity immediately is not optimal ($y \le \bar{y}(x)$), then wealth is derived from (a) the perpetuity value of existing capacity, $v_L(y,x)$, and (b) the real option value of future capacity expansions captured by $A(x)y^{\beta}$ in (19d). Essentially $A(x)y^{\beta}$ corresponds to the net discounted (perpetuity) value of acquiring supramarginal capacity units (above x) whenever the commodity price $(Y_t; t \ge 0)$ reaches the boundary $\bar{y}(\cdot)$. The inertial effect on the real option value of future capacity expansions is implicitly accounted for via the conditions (19a)-(19b) that determine the threshold $\bar{y}(\cdot)$ (and its inverse $\bar{x}(\cdot)$). It is also interesting that, whereas the term $A(x)y^{\beta}$ depends explicitly on the variable cost c, its dependence on the fixed cost k is not explicit, but rather implicit from conditions (19a)-(19b). If the firm installs capacity at the outset ($y \ge \bar{y}(x)$), then its total capacity after that investment is $z(y)$ units; hence the investment "resets" gross perpetuity value to $v_L(y, z(y))$ at a cost of $c[z(y) - x] + k$, and the real option value of subsequent capacity expansions is adjusted to $A(z(y))y^{\beta}$.

6.2. Explicit expressions for $\bar{x}(\cdot)$ and $z(\cdot)$

The functions $\bar{x}(\cdot)$ and $z(\cdot)$ are defined implicitly by the system (19a)-(19b). We now explain how to express these functions explicitly. For this purpose, we introduce the functions F and G, defined as

$$F(y,x) := (\beta - 1)v_L(y,x) - \beta c x, \tag{21a}$$

$$G(y,x) := -\left[\frac{\partial v_L}{\partial x}(y,x) - c\right] y^{-\beta}. \tag{21b}$$

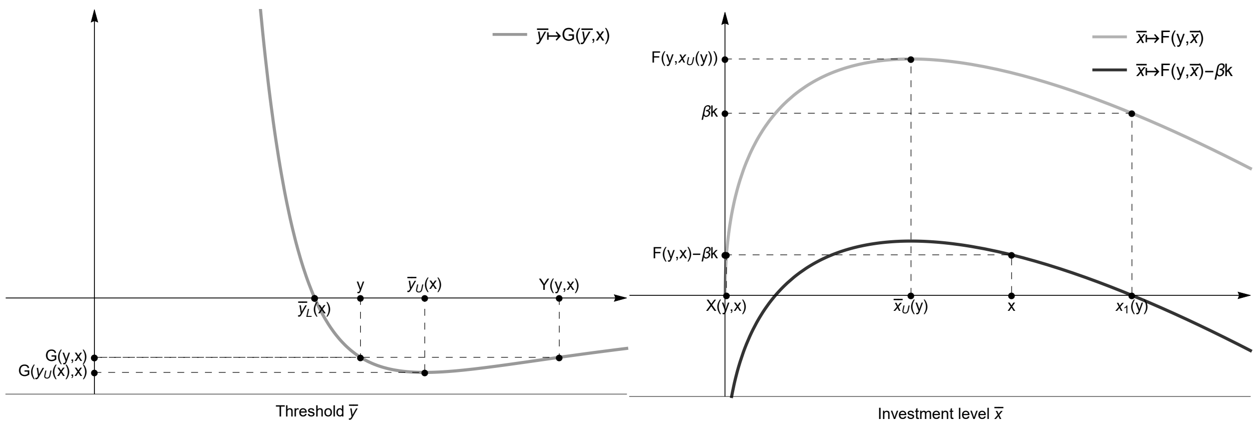
These definitions, in turn, allow us to state conditions (19a)-(19b) as

$$F(y, z(y)) = F(y, \bar{x}(y)) + \beta k, \tag{22a}$$

$$G(y, z(y)) = G(\bar{y}(z(y)), z(y)). \tag{22b}$$

Finding the solutions $\bar{x}(\cdot)$ and $z(\cdot)$ of the system (22a)–(22b) is certainly nontrivial because the unknown function $z(\cdot)$ appears in (22b) as an argument of $\bar{y}(\cdot)$, the inverse of $\bar{x}(\cdot)$ which is itself an unknown function. Lemmas 2.A and 2.B introduce two functions, Y and X, that will be useful for determining the boundary curves $\bar{x}(\cdot)$ and $z(\cdot)$.

Figure 2 Illustration of functions Y and X.



(a) Function Y (b) Function X

Note. Assuming $\delta(x)=\sqrt{x}$ with $r=0.05$, $\sigma=0.05$, $c=10$, and $k=15$. The marked points correspond to the choice of initial state $(y,x)=(1.5,3)$; here $y^\star\approx 1.3$.

LEMMA 2.A. *For any given $x>0$ there exists a unique solution, denoted $Y(y,x)\in[\bar{y}_L(x),\infty)$, of*

$$G(y,x)=G(Y(y,x),x) \tag{23}$$

that is distinct from y. This solution satisfies the following properties:

$$Y(\bar{y}_L(x),x)=\infty \qquad \textit{and} \qquad Y(y,x)\lesseqgtr\bar{y}_U(x)\iff y\gtreqless\bar{y}_U(x).$$

Panel (a) of Figure 2 helps us interpret the threshold $Y(y,x)$ by plotting the function $\bar{y}\mapsto G(\bar{y},x)$, where $\bar{y}>0$ denotes an arbitrary price threshold (possibly distinct from the current commodity price y). From the definitions of $\bar{y}_L(x)$ and G in (11a) and (21b), we can easily see that the discounted value of a marginal unit, or $\left[\frac{\partial v_L}{\partial x}(\bar{y},x)-c\right]\left(\frac{y}{\bar{y}}\right)^\beta$, is positive if the threshold $\bar{y}>y$ at which the firm supposedly invests *exceeds* the point $\bar{y}_L(x)$ in the reversible case but is negative otherwise. Threshold $\bar{y}_U(x)$ is the commodity price at which the net present (perpetuity) value of the marginal unit is largest. Equation (23) states that, if we ignore fixed costs k, then the threshold $Y(y,x)$ is "selected" such that the firm is indifferent (from a present value perspective) between investing marginally in capacity at the outset when the commodity price is y and investing in the future when that price $(Y_t;t\geq 0)$ reaches threshold $Y(y,x)>y$. It would be suboptimal to follow an investment rule based on a one-sided economic interpretation of $\bar{y}\mapsto G(\bar{y},x)$ because that approach disregards the fixed-cost parameter $k>0$. The function X is used to refine our approach.

LEMMA 2.B. *For any $y\geq y^\star$ with $y^\star$ given implicitly by*

$$F(y^\star,\bar{x}_U(y^\star))=\beta k, \tag{24a}$$

there exists a unique solution, denoted $X(y,x) \in [0, \bar{x}_U(y)]$*, of*

$$F(y, X(x,y)) = F(y,x) - \beta k. \tag{24b}$$

The function $x \mapsto X(y,x)$ *is defined on the domain* $[\bar{x}_U(y), x_1(y)]$*, with* $x_1(y)$ *given implicitly by*

$$F(y, x_1(y)) = \beta k. \tag{24c}$$

Panel (b) of Figure 2 illustrates the functions $\bar{x} \mapsto F(y,\bar{x})$ and $\bar{x} \mapsto F(y,\bar{x}) - \beta k$ for an arbitrary capacity level $\bar{x}$; it reveals that $X(y,x)$ is the implicit solution of (24b). Recall that $F(y,\bar{x})/[\beta - 1]$ in (21a) is the *net* perpetuity value of inframarginal capacity units (i.e., of all units below $\bar{x}$). The *gross* value of the marginal capacity unit decreases with capacity x because v_L is concave increasing in x. The net perpetuity value of $\bar{x}$ units of capacity (function F) increases as long as $\bar{x}$ is below the threshold, $\bar{x}_U(y)$, at which the value of the marginal unit falls to C and F attains its maximum. Investing more than $\bar{x}_U(y)$ units of capacity actually reduces the aggregate value $F(y,\bar{x})/[\beta - 1]$ because the value of each additional unit is less than C. The function $\bar{x} \mapsto F(y,\bar{x}) - \beta k$ allows consideration of the additional opportunity cost K arising from the presence of a fixed cost $k > 0$, which motivates our use of $X(y,x)$ as defined in (24b).

Using either function X or Y is not sufficient to determine $\bar{x}(\cdot)$ and $z(\cdot)$. Therefore, we consider conditions based on both functions in restating Proposition 3 as follows.

Restatement of Proposition 3. *In Proposition 3, the problem of finding boundary curves* $\bar{x}(\cdot)$ *and* $z(\cdot)$ *that satisfy* (19a)-(19b) *amounts to finding monotone increasing functions* $\bar{x}(\cdot)$ *and* $z(\cdot)$ *on the domain* $[y^\star, \infty)$ *such that*

$$0 \le \bar{x}(y) \le X(y, \bar{x}_U(y)) \le \bar{x}_U(y) \le z(y) \le \min\{\bar{x}_L(y), x_1(y)\}, \tag{25a}$$

$$z(y) = \bar{x}\big(Y(y, z(y))\big), \tag{25b}$$

$$\bar{x}(y) = X(y, z(y)). \tag{25c}$$

Because of the inequality $\bar{x}_U(y) \le z(y)$ in (25a), if the firm invests at the outset, then it installs more capacity when $k > 0$ than when $k = 0$, i.e., if $(y,x) \in \mathcal{S}$, it makes a larger lump-sum investment at the outset in the impulse control case than in the singular control case. This finding may seem counterintuitive in that a firm is usually believed to be reluctant to invest when doing so requires the commitment of sunk resources. Yet that perception is misplaced when $(y,x) \in \mathcal{S} \subset \mathcal{S}_U$ because the firm has already decided to invest; it is only the magnitude of the additional capacity that remains to be determined. The reason for our somewhat surprising result is that, given a fixed cost $k > 0$, the firm will *not* stop raising capacity as soon as the perpetuity value of an extra unit falls below the opportunity unit cost C, i.e., at a capacity level of $\bar{x}_U(y)$ units. Above the stock level $\bar{x}_U(y)$, the (perpetuity) value of an extra unit of capacity is still greater than the unit cost c—provided the firm does not hold more than $\bar{x}_L(y)$ capacity units in total. Hence the firm will invest "in bulk" until the capacity stock reaches a level $z(y)$, which is located between $\bar{x}_U(y)$ and $\bar{x}_L(y)$ as per the inequalities $\bar{x}_U(y) \leq z(y) \leq \min\{\bar{x}_L(y), x_1(y)\}$ in (25a); the investment level $z(y)$ is such that the value of the extra capacity is sufficient to recoup the fixed cost $k > 0$. If the fixed cost is negligible ($k = 0$) then we already know that the boundaries $\bar{x}(\cdot)$ and $z(\cdot)$ coincide, in which case we can infer from the inequalities $\bar{x}(y) \leq \bar{x}_U(y) \leq z(y)$ in (25a) that $\bar{x}(\cdot) \equiv \bar{x}_U(\cdot) \equiv z(\cdot)$ if $k = 0$. In other words, the two (s, S) boundaries degenerate to the single singular control boundary $\bar{x}_U(\cdot)$ when the fixed cost vanishes. This degeneracy ultimately makes the QVI of the singular control problem in (13a)-(13e) easier to solve.

7. Constructing an explicit solution for a particular case

Proposition 3 gives an explicit, regular solution (19c)-(19d) to the free boundary problem in (9a)-(9d). To establish that result, we need to find monotone increasing curves $\bar{x}(\cdot)$ and $z(\cdot)$ that satisfy (25a)–(25c). Unfortunately, it is a daunting task to proceed from that, i.e., (i) to state the conditions (25a)–(25c) in terms of model primitives and (ii) to establish that the solution to the FBP coincides with the minimum solution of the QVI (6a)-(6e) and ultimately with the value function (1).

We shall thus adopt a constructive approach and consider the following specific form for the operating profit function together with restrictions on the model parameters:

$$\delta(x) = \sqrt{x} \quad \text{and} \quad 3\sigma^2 = r - 3\mu > 0. \tag{26}$$

For this particular case, we construct an (s, S)-type policy, which we prove to be optimal, and also provide an explicit solution for the value function v.

Here the thresholds $\bar{x}_L(y)$ and $\bar{x}_U(y)$ of the benchmark cases, as given in (11b) and (14c), specialize to

$$\bar{x}_L(y) = \frac{y^2}{4c^2(r-\mu)^2} \quad \text{and} \quad \bar{x}_U(y) = \frac{y^2}{9c^2(r-\mu)^2}, \tag{27a}$$

while $y^\star$ and $x_1(y)$ as defined in (24a) and (24c) are now

$$y^\star = 3(r-\mu)\sqrt{kc} \quad \text{and} \quad x_1(y)^{1/2} = \frac{y + \sqrt{y^2 - 9kc(r-\mu)^2}}{3c(r-\mu)}, \tag{27b}$$

respectively.

7.1. The (s, S)-type impulse control policy

We are able to determine the curves $\bar{x}(\cdot)$ and $z(\cdot)$ for this particular case (see Appendices for details). When $\beta = 3$, $Y(y, x)$—which we defined implicitly in (23)—happens to be a root of a polynomial function of degree 3 that is distinct from y, which is also a root. This fact helps us express $Y(y,x)$ uniquely as the (positive) root of a quadratic function. Furthermore, the function X defined in (24b) also admits an explicit expression.

We introduce the following notation:

$$\psi(y) := \sqrt{\frac{z(y)}{\bar{x}_U(y)}} - 1, \tag{28a}$$

$$\varphi(y) := \frac{Y(y, z(y))}{y}, \tag{28b}$$

$$\rho(y) := \frac{3(r-\mu)\sqrt{kc}}{y}, \tag{28c}$$

where $\bar{x}_U(\cdot)$ is defined as in (27a). Doing so allows us to make changes of variables in order to express the conditions on $\bar{x}(\cdot)$ and $z(\cdot)$ in (25a)–(25c) in a more suitable manner. Proposition 4 (to follow) proves the existence of functions $\psi(\cdot)$, $\psi(\cdot)$, and $\rho(\cdot)$ and identifies the curves $\bar{x}(\cdot)$ and $z(\cdot)$ explicitly. Proposition 4 also states, in (30), that the (s,S) boundaries $\bar{x}(\cdot)$ and $z(\cdot)$ are asymptotically equivalent to the boundary $\bar{x}_U(\cdot)$ of the singular control problem as $y \to \infty$. In other words, the singular control policy ($k=0$) yields a benchmark that becomes increasingly reasonable when compared with the impulse control policy ($k>0$) as the commodity price y increases. The reason is that, as the commodity price becomes very large, the perpetuity value of additional capacity units also increases while the fixed cost term ($k>0$) becomes comparatively negligible. We asserted previously that $\bar{x}(\cdot) \equiv z(\cdot) \equiv \bar{x}_U(\cdot)$ when the fixed cost k vanishes; this relationship allows us to view the singular control boundary $\bar{x}_U(\cdot)$ as a degenerate case.

PROPOSITION 4. *The set of functions* $\psi : [y^\star, \infty) \to [0, 1/2]$ *that solve*

$$\sqrt{\frac{1+2\psi(y)/3}{1-2\psi(y)}} = 1 + \frac{2}{3}\psi(y) + \frac{2}{3}\sqrt{\rho^2(y) + \varphi^2(y)\psi^2(y\varphi(y))} \tag{29a}$$

and that also satisfy $\psi'(\cdot) \le 0$ *and* $(y\psi(y))' \ge 0$*—with* $\varphi : [y^\star, \infty) \to [1, \infty)$ *and* $\rho : [y^\star, \infty) \to (0, 1]$ *given by*

$$\varphi(y) = -\frac{1}{2} + \frac{3}{2}\sqrt{\frac{1+2\psi(y)/3}{1-2\psi(y)}} \tag{29b}$$

and (28c)*, respectively—is nonempty; that set admits a minimum and a maximum. The functions* $\psi(\cdot)$ *and* $\rho(\cdot)$ *vanish as* $y \to \infty$.

Considering the minimum (ψ) *among the set of solutions, the curves* $z : [\bar{y}_0, \infty) \to [0, \infty)$ *and* $\bar{x} : [\bar{y}_0, \infty) \to [0, \infty)$ *obtain explicitly from* (28a) *and*

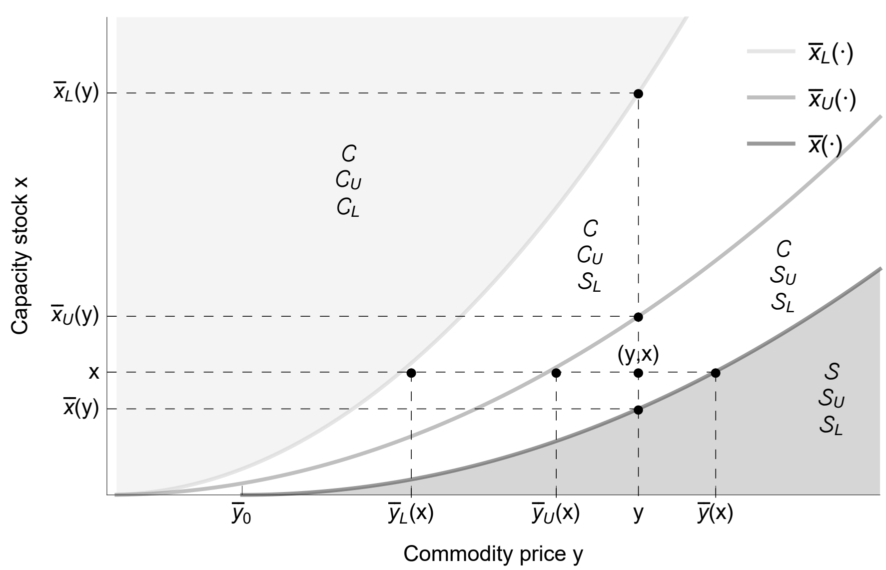
$$\sqrt{\frac{\bar{x}(y)}{\bar{x}_U(y)}} = 1 - \sqrt{\psi^2(y) + \rho^2(y)}, \tag{29c}$$

respectively, where $\bar{y}_0 > y^\star$ *is the unique value such that* $\psi(y) \lesseqgtr \sqrt{1-\rho(y)^2}$ *for all* $y \gtreqless \bar{y}_0$*. The inequalities in* (25a) *hold. In addition, the functions* $\bar{x}(\cdot)$ *and* $z(\cdot)$ *are monotone increasing, and we have*

$$\sqrt{\frac{\bar{x}(y)}{\bar{x}_U(y)}} \to 1 \ \ \textit{and} \ \ \sqrt{\frac{z(y)}{\bar{x}_U(y)}} \to 1 \quad \textit{as } y \to \infty. \tag{30}$$

Figures 3 and 4 illustrate the generalized impulse control policy obtained in Proposition 4.[9] Using as benchmarks the previously discussed cases of reversible investment (subscript L) and singular control investment (subscript U), Figure 3 presents the investment and waiting regions of the impulse control problem (*no* subscript) for a particular set of parameter values. The boundaries seen in this graph are $\bar{x}_L(\cdot)$ and $\bar{x}_U(\cdot)$ [from (27a)] as well as $\bar{x}(\cdot)$ [from (29c)]. The dark gray area below the $\bar{x}(\cdot)$ curve corresponds to the stopping set $\mathcal{S}$ of the optimal impulse control policy, which is a subset of $\mathcal{S}_U$ (if $k > 0$) and also of $\mathcal{S}_L$. The light gray area—above the $\bar{x}_L(\cdot)$ curve—is a noninvestment region for all three cases. The white area corresponds to the hysteresis set, of which the depicted state (y, x) is an element. Given a capacity stock of x units, the commodity price y represented in Figure 3 is above the price $\bar{y}_L(x)$ above which the firm would make a reversible investment but below the price $\bar{y}(x)$ below which a firm facing irreversibility and a fixed cost k would stay put. The converse statement also holds: for a given commodity price y, the stock of x capacity units represented in the graph is below $\bar{x}_L(y)$ (the threshold below which the firm would make a reversible investment) but above $\bar{x}(y)$ (the threshold above which the firm facing irreversibility and a fixed cost k would stay put).

Figure 3 Comparison of regions $\mathcal{S}_L$, $\mathcal{S}_U$, and $\mathcal{S}$.

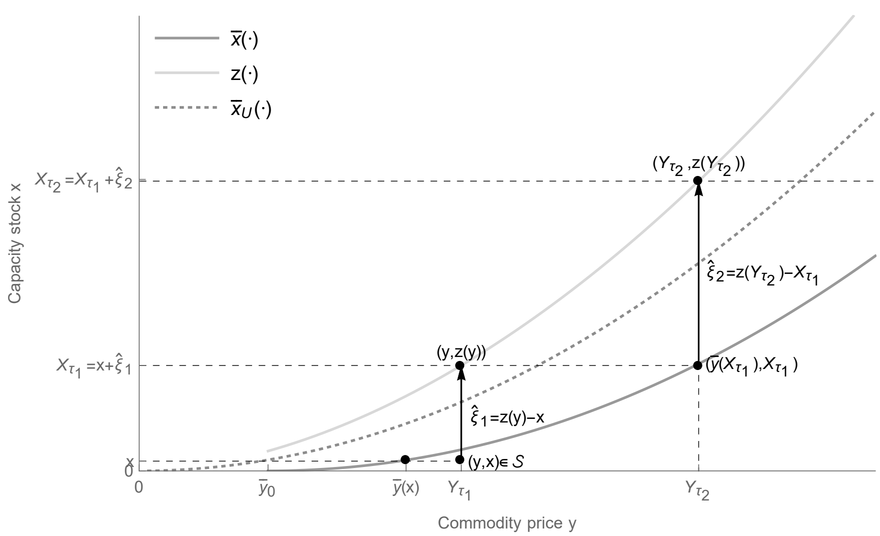


Note. Assuming $\delta(x) = \sqrt{x}$ with $r = 0.05$, $\sigma = 0.05$, $c = 10$, and $k = 15$; we obtain $\bar{y}_0 \approx 1.4$. The marked points correspond to the choice of initial state $(y, x) = (5.5, 18)$.

In this case, two effects lead to hysteresis. First, investment irreversibility creates hysteresis even when the firm does not incur a fixed adjustment cost ($k > 0$); this root cause of hysteresis was illustrated earlier in the singular control case (see Section 4.2). Second, we acknowledge that incurring a fixed cost k here expands the hysteresis set (widens the figure's white area), a feature reminiscent of key insights from the real options literature. The point (y, x) in Figure 3 is of particular interest in this regard. A firm that does not incur a fixed cost k would make an irreversible investment (because $(y, x) \in \mathcal{S}_U$), but that decision would be suboptimal for a firm that incurs such a cost (because $(y, x) \notin \mathcal{S}$).

Figure 4 focuses on the two boundaries of the generalized (s, S)-type policy, $\bar{x}(\cdot)$ and $z(\cdot)$, and illustrates two stages of capacity expansion for when the state (y, x) is initially (i.e., at $t = 0-$) in the stopping set $\mathcal{S}$. In this case, the firm would adjust its capacity stock ($X_{0-} = x$) immediately by making a lump-sum investment $\hat{\xi}_1 = z(y) - x$ in capacity and thereby attaining a stock of $X_{\hat{\tau}_1} = z(y)$ capacity units. The first investment takes place at time $\hat{\tau}_1 = 0$; the next capacity installment occurs at time $\hat{\tau}_2$, when the commodity price (Y_t; $t \geq 0$) reaches the (deterministic) level $\bar{y}(X_{\hat{\tau}_1})$ (which the price "hits from below"). At that (stopping) time $\hat{\tau}_2 = \inf\{t \geq \hat{\tau}_1 \mid Y_t = \bar{y}(X_{\hat{\tau}_1})\}$, the amount of investment is deterministic and is given by $\hat{\xi}_2 = z(Y_{\hat{\tau}_2}) - X_{\hat{\tau}_1}$. The evolution of capacity stock at later stages (i.e., at times $\hat{\tau}_3$, $\hat{\tau}_4$, ...) is not shown in Figure 4, but can easily be inferred following the same logic. Now if the state (y, x) starts instead in the continuation set $\mathcal{C}$ (a case also not illustrated in Figure 4), then the first investment will take place at some unknown time $\hat{\tau}_1$ in the future—namely, when the (stochastic) commodity price (Y_t; $t \geq 0$) reaches the (deterministic) level $\bar{y}(x)$.

Figure 4 **Generalized (s, S) impulse control policy with boundaries $\bar{x}(\cdot)$ and $z(\cdot)$.**



Note. Assuming $\delta(x) = \sqrt{x}$ with $r = 0.05$, $\sigma = 0.05$, $c = 10$, $k = 15$; we obtain $\bar{y}_0 \approx 1.4$. The marked points correspond to the choice of initial state $(y, x) = (3.5, 1.5)$.

Overall, the capacity stock $(\hat{X}_t; t \geq 0)$ evolves stochastically in stages characterized by "investment bursts." Although the timing of future capacity installments remains unknown (since the $\hat{\tau}_n$ are nondegenerate stopping times), the capacity stock reached by the end of each investment stage, or the $X_{\hat{\tau}_n}$, *is* known. The effect of a (stock) fixed cost on the optimal investment strategy was correctly intuited by Dixit and Pindyck (1994): "if there is a stock fixed cost . . . the optimal policy will allow the capital stock to jump in discrete steps at isolated moments" (p. 386). To the best of our knowledge, we are the first to substantiate their intuition by explicitly constructing the (s, S)-type policy.

7.2. Value of sequential capacity expansion options

For the particular case of (26), we already identified in Proposition 4 the (s, S) curves $\bar{x}(\cdot)$ and $z(\cdot)$. Our next proposition completes the analysis by identifying a function v that coincides with the value function (1), thereby proving the optimality of the (s, S)-type policy we just described.

PROPOSITION 5. *For the particular case of* (26)*, the explicit function* v *defined by*

$$v(y,x) = \begin{cases} v_L(y,x) + A(x)y^3, & x \geq \bar{x}(y), \\ v_L(y,\bar{x}(y)) + A(\bar{x}(y))y^3 - c[\bar{x}(y) - x], & x \leq \bar{x}(y), \end{cases} \tag{31a}$$

with $v_L(y) = y\sqrt{x}/[r-\mu]$ *and*

$$A(x)y^3 = \int_x^\infty \left(\frac{1}{2}\frac{\bar{y}(\xi)\xi^{-1/2}}{r-\mu} - c\right)\left(\frac{y}{\bar{y}(\xi)}\right)^3 \mathrm{d}\xi \geq 0, \tag{31b}$$

satisfies the regularity (8) *and is a solution of the QVI* (6a)-(6e) *in the interval* $[v_L, v_U]$*. This function* v *coincides with the value function* (1)*.*

Thus the expression in (31a)-(31b) substitutes for the one in (19c)-(19d) in this particular case. That substitutability is especially interesting because (19c) explicitly incorporates the fixed cost component $k > 0$ and the investment level $z(\cdot)$, whereas (31a) does not. Yet we remark that, for $y \geq \bar{y}(x)$ or $x \leq \bar{x}(y)$, (31a) involves a capacity level (namely, $\bar{x}(y)$) that is lower than $z(y)$ in (19c). The expression (31a) shares some features of (14a), but the terms for capacity levels ($\bar{x}(\cdot)$ vs. $\bar{x}_U(\cdot)$) and for the value of capacity expansion options (A vs. A_U) differ significantly. This follows because our more general setting involves a fixed cost, which reduces firm value (i.e., $v \leq v_U$).

8. Conclusion

McDonald and Siegel (1986) proposed a seminal approach to modeling the launch of an uncertain project as an optimal timing problem in which the fixed investment cost plays the role of the exercise price of a perpetual American call option. However, a project's value may depend on decisions (about, e.g., the production scale) that are made at the time(s) of exercising that option. We modeled a firm's staged capacity investments under uncertainty as the sequential exercising of compound capacity expansion options with optimal capacity choices at each investment time.

We derived and solved a quasi-variational inequality involving two state variables. We constructed a generalized (s, S) impulse control policy and determined the value function in closed form for a specific class of operating profit functions and under some restrictions on the model parameters. We then compared our results with benchmarks in the economics literature. Traditional models of capacity accumulation are concerned with the endogenous, gradual buildup of capital at the macroeconomic level. However, they disregard fixed costs—which matter at the microeconomic firm level and are central to real options theory—and offer no predictions about the level of lump-sum investment in capacity. We showed that the hysteresis or inertial region expands when fixed costs are relevant. More specifically, we showed that the firm does not acquire a marginal capacity unit as soon as the marginal (perpetuity) value exceeds the marginal (opportunity) cost; instead, it delays until the net aggregate value to be gained from the extra capacity exceeds the fixed cost by a margin sufficient to cover the opportunity cost of "killing the delay option." This setting gives rise to a lump-sum capacity installment at each investment time, an outcome that better represents investment behavior at the microeconomic (firm) level (as argued by Pindyck 1988). We also showed that a firm investing at the outset will invest more if it does than if it does not incur a fixed cost. The reason is that any capacity unit (above $\bar{x}_U(y)$) that helps generate a value greater than the linear cost c should be utilized to recoup the fixed opportunity investment cost K.

Our approach is novel and rigorous, yet we believe that extending it further would enhance our understanding of investment behaviors. In particular, the assumptions about the firm's profit function and parameter values are fairly restrictive and meant to ensure the model's tractability. Investigating larger classes of profit functions may serve to establish the generality of our approach. For instance, one could interpret the capacity state as an upper bound on the output rate, thereby allowing for production flexibility with the firm adjusting output in view of the realized commodity price. It may also be fruitful to consider the case where a firm's output affects the price-setting mechanism—that is, with price dynamics that are at least partly endogenous. More generally, we have assumed that the stochastic differential equation driving the price process is beyond the firm's control; also, we did not allow the capacity stock to depreciate, be sold, or be driven by a diffusion term. It would also be interesting to establish whether our key results on the structure of the optimal generalized impulse control policy still hold when considering, say, a mean-reverting diffusion process. Finally, although there is an expanding literature that addresses real options in imperfectly competitive settings (for a review, see Chevalier-Roignant et al. 2011), hardly any scholars have considered strategic capacity expansions in a continuous-time setting with fixed costs. A notable exception is the contribution by Huisman and Kort (2015), but these authors did not consider a *sequence* of capacity expansions. The incentive to overinvest in capacity to deter and/or accommodate a rival's installing more capacity may be reduced if investments are staged instead of being made only once.

Endnotes

1. Economists distinguish between *stock* fixed costs (which are paid at once) and *flow* fixed costs (which are paid over a longer time span). Only the former are relevant here because, by construction, the firm invests only "at isolated moments," meaning that the set of times at which the firm invests has probability measure 0.

2. The early literature on irreversible capital investment (see, e.g., Arrow 1968, Arrow and Kurz 1970), which ignored uncertainty, used Pontryagin's maximum principle for continuous control. Later models considered (partial) irreversibility under uncertainty and formally addressed a singular control problem (see Section 4.2 for a definition).

3. Our assumption of the firm's risk neutrality is not necessary for the results derived in this paper to hold *provided* the market is complete and agents have no arbitrage opportunities. Our analysis would then apply with the probability measure $\mathbb{P}$ being interpreted as the equivalent martingale measure. The assumption of market completeness could be problematic in real options analysis because the underlying asset may not even be tradable or there might not be traded assets with an equivalent risk profile. We shall ignore this issue because it is not central to the problem being addressed.

4. The nonlinear operator $\mathcal{M}$ is monotone increasing in the sense that $\varphi(\cdot) \leq \psi(\cdot)$ implies $\mathcal{M}\varphi(\cdot) \leq \mathcal{M}\psi(\cdot)$. The constraint $z \geq x$ in (3) ensures that $\xi_n > 0$.

5. The function v_L is a lower bound because $\mathcal{L}(v - v_L)(y,x) \geq 0$, $(v - v_L)(0,x) = 0$, and $\liminf_{y\to\infty}\{v(y,x) - v_L(y,x)\} \geq 0$.

6. The term "càdlàg" is an acronym used to characterize a process that is "continu à droite avec des limites à gauche," meaning in French "right continuous with left limits." A singular control is a mathematical abstraction. Formally, it corresponds to a case where the firm invests repeatedly by infinitesimal "lumps" or impulses. Continuous control does not allow for impulse. The second integral in (12) is defined in the sense of Lebesgue-Stieltjes, which is meaningful here because ξ is monotone.

7. Inequality (13b) follows intuitively from limit considerations: (6b) reduces to $v_U(y,x) - cx \geq \sup_{z\geq x}\{v_U(y,z) - cz\}$ as $k \to 0$; hence $z \mapsto v_U(y,z) - cz$ decreases, which leads to (13b).

8. This classical construction anchored in dynamic programming was introduced by Bensoussan and Lions (1984, chap. 7) in the context of diffusion processes. Øksendal and Sulem (2007, chap. 7) used a comparable scheme generalized to jump-diffusion processes using probabilistic techniques.

9. To generate the curves, we define a sequence $\{\zeta^n(\cdot)\}_n$ as in the proof of Lemma 10 (see Appendix F). In that lemma, we prove analytically that this sequence converges. We set $\psi(\cdot) = \zeta_N(\cdot)$ for some large N; here, $N = 13$, above which the improvement (in terms of accuracy) from $\zeta_{n-1}(y)$ to $\zeta_n(y)$ (for $y = 5$) falls below 2.3×10^{-4}. Thresholds are determined by using (29c) and (28a). Mathematica codes are available upon request.

Acknowledgments

Research supported by the National Science Foundation under Grant nos. DMS-1303775 and DMS-1612880 and the Research Grants Council of the Hong Kong Special Administrative Region (CityU-500113, CityU-11303316). The authors thank Christoph Flath and participants at the 2017 YARO workshop, Leipzig.

References

Abel, A. B. and J. C. Eberly (1994). A unified model of investment under uncertainty. *American Economic Review 84*(5), 1369–1384.

Abel, A. B. and J. C. Eberly (1997). An exact solution for the investment and value of a firm facing uncertainty, adjustment costs, and irreversibility. *Journal of Economic Dynamics and Control 21*(4-5), 831–852.

Alvarez, L. H. (2011). Optimal capital accumulation under price uncertainty and costly reversibility. *Journal of Economic Dynamics and Control 35*(10), 1769–1788.

Arrow, K. J. (1968). Optimal capital policy with irreversible investment. In J. N. Wolfe (Ed.), *Value, Capital and Growth*. Edinburgh University Press.

Arrow, K. J. and M. Kurz (1970). *Public investment, the rate of return and optimal fiscal policy.* John Hopkins Press.

Bar-Ilan, A., D. Perry, and W. Stadje (2004). A generalized impulse control model of cash management. *Journal of Economic Dynamics and Control 28*(6), 1013–33.

Bellman, R. (1957). *Dynamic Programming.* Princeton University Press.

Bensoussan, A. (2011). *Dynamic Programming and Inventory Control*, Volume 3 of *Studies in Probability, Optimization and Statistics.* IOS Press.

Bensoussan, A. and J.-L. Lions (1982). *Applications of Variational Inequalities in Stochastic Control*, Volume 13 of *Studies in Mathematics and Its Applications.* New York, NY: North-Holland.

Bensoussan, A. and J.-L. Lions (1984). *Impulse Control and Quasi-Variational Inequalities.* Paris, France: Gauthiers-Villars.

Chevalier-Roignant, B., C. M. Flath, A. Huchzermeier, and L. Trigeorgis (2011). Strategic investment under uncertainty: A synthesis. *European Journal of Operational Research 215*(3), 639–650.

Constantinides, G. M. (1976). Stochastic cash management with fixed and proportional transaction costs. *Management Science 22*(12), 1320–1331.

Constantinides, G. M. and S. F. Richard (1978). Existence of optimal simple policies for discounted-cost inventory and cash management in continuous time. *Operations Research 26*(4), 620–636.

Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy 97*(3), 620–638.

Dixit, A. K. and R. S. Pindyck (1994). *Investment under Uncertainty.* Princeton, NJ: Princeton University Press.

Ferrari, G. (2015). On an integral equation for the free-boundary of stochastic, irreversible investment problems. *Annals of Applied Probability 25*(1), 150–176.

Guo, X. and G. Wu (2009). Smooth fit principle for impulse control of multidimensional diffusion processes. *SIAM Journal on Control and Optimization 48*(2), 594–617.

Harrison, J. M., T. M. Sellke, and A. J. Taylor (1983). Impulse control of Brownian motion. *Mathematics of Operations Research 8*(3), 454–466.

Hubbard, R. G. (1994). Investment under uncertainty: Keeping one's options open. *Journal of Economic Literature 32*(4), 1816–1831.

Huisman, K. J. and P. M. Kort (2015). Strategic capacity investment under uncertainty. *RAND Journal of Economics 46*, 376–408.

Kobila, T. (1993). A class of solvable stochastic investment problems involving singular controls. *Stochastics and Stochastic Reports 43*(1-2), 29–63.

McDonald, R. and D. Siegel (1986). The value of waiting to invest. *Quarterly Journal of Economics 101*(4), 707–728.

Øksendal, A. (2000). Irreversible investment problems. *Finance and Stochastics 4*(2), 223–250.

Øksendal, B. K. and A. Sulem (2007). *Applied stochastic control of jump diffusions*, Volume 498. Springer.

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

Riedel, F. and X. Su (2011). On irreversible investment. *Finance and Stochastics 15*(4), 607–633.

Trigeorgis, L. (1996). *Real options: Managerial flexibility and strategy in resource allocation.* Cambridge, MA: The MIT Press.

Appendix A: Proof of Proposition 1

Free-boundary problem

We look for a solution v_U of the QVI (13a)-(13e) that satisfies the regularity (8). We consider the free boundary problem with threshold $\bar{y}_U(\cdot)$ given by

$$\mathcal{L}v_U(y,x) = y\delta(x), \qquad y \le \bar{y}_U(x), \tag{32a}$$

$$\frac{\partial v_U}{\partial x}(y,x) = c, \qquad y \ge \bar{y}_U(x). \tag{32b}$$

After a few calculations, we find that the function v_U given in (14a) satisfies the regularity (8) and also solves the FBP in (32a)-(32b) together with the boundary conditions (13d)-(13e)

Quasi-variational inequality

Note that a solution v_U of the FBP in (32a)-(32b) must also satisfy the inequalities

$$\mathcal{L}v_U(y,x) \ge y\delta(x), \qquad \text{a.e. } y \ge \bar{y}_U(x), \tag{33a}$$

$$\frac{\partial v_U}{\partial x}(y,x) \le c, \qquad \text{a.e. } y \le \bar{y}_U(x), \tag{33b}$$

if it is to solve the QVI (13a)-(13e).

Case 1: $y \ge \bar{y}_U(x)$. Verifying inequality (33a) is equivalent to showing that $\mathcal{L}v_U(y,x) \ge y\delta(x)$ if $x \le \bar{x}_U(y)$. From (14a), the smooth-fit conditions, and our definition of $\bar{x}_U(\cdot)$ as the inverse of $\bar{y}_U(\cdot)$, it follows that

$$\mathcal{L}v_U(y,x) - y\delta(x) = y[\delta(\bar{x}_U(y)) - \delta(x)] - rc[\bar{x}_U(y) - x], \quad \forall x < \bar{x}_U(y).$$

Hence inequality (33a) is satisfied if $x \mapsto rcx - y\delta(x)$ decreases on the interval $(0, \bar{x}_U(y)]$. In other words: inequality (33a) is satisfied if $rc \le y\delta'(x)$ for all $x \le \bar{x}_U(y)$ or $y > rc/\delta'(x)$ for all $y \ge \bar{y}_U(x)$. By our definition of threshold $\bar{y}_U(x)$ in (14b), the latter condition is satisfied if $(r-\mu)\beta \ge r(\beta - 1)$, which holds because $\beta < r/\mu$.

Case 2: $y \le \bar{y}_U(x)$. Here we must check that $y\delta'(x)/[r-\mu] + A'_U(x)y^\beta \le c$ for all $y \ge \bar{y}_U(x)$. From (14e) we obtain that the inequality (33b) is satisfied if

$$\frac{y\delta'(x)}{r-\mu} - \left(\frac{\beta-1}{c}\right)^{\beta-1}\left[\frac{y\delta'(x)}{(r-\mu)\beta}\right]^{\beta} \le c, \quad \forall y \ge \bar{y}_U(x). \tag{34}$$

Setting $u = (\beta-1)y\delta'(x)/[c(r-\mu)]$ and $\alpha = 1/[\beta - 1]$, we can rewrite the inequality in (34) as $\beta\alpha u - \alpha u^\beta - 1 \le 0$ for all $u \in [0,1]$, which is easily verified. This completes the proof that the function v_U in (14a) solves the QVI (13a)-(13e) with the regularity (8).

We now establish a useful relationship between v_L and v_U.

Lemma 3. *The solution of the QVI of the singular-control problem, v_U, majorizes the perpetuity value of capacity, v_L, in* (7).

Proof. For the case $y \leq \bar{y}_U(x)$, it is clear from (14a) that inequality $v_U(y,x) > v_L(y,x)$ is satisfied. For the case $y \geq \bar{y}_U(x)$, we must show that

$$cx + \frac{y\delta(\bar{x}_U(y))}{r-\mu} + A_U(\bar{x}_U(y))y^\beta - c\bar{x}_U(y) \geq \frac{y\delta(x)}{r-\mu}. \tag{35}$$

Because $A_U(x) \geq 0$, it is sufficient to check that

$$\frac{y\delta(\bar{x}_U(y))}{r-\mu} - c\bar{x}_U(y) \geq \frac{y\delta(x)}{r-\mu} - cx \quad \forall x \leq \bar{x}_U(y).$$

This inequality holds whenever $x \mapsto y\delta(x)/[r-\mu] - cx$ is monotone increasing on $(0, \bar{x}_U(y)]$. Therefore, it holds if $y \geq c(r-\mu)/\delta'(x)$ when $y \geq \bar{y}_U(x)$, a property which is easily established. It follows immediately that $v_U \geq v_L$. $\square$

Here we did *not* establish that the function v_U given in (14a) coincides with the value function of the singular control problem in (12). Proving that result would require a verification theorem and the rigorous construction of a reflected process along the boundary curve (i.e., of a singular control). This result is known from the mathematical literature (see, e.g., Kobila 1993). In any event, we interpret function v_U not as the value function of the singular control problem but rather as an upper bound for the solution of the QVI of the impulse control problem.

Appendix B: Proof of Lemma 1

We prove the lemma by induction. Assume that $v_n \leq v_U$, which holds if $n = 0$ (see Lemma 3). We want to show that $v_{n+1} \leq v_U$ follows. Toward that end we consider the variational inequality in (15a)-(15e), for which $v_{n+1}(y,x) = \mathcal{M}v_n(y,x)$ in the stopping region and $\mathcal{L}v_{n+1}(y,x) = y\delta(x)$ in the continuation region. In the *stopping* region, "Property A"—that is, $v_{n+1}(y,x) \leq \mathcal{M}v_U(y,x) < v_U(y,x)$—holds because (i) $v_n \leq v_U$ (by the induction assumption) and so $\mathcal{M}v_n \leq \mathcal{M}v_U$ (see note 4) and (ii) $v_U(y,x) > \mathcal{M}v_U(y,x)$ if $k > 0$. In the *continuation* region, "Property B," — i.e., $\mathcal{L}(v_U - v_{n+1})(y,x) \geq 0$ — is immediate. Together, Properties A and B imply that $(v_U - v_{n+1})(y,x) \geq 0$. This is because, if there exists a point y such that $(v_U - v_{n+1})(y,x) < 0$ (recall that x is fixed) then it can occur only inside the continuation region. But in that case, there would be a negative local minimum in this region, which is impossible because $\mathcal{L}(v_U - v_{n+1})(y,x) \geq 0$. The proof of $v_n \geq v_L$ proceeds similarly.

Let us now use induction to establish that $v_n \leq v_{n+1}$, which holds for $n = 0$. Suppose the inequality holds for n, and consider the problem defining v_{n+2}. By the induction assumption and the monotonicity of the operator $\mathcal{M}$ (see note 4), we have $v_{n+2} \geq \mathcal{M}v_{n+1} \geq \mathcal{M}v_n$. Hence $v_{n+2}(y,x) \geq v_{n+1}(y,x) = \mathcal{M}v_n(y,x)$ in the stopping region and $\mathcal{L}(v_{n+2} - v_{n+1})(y,x) \geq 0$ in the continuation region. These statements imply that $(v_{n+2} - v_{n+1})(y,x) \geq 0$ and so the induction goes through.

We have therefore established the inequalities in (16). Passing to the limit in (15a)-(15e), we find that the limit of the sequence $\{v_n\}_n$, denoted $\underline{v}$, solves the QVI (6a)-(6e) in the classical sense if $\underline{v}$ satisfies the regularity (8).

Appendix C: Proof of Proposition 2

Our proof proceeds in several steps articulated as lemmas. We also introduce a new problem related to the QVI (6a)-(6c). This problem amounts to solving the inequalities in (6a) and (6b) but *not* (necessarily) the complementary slackness condition (6c):

$$\min\{\mathcal{L}v(y,x) - y\delta(x), v(y,x) - \mathcal{M}v(y,x)\} \geq 0, \quad \text{a.e. } y \in \mathbb{R}_+,\ \forall x \in \mathbb{R}_+. \tag{36}$$

We call a function with the regularity (8) that solves (36) and satisfies the boundary conditions (6d)-(6e) an *upper solution* of the QVI (6a)-(6e). To limit notations, we use v to signify both the solution and the upper solution of (6a)-(6e). By definition, an upper solution of (6a)-(6e) need not be a solution of (6a)-(6e). Of course, any solution of (6a)-(6e) is an upper solution. The function v_U in (14)—which is a regular solution of the QVI (13a)-(13e)—is an example of an upper solution. In fact, the inequalities (13a) and (6a) are identical. Besides, since $x \mapsto v_U(y,x) - cx$ decreases, it follows that $v_U(y,x) = \sup_{\xi \geq 0}[v_U(y, x+\xi) - c\xi]$. Hence $v_U(y,x) > \mathcal{M}v_U(y,x)$ as in (6b). Lemma 4 gives an interpretation of the minimum solution, $\underline{v}$, in terms of upper solutions, while Lemma 5 substantiates the terminology "upper solution."

Lemma 4. *A minimum solution $\underline{v}$ that satisfies the regularity* (8) *is the smallest upper solution of* (6a)-(6e).

Proof. Suppose a function v with the regularity (8) majorizes v_L, solves (36) and satisfies (6d)-(6e); v is by definition an upper solution. Consider the sequence $\{v_n\}_{n\geq 0}$ defined in (15a)-(15e), which we proved was both increasing and converging. We wish to establish the inequality $v_n \leq v$ by induction. The proof follows similarly to the one given for $v_n \leq v_U$ in Lemma 1, except now in the stopping region we write $v_{n+1}(y,x) = \mathcal{M}v_n(y,x) \leq \mathcal{M}v(y,x) \leq v(y,x)$, which follows from the induction property and the definition of v as an upper solution. We have thus established the induction property. We complete the proof by recalling that the minimum solution is the limit of the sequence $\{v_n\}_n$. □

Lemma 5. *If a function v with the regularity* (8) *majorizes v_L, solves* (36) *and satisfies* (6d)-(6e)*, then it majorizes the value function in* (1)*; hence the terminology "upper solution."*

Proof. Let the control $\nu = \{\tau_n, \xi_n\}_n$ be admissible as per (2). The function v satisfies the regularity (8). The classical Itô's lemma applies to functions that are C^2, i.e., more regular. We here employ a generalized version (see Bensoussan and Lions 1982, Thm. 8.5, pp. 185–86); this theorem applies because functions that are regular in the sense of (8) belongs to the relevant Sobolev space. We use this generalized version between τ_n and τ_{n+1} and use inequality (6a) to obtain

$$\mathbb{E}[v(Y_{\tau_n}, X_{\tau_n})e^{-r\tau_n}] \geq \mathbb{E}\left[\int_{\tau_n}^{\tau_{n+1}} Y_t\delta(X_t)e^{-rt}\,\mathrm{d}t\right] + \mathbb{E}[v(Y_{\tau_{n+1}}, X_{\tau_n})e^{-r\tau_{n+1}}]. \tag{37}$$

By (6b), we have

$$v(Y_{\tau_{n+1}}, X_{\tau_n}) \geq v(Y_{\tau_{n+1}}, X_{\tau_{n+1}}) - k - c\xi_{n+1};$$

therefore,

$$\mathbb{E}[v(Y_{\tau_n}, X_{\tau_n})e^{-r\tau_n}] \geq \mathbb{E}\left[\int_{\tau_n}^{\tau_{n+1}} Y_t\delta(X_t)e^{-rt}\,\mathrm{d}t\right] \\ + \mathbb{E}[v(Y_{\tau_{n+1}}, X_{\tau_{n+1}})e^{-r\tau_{n+1}}] - \mathbb{E}[(k + c\xi_{n+1})e^{-r\tau_{n+1}}].$$

Because $v \geq 0$, this inequality holds for any admissible control ν. Summing from $n = 0$ to $N - 1$ now yields

$$v(y,x) \geq \mathbb{E}\left[\int_0^{\tau_N} Y_t\delta(X_t)e^{-rt}\,\mathrm{d}t - \sum_{n=1}^{N}(k + c\xi_n)e^{-r\tau_n}\right] + \mathbb{E}[v(Y_{\tau_N}, X_{\tau_N})e^{-r\tau_N}], \tag{38}$$

and since $v \geq 0$ it follows that

$$v(y,x) \geq \mathbb{E}\left[\int_0^{\tau_N} Y_t\delta(X_t)e^{-rt}\,\mathrm{d}t - \sum_{n=1}^{N}(k + c\xi_n)e^{-r\tau_n}\right].$$

If we let $N \to \infty$ and recall that the control ν is admissible, then it follows immediately that

$$v(y,x) \geq J_{y,x}(\nu) := \mathbb{E}\left[\int_0^{\infty} Y_t\delta(X_t^{\nu})e^{-rt}\,\mathrm{d}t - \sum_{n\geq 1}(k + c\xi_n)e^{-r\tau_n}\right].$$

Since the control ν is arbitrary, an upper solution on the LHS majorizes the value function given in (1). $\square$

It is now immediate from Lemmas 4 and 5 that, if the minimum solution $\underline{v}$ satisfies the regularity (8), then it majorizes the value function in (1). We shall prove the following lemma.

Lemma 6. *Let the assumptions listed in Proposition 2 hold, where $\underline{v}$ and $\hat{\nu}$ denote (respectively) the regular minimum solution and the admissible impulse control corresponding to the (s,S) policy. Then*

$$\underline{v}(y,x) = J_{y,x}(\hat{\nu}). \tag{39}$$

An immediate consequence of this lemma is that $\underline{v}(y,x) = \sup_{\nu} J_{y,x}(\nu)$. This equality proves that the (regular) minimum solution coincides with the value function.

Proof of Lemma 6. Equation (6c) implies that $\mathcal{L}\underline{v}(Y_t, X_{\hat{\tau}_n\wedge T}) = Y_t\delta(\hat{X}_t)$ for $t \in (\hat{\tau}_n \wedge T, \hat{\tau}_{n+1} \wedge T)$. If $(\hat{\tau}_n \wedge T, \hat{\tau}_{n+1} \wedge T) \neq \emptyset$, then $\hat{\tau}_n < \hat{\tau}_{n+1} \leq T$. Using the aforementioned generalized Itô's lemma to the regular minimum solution $\underline{v}$ between $\hat{\tau}_n$ and $\hat{\tau}_{n+1}$, we can write

$$\underline{v}(Y_{\hat{\tau}_{n+1}\wedge T}, X_{\hat{\tau}_n\wedge T})e^{-r(\hat{\tau}_{n+1}\wedge T)} + \int_{\hat{\tau}_n\wedge T}^{\hat{\tau}_{n+1}\wedge T} Y_t\delta(X_{\hat{\tau}_n\wedge T})e^{-rt}\,\mathrm{d}t \\ = \int_{\hat{\tau}_n\wedge T}^{\hat{\tau}_{n+1}\wedge T} \sigma\frac{\partial\underline{v}}{\partial y}(Y_t, X_{\hat{\tau}_n\wedge T})Y_te^{-rt}\,\mathrm{d}W_t + \underline{v}(Y_{\hat{\tau}_n\wedge T}, X_{\hat{\tau}_n\wedge T})e^{-r(\hat{\tau}_n\wedge T)}.$$

This relation is obviously true when $\hat{\tau}_n \wedge T = \hat{\tau}_{n+1} \wedge T$. Therefore,

$$\mathbb{E}\left[\underline{v}(Y_{\hat{\tau}_{n+1}\wedge T}, X_{\hat{\tau}_n\wedge T})e^{-r(\hat{\tau}_{n+1}\wedge T)} + \int_{\hat{\tau}_n\wedge T}^{\hat{\tau}_{n+1}\wedge T} Y_t \delta(X_{\hat{\tau}_n\wedge T})e^{-rt}\,\mathrm{d}t\right] = \mathbb{E}[\underline{v}(Y_{\hat{\tau}_n\wedge T}, X_{\hat{\tau}_n\wedge T})e^{-r(\hat{\tau}_n\wedge T)}].$$

We can now use the definitions of $\hat{X}_t$ and $X_{\hat{\tau}_{n+1}\wedge T}$ to write

$$\begin{aligned}\mathbb{E}[\underline{v}(Y_{\hat{\tau}_n\wedge T}, X_{\hat{\tau}_n\wedge T})e^{-r(\hat{\tau}_n\wedge T)}] = \mathbb{E}\Big[&\underline{v}(Y_{\hat{\tau}_{n+1}\wedge T}, X_{\hat{\tau}_{n+1}\wedge T})e^{-r(\hat{\tau}_{n+1}\wedge T)} \\ &+ \int_{\hat{\tau}_n\wedge T}^{\hat{\tau}_{n+1}\wedge T} Y_t\delta(\hat{X}_t)e^{-rt}\,\mathrm{d}t - (k + c\hat{\xi}_{n+1})e^{-r\hat{\tau}_{n+1}}\mathbb{1}_{\{\hat{\tau}_{n+1}<T\}}\Big].\end{aligned}$$

Summing between $n = 0$ and $N - 1$, we obtain

$$\begin{aligned}\underline{v}(y,x) = \mathbb{E}\Big[&\underline{v}(Y_{\hat{\tau}_N\wedge T}, X_{\hat{\tau}_N\wedge T})e^{-r(\hat{\tau}_N\wedge T)} \\ &+ \int_0^{\hat{\tau}_N\wedge T} Y_t\delta(\hat{X}_t)e^{-rt}\,\mathrm{d}t - \sum_{n=0}^{N-1}(k + c\hat{\xi}_{n+1})e^{-r\hat{\tau}_{n+1}}\mathbb{1}_{\{\hat{\tau}_{n+1}<T\}}\Big].\end{aligned}$$

We now let $N \to \infty$ and use that $\hat{\tau}_N \to \infty$ almost surely. Then

$$\underline{v}(y,x) = \mathbb{E}\left[v(Y_T, \hat{X}_T)e^{-rT} + \int_0^T Y_t\delta(\hat{X}_t)e^{-rt}\,\mathrm{d}t - \sum_{n=0}^{\infty}(k + c\hat{\xi}_{n+1})e^{-r(\hat{\tau}_{n+1}\wedge T)}\mathbb{1}_{\{\hat{\tau}_{n+1}<T\}}\right]. \tag{40}$$

We note that, by (16),

$$0 \le \mathbb{E}[\underline{v}(Y_T, \hat{X}_T)e^{-rT}] \le \mathbb{E}[v_U(Y_T, \hat{X}_T)e^{-rT}].$$

The RHS term vanishes as $T \to \infty$ because of the transversality condition (18) and so $\mathbb{E}[\underline{v}(Y_T, \hat{X}_T)e^{-rT}] \to 0$. Because $\hat{\nu}$ is admissible as per (2), we have

$$\lim_{T\to\infty} \mathbb{E}\left[\sum_{n=0}^{\infty}(k + c\hat{\xi}_{n+1})e^{-r(\hat{\tau}_{n+1}\wedge T)}\mathbb{1}_{\{\hat{\tau}_{n+1}<T\}}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty}(k + c\hat{\xi}_{n+1})e^{-r\hat{\tau}_{n+1}}\right].$$

Hence we can pass to the limit in (40) as $T \to \infty$ and thereby obtain $\underline{v}(y,x) = J_{y,x}(\hat{\nu})$. $\square$

In the proof of Lemma 6, we can replace the minimum solution $\underline{v}$ by any solution of the QVI (6a)-(6e) along with an impulse control ν provided the conditions listed in Proposition 2 are satisfied. It follows that any such solution is unique and coincides with both the value function and the smallest upper solution.

Appendix D: Proof of Proposition 3

We assume that the function $\bar{y}(\cdot)$ is monotone increasing and set $\bar{y}_0 = \bar{y}(0)$. The inverse of $\bar{y}(\cdot)$, denoted $\bar{x}(\cdot)$, is defined on $[\bar{y}_0, \infty)$. By arguments similar to those in the proof of Proposition 1, the FBP (9a)-(9c) translates into

$$v(y,x) = \begin{cases}\frac{y\delta(x)}{r-\mu} + A(x)y^{\beta}, & x \in [\bar{x}(y), \infty),\\ \sup_{z>x}[v(y,z) - cz] + cx - k, & x \in [0, \bar{x}(y)].\end{cases} \tag{41}$$

Equation (41) can be written as $v(y,x) = \sup_{z \geq \bar{x}(y)}[v(y,z) - cz] + cx - k$ for $x \in [0, \bar{x}(y)]$. Consider the point of maximum, denoted $z(y)$. Necessarily, $z(y) > \bar{x}(y)$. We can use (41) to write the optimality condition as

$$\frac{y\delta'(z(y))}{r-\mu} + A'(z(y))y^{\beta} = c. \tag{42}$$

Hence (41) becomes

$$v(y,x) = \begin{cases} \frac{y\delta(x)}{r-\mu} + A(x)y^{\beta}, & x \in [\bar{x}(y), \infty), \\ \frac{y\delta(z(y))}{r-\mu} + A(z(y))y^{\beta} - c[z(y) - x] - k, & x \in [0, \bar{x}(y)]. \end{cases} \tag{43}$$

The solution thus depends on three unknowns: $z(\cdot)$, $\bar{x}(\cdot)$, and $A(\cdot)$. One condition is (42). The other two are the smooth-fit conditions at $\bar{x}(y)$:

$$\frac{y\delta(\bar{x}(y))}{r-\mu} + A(\bar{x}(y))y^{\beta} = \frac{y\delta(z(y))}{r-\mu} + A(z(y))y^{\beta} - c[z(y) - \bar{x}(y)] - k, \tag{44a}$$

$$\frac{y\delta'(\bar{x}(y))}{r-\mu} + A'(\bar{x}(y))y^{\beta} = c. \tag{44b}$$

The continuity of v_y follows automatically. Thanks to (42), we can verify that the continuity of v_y at $\bar{x}(y)$ is equivalent to

$$\frac{\delta(\bar{x}(y))}{r-\mu} + A(\bar{x}(y))\beta y^{\beta-1} = \frac{\delta(z(y))}{r-\mu} + A(z(y))\beta y^{\beta-1}. \tag{44c}$$

We can alternatively obtain (44c) by first differentiating (44a) with respect to y and then employing the relationships established in (42) and (44b). Hence, the three conditions (42), (44a), and (44b) suffice to guarantee that v—as defined by (43)—is continuously differentiable at the boundary. We eliminate $A(\cdot)$ by using (44a) and (44c).

If we multiply (44c) by y and multiply (44a) by β, we eventually obtain (22a). Furthermore, from (42) and (44b) it follows that $A'(z(y)) = G(y, z(y))$ and $A'(\bar{x}(y)) = G(y, \bar{x}(y))$ for G as given by (21b). The equality $A'(\bar{x}(y)) = G(y, \bar{x}(y))$ implies that $A'(x) = G(\bar{y}(x), x)$, so the expression $A'(z(y)) = G(y, z(y))$ can be restated as (22b).

Equations (22a)-(22b) characterize the curves $z(\cdot)$ and $\bar{x}(\cdot)$. We want these curves to be monotone increasing on $[\bar{y}_0, \infty)$ and to satisfy $z(\cdot) > \bar{x}(\cdot)$. If we obtain them, then $A'(x) = G(\bar{y}(x), x)$. To define $A(\cdot)$, we use the boundary condition (9d), which implies $A(\infty) = 0$. Therefore, $A(x)$ is expressed as in (19d).

Appendix E: Proofs of Lemmas 2.A and 2.B

We start by proving **Lemma 2.A**. From (22b) and (21b) it follows that

$$G_y(y,x) = \left[-\beta c + \frac{(\beta-1)y\delta'(x)}{r-\mu}\right] y^{-(\beta+1)}.$$

The function $y \mapsto G(y,x)$ attains a unique minimum at $\bar{y}_U(x)$, at which point

$$G(\bar{y}_U(x),x) = -c(\beta-1)^{-1}\bar{y}_U(x)^{-\beta},$$

an amount which is strictly negative. We know that $y \mapsto G(y,x)$ decreases from ∞ to $G(\bar{y}_U(x),x)$ on $[0,\bar{y}_U(x)]$ and increases from $G(\bar{y}_U(x),x)$ to 0 on $[\bar{y}_U(x),\infty)$. Also, $y \mapsto G(y,x)$ reaches zero at a unique $\bar{y}_L(x) < \bar{y}_U(x)$. In short: $Y(\cdot,x)$ is well-defined on $[\bar{y}_L(x),\infty)$; and it satisfies $Y(\bar{y}_L(x),x) = \infty$, $Y(\bar{y}_U(x),x) = \bar{y}_U(x)$, and $Y(y,x) \lesseqqgtr \bar{y}_U(x)$ if and only if $y \gtreqqless \bar{y}_U(x)$—as stated in Lemma 2.A.

We now turn to **Lemma 2.B**. Note that

$$F_x(y,x) = \frac{(\beta-1)y}{r-\mu}\delta'(x) - \beta c \quad \text{and} \quad F_{xx}(y,x) = \frac{(\beta-1)y}{r-\mu}\delta''(x) < 0.$$

Thus $x \mapsto F(y,x)$ is concave, and its maximum is attained at $\bar{x}_U(y)$ as given by (14c). By (21a),

$$\frac{\mathrm{d}F}{\mathrm{d}y}\Big(y,\bar{x}_U(y)\Big) = F_y(y,\bar{x}_U(y)) = \frac{\beta-1}{r-\mu}\delta'(\bar{x}_U(y)) > 0.$$

Moreover, $F(0,x) = 0$ and $F(\infty,x) = \infty$ for all x. Hence $y \mapsto F(y,\bar{x}_U(y))$ is increasing from 0 to ∞. There exists a unique $y^\star$ such that (24) holds. For $y > y^\star$, we have $F(y,\bar{x}_U(y)) > \beta k$. Since $F_x(y,\infty) = -\beta c$, it follows that $F(y,\infty) = -\infty$. Therefore, $x \mapsto F(y,x)$ decreases on $[\bar{x}_U(y),\infty)$ from $F(y,\bar{x}_U(y)) > \beta k$ to $-\infty$. Hence there exists a unique $x_1(y)$ such that (24) holds. Clearly, $x_1(y^\star) = \bar{x}_U(y^\star)$. In summary, $X(y,\cdot) : [\bar{x}_U(y), x_1(y)] \to [0,\bar{x}_U(y)]$ is well-defined for $y \in [y^\star,\infty)$, as stated in Lemma 2.B.

Appendix F: Restatement of Proposition 3

Condition (22a) implies that, for all $y \geq y^*$,

$$\bar{x}(y) = X(y,z(y)) \quad \text{and} \quad 0 \leq \bar{x}(y) \leq \bar{x}_U(y) \leq z(y) \leq x_1(y); \tag{45}$$

furthermore, $\bar{x}(y^*) = 0$ and $z(y^*) = x_1(y^*) = \bar{x}_U(y^*)$. If $z(y) = x_1(y)$ then $\bar{x}(y) = 0$, and if $z(y) = \bar{x}_U(y)$ then $\bar{x}(y) = X(y,\bar{x}_U(y))$. Given that $Xt(y,\bar{x}_U(y)) \leq \bar{x}_U(y)$, we have the inequalities

$$0 \leq \bar{x}(y) \leq X(y,\bar{x}_U(y)) \leq \bar{x}_U(y) \leq z(y) \leq x_1(y). \tag{46}$$

Condition (22b) implies that $\bar{y}(z(y)) = Y(y,z(y))$, which holds only if $y \geq \bar{y}_L(z(y))$. That inequality is equivalent to $z(y) \leq \bar{x}_L(y)$, where $\bar{x}_L(y) = (\delta')^{-1}(c(r-\mu)/y)$ is the inverse of $\bar{y}_L(x)$. We can write $\bar{y}(z(y)) = Y(y,z(y))$ as $z(y) = \bar{x}(Y(y,z(y)))$. It follows the two equalities and the inequalities in (25). Besides, $\bar{x}_U(y) \leq \bar{x}_L(y)$ because $\delta'(\bar{x}_U(y)) = C(r-\mu)/y \geq c(r-\mu)/y = \delta'(\bar{x}_L(y))$. Finally, $\bar{x}(y^\star) = 0$ and $z(y^\star) = \bar{x}_U(y^\star)$.

Appendix G: Proof of Proposition 4

The proof is technical and will require several steps, which are articulated as lemmas.

LEMMA 7. *In the special case of* (26), *functions X and Y are defined explicitly on* $[y^\star,\infty)\times[\bar{x}_U(y),\min\{\bar{x}_L(y),x_1(y)\}]$ *by*

$$X(y,x)^{1/2}=\frac{y-\sqrt{[3c(r-\mu)\sqrt{x}-y]^2+9kc(r-\mu)^2}}{3c(r-\mu)}, \tag{47a}$$

$$Y(y,x)=\frac{y}{2}\left[-1+\sqrt{\frac{y+6c(r-\mu)\sqrt{x}}{y-2c(r-\mu)\sqrt{x}}}\right]; \tag{47b}$$

here $y^\star$, $\bar{x}_U(\cdot)$, $\bar{x}_L(\cdot)$, *and* $x_1(\cdot)$ *are as given in* (27). *Moreover,* $0\le X(y,x)^{1/2}\le X(y,\bar{x}_U(y))^{1/2}$ *and* $Y(y,x)\ge 3c(r-\mu)\sqrt{x}\ge y$.

Proof. The function X is defined for state values (y,x) such that $\bar{x}_U(y)\le x\le x_1(y)$ and $y\ge y^\star$ (see Lemma 2.B). These inequalities imply that

$$y\ge\sqrt{(3c(r-\mu)\sqrt{x}-y)^2+9kc(r-\mu)^2}.$$

We obtain (47a) from (24b). In addition,

$$0\le X^{1/2}(y,x)\le x_1^{1/2}(y)=X^{1/2}(y,\bar{x}_U(y))=\frac{y-y^*}{3c(r-\mu)}. \tag{48}$$

We turn to Y. Following (21b) and (23), we need to solve for a given $y\ge\bar{y}_L(x):=2c(r-\mu)\sqrt{x}$

$$\left[c-\frac{y}{2(r-\mu)\sqrt{x}}\right]\frac{1}{y^3}=\left[c-\frac{Y(y,x)}{2(r-\mu)\sqrt{x}}\right]\frac{1}{Y(y,x)^3} \tag{49}$$

for the point $Y(y,x)$, thereby obtaining (47b). Since $\bar{y}_U(x)=3c(r-\mu)\sqrt{x}$, we have that $Y(\bar{y}_U(x),x)=\bar{y}_U(x)$ such that the inequality $\bar{y}_L(x)\le y\le\bar{y}_U(x)$ implies the relationship $Y(y,x)\ge Y(\bar{y}_U(x),x)=\bar{y}_U(x)\ge y$. □

For the particular case, we want to find expressions for the functions $\bar{x}(\cdot)$ and $z(\cdot)$ that are more convenient than the general explicit expressions in the "restatement of Proposition 3." We introduce the functions $\varphi(\cdot)$, $\psi(\cdot)$, and $\rho(\cdot)$ in (28a)-(28c) and propose (yet) another statement of Proposition 3 in Lemma 8 following the change of variables.

LEMMA 8. *The (restated) problem in Proposition 3 boils down to finding functions* $\psi:[y^\star,\infty)\to[0,1/2]$ *that solve* (29a)*—provided that* $\rho:[y^\star,\infty)\to(0,1]$ *and* $\varphi:[y^\star,\infty)\to[1,\infty)$ *(as defined by* (28c) *and* (29b)*, respectively) are satisfied.*

Proof. We can restate (25b)–(25c) as the fixed-point equation

$$z(y)=X\big(Y(y,z(y)),z\big(Y(y,z(y))\big)\big). \tag{50}$$

On the definition domain, we have

$$y \geq y^\star \quad \text{and} \quad 0 \leq \bar{x}(y) \leq X(y, \bar{x}_U(y)) \leq \bar{x}_U(y) \leq z(y) \leq \min\{\bar{x}_L(y), x_1(y)\}. \tag{51}$$

Employing the notation from (28a)-(28c), we can restate constraints (51) as $\varphi(y) \geq 1$ and $\psi(y) \in [0,1/2]$; then (29b) follows from (47b). Furthermore, (50) now reads $z(y) = X(y\varphi(y), z(y\varphi(y)))$. By (28b) we have

$$1 + \psi(y) = \varphi(y) - \sqrt{\psi^2(y\varphi(y))\varphi^2(y) + \rho^2(y)},$$

where $\rho : [y^\star, \infty) \to (0,1]$ is as given in (28c). Finally, we obtain (29a) by using (29b). $\square$

Note that if we find $\psi(\cdot)$, then $\varphi(\cdot)$ is explicitly defined by (29b). We want to prove that the set of solutions $\psi : [y^\star, \infty) \to [0,1/2]$ of (29a), such that $\psi'(\cdot) \leq 0$ and $(y\psi(y))' \geq 0$, is nonempty. For that purpose, we define $f : [0,1/2] \to [0,\infty)$ and $g : [0,1/2] \to [1,\infty)$ as

$$f(u) := \sqrt{\frac{1+2u/3}{1-2u}} - 1 - \frac{2}{3}u \quad \text{and} \quad g(u) := -\frac{1}{2} + \frac{3}{2}\sqrt{\frac{1+2u/3}{1-2u}}. \tag{52}$$

We consider functions $\theta : [y^\star, \infty) \to [0,1/2]$ such that $\theta'(\cdot) \leq 0$ and $(y\theta(y))' \geq 0$ and define a map $\mathcal{T} : \theta(\cdot) \mapsto \zeta(\cdot)$, where $\zeta(\cdot)$ solves

$$f(\zeta(y)) = \frac{2}{3}\sqrt{\rho^2(y) + \chi^2(y)\theta^2(y\chi(y))} \tag{53}$$

for $\chi(\cdot) = g(\zeta(\cdot))$. Note that a solution ψ of (29a)–(29b) is a fixed point of the map $\mathcal{T}$.

Lemma 9. *There exists a solution* $\zeta(y)$ *of* (53) *that satisfies* $0 \leq \zeta(y) \leq 1/2$, $\zeta'(y) \leq 0$, *and* $(y\zeta(y))' \geq 0$.

Proof. We shall prove the lemma by constructing a monotone sequence $\{\zeta^k(y), \chi^k(y)\}_k$ whose limit will be a solution of (53). Given a $\zeta^k(y) \in [0,1/2]$, we define $\chi^k(y) = g(\zeta^k(y))$. We start with $\zeta^1(y) = 0$ and define $\zeta^{k+1}(y)$ as a solution of the recurrence relation

$$f(\zeta^{k+1}(y)) = \frac{2}{3}\sqrt{\rho^2(y) + (\chi^k)^2(y)\theta^2(y\chi^k(y))}. \tag{54}$$

Since

$$f'(u) = \frac{2}{3}\left[\frac{2}{(1-2u)\sqrt{(1+2u/3)(1-2u)}} - 1\right] \geq 0, \tag{55}$$

function f is increasing on $[0,1/2]$ from 0 to ∞; it follows that the solution $\zeta^{k+1}(y)$ of (54) is unique and belongs to f's domain of definition $[0,1/2]$.

We now make the induction hypothesis that $\zeta^k(y) \geq \zeta^{k-1}(y)$ for $k \geq 2$; then $\chi^k(y) \geq \chi^{k-1}(y)$ and $y\chi^k(y) \geq y\chi^{k-1}(y)$. Since $y \mapsto y\theta(y)$ increases (by the definition of θ), we have

$$y\chi^k(y)\theta(y\chi^k(y)) \geq y\chi^{k-1}(y)\theta(y\chi^{k-1}(y)).$$

Therefore, $\chi^k(y)\theta(y\chi^k(y)) \geq \chi^{k-1}(y)\theta(y\chi^{k-1}(y))$. It follows that $f(\zeta^{k+1}(y)) \geq f(\zeta^k(y))$, which implies $\zeta^{k+1}(y) \geq \zeta^k(y)$ because $f(\cdot)$ is increasing. This proves the induction.

Because the sequence $\{\zeta^k(y)\}_k$ is monotone and remains in $[0,1/2]$, it necessarily converges to a $\zeta(y) \in [0,1/2]$. It is now immediate that the sequence $\{\chi^k(y)\}_k$ also converges; its limit is $\chi(y) = g(\zeta(y))$. The limits $\big(\zeta(y),\chi(y)\big)$ that we just constructed thus solve (53).

A key remaining issue is to prove that $\zeta'(y) \geq 0$ and $(y\zeta(y))' \geq 0$. By applying the chain rule to (53) the derivative of $\zeta(\cdot)$ satisfies

$$\zeta'(y)f'(\zeta(y)) = \frac{2}{3}\frac{\rho(y)\rho'(y) + \chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}y}\Big(\chi(y)\theta(y\chi(y))\Big)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}.$$

Using $\rho'(y) = -\rho(y)/y$, we obtain

$$y\zeta'(y)f'(\zeta(y)) = \frac{2}{3}\frac{-\rho^2(y) + y\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}y}\Big(\chi(y)\theta(y\chi(y))\Big)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}.$$

By re-arranging terms,

$$\begin{aligned} y\zeta'(y)f'(\zeta(y)) &= -\frac{2}{3}\frac{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} \\ &\quad + \frac{2}{3}\frac{\chi(y)\theta(y\chi(y))\Big[\chi(y)\theta(y\chi(y)) + y\frac{\mathrm{d}}{\mathrm{d}y}\Big(\chi(y)\theta(y\chi(y))\Big)\Big]}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} \\ &= -\frac{2}{3}\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))} + \frac{2}{3}\frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}y}\Big(y\chi(y)\theta(y\chi(y))\Big)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}. \end{aligned}$$

Recalling the expression for $\zeta(\cdot)$ in (53) and the definition of f in (52), the last equation becomes

$$y\zeta'(y)f'(\zeta(y)) = \frac{2}{3}\zeta(y) + 1 - \sqrt{\frac{1+2\zeta(y)/3}{1-2\zeta(y)}} + \frac{2}{3}\frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}y}\Big(y\chi(y)\theta(y\chi(y))\Big)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}. \tag{56}$$

Besides, by (55), we have

$$\zeta(y)f'(\zeta(y)) = -\frac{2}{3}\zeta(y) + \frac{4}{3}\frac{\zeta(y)}{\sqrt{(1+2\zeta(y)/3)(1-2\zeta(y))^3}}. \tag{57}$$

Summing (56) and (57) yields

$$\begin{aligned} \Big[\zeta(y) + y\zeta'(y)\Big]f'(\zeta(y)) &= 1 - \sqrt{\frac{1+2\zeta(y)/3}{1-2\zeta(y)}} + \frac{4}{3}\frac{\zeta(y)}{\sqrt{(1+2\zeta(y)/3)(1-2\zeta(y))^3}} \\ &\quad + \frac{2}{3}\frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}y}\Big(y\chi(y)\theta(y\chi(y))\Big)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}. \end{aligned} \tag{58}$$

Next,

$$\frac{\mathrm{d}}{\mathrm{d}y}\Big(y\chi(y)\theta(y\chi(y))\Big) = \frac{\mathrm{d}}{\mathrm{d}\eta}\Big(\eta\theta(\eta)\Big)\big|_{\eta=y\chi(y)}\frac{\mathrm{d}}{\mathrm{d}y}(y\chi(y))$$

and

$$\frac{\mathrm{d}}{\mathrm{d}y}(y\chi(y)) = g(\zeta(y)) - \frac{2\zeta(y)}{\sqrt{(1+2\zeta(y)/3)(1-2\zeta(y))^3}} + \frac{2\Big(\zeta(y)+y\zeta'(y)\Big)}{\sqrt{(1+2\zeta(y)/3)(1-2\zeta(y))^3}}.$$

To reduce the notation, we put $\gamma(u) := [(1+2u/3)(1-2u)^3]^{-1/2}$. Then we may write

$$\frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}y}\Big(y\chi(y)\theta(y\chi(y))\Big)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} = \frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} \times\Big\{1-\frac{3}{2}+\frac{3}{2}\sqrt{\frac{1+2\zeta(y)/3}{1-2\zeta(y)}} - 2\zeta(y)\gamma(\zeta(y)) + 2(\zeta(y)+y\zeta'(y))\gamma(\zeta(y))\Big\}.$$

From (58) it follows that

$$\begin{aligned}(\zeta(y)+y\zeta'(y))\Big[2\gamma(\zeta(y))-1-\gamma(\zeta(y))\frac{2\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}\Big]\\ = \Big[\frac{3}{2}-\frac{3}{2}\sqrt{\frac{1+2\zeta(y)/3}{1-2\zeta(y)}}+2\zeta(y)\gamma(\zeta(y))\Big]\Big[1-\frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}\Big]\\ + \frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}.\end{aligned} \tag{59}$$

We use the property $0 \le \frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta)) = \theta(\eta)+\eta\theta'(\eta) \le 1/2$, which implies that

$$0 \le \frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} \le \frac{1}{2}$$

and

$$2\gamma(\zeta(y))\Big[1-\frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}\Big] - 1 \ge \gamma(\zeta(y)) - 1 \ge 0.$$

Next, we have

$$\frac{3}{2}-\frac{3}{2}\sqrt{\frac{1+2\zeta(y)/3}{1-2\zeta(y)}}+2\zeta(y)\gamma(\zeta(y)) = \frac{3}{2}h(\zeta(y))\gamma(\zeta(y)), \tag{60}$$

where

$$h(u) := \sqrt{(1+2u/3)(1-2u)^3} - 1 + 8u/3 + 4u^2/3, \quad u \in [0,1/2].$$

We want to show that the expression in (60) is positive when $\zeta(y) \in [0,1/2]$. We observe that the RHS of (59) is positive and that the coefficient of $\zeta(y)+y\zeta'(y)$ on the LHS of (59) is also positive; therefore, the term $\zeta(y)+y\zeta'(y)$ on the LHS is necessarily positive. To check that (60) is positive, we show that $h(u) \ge 0$ for all $u \in [0,1/2]$. Since $-1+\sqrt{7}/2 < 1/2$ is the positive root of $-1+8u/3+4u^2/3 = 0$, it follows that $-1+8u/3+4u^2/3 > 0$ for $u \in [-1+\sqrt{7}/2, 1/2]$; hence $h(u) \ge 0$ in this interval. Now consider $u \in [0, -1+\sqrt{7}/2]$. Because $1-8u/3-4u^2/3 \ge 0$, we need only show that

$$e(u) := (1+2u/3)(1-2u)^3 - (1-8u/3-4u^2/3)^2 \ge 0, \quad u \in [0, -1+\sqrt{7}/2].$$

But $e'(u) = 64u(1+u)(1-4u)/9$. Since $e(\cdot)$ has a unique root on $(0,1/2]$ that happens to be in $(-1+\sqrt{7}/2, 1/2)$, we have that $e(u) \ge 0$ for $u \in [0, -1+\sqrt{7}/2]$. Therefore, $h(u) \ge 0$ for all $u \in [0,1/2]$.

We now consider $\zeta'(\cdot)$. It follows from the previous discussion that

$$\zeta'(y)[2\gamma(\zeta(y))-1] = -y^{-1}\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))} + \frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}\chi'(y)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} + y^{-1}\frac{\chi^2(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}.$$

Given that $\chi'(y)=2\zeta'(y)\gamma(\zeta(y))$, the preceding display can be written as

$$\begin{aligned}
&\zeta'(y)\Big\{2\gamma(\zeta(y))\Big[1-\frac{\chi(y)\theta(y\chi(y))\frac{\mathrm{d}}{\mathrm{d}\eta}(\eta\theta(\eta))|_{\eta=y\chi(y)}}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}\Big]-1\Big\}\\
&= -y^{-1}\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))} + y^{-1}\frac{\chi^2(y)\theta^2(y\chi(y))}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} + \frac{\chi^3(y)\theta^2(y\chi(y))\theta'(y\chi(y))}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}}\\
&= -y^{-1}\frac{\rho^2(y)}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} + \frac{\chi^3(y)\theta^2(y\chi(y))\theta'(y\chi(y))}{\sqrt{\rho^2(y)+\chi^2(y)\theta^2(y\chi(y))}} \le 0. \qquad (61)
\end{aligned}$$

Because the coefficient for $\zeta'(y)$ is positive, as seen in (59), we have proved that $\zeta'(y)\le 0$. This completes the proof of Lemma 9. □

LEMMA 10. *The set of functions* $\psi:[y^\star,\infty)\to[0,1/2]$ *and* $\varphi:[y^\star,\infty)\to[1,\infty]$ *that satisfy* (29a)–(29b) *and are such that* $\psi(y)\in(0,1/2)$, $\psi'(y)\le 0$, *and* $(y\psi(y))'\ge 0$ *is nonempty; that set has a minimum and a maximum.*

Proof. To prove this result, we first show that the map $\mathcal{T}:\theta(\cdot)\mapsto\zeta(\cdot)$, with $\zeta(y)$ defined as an implicit solution of the equation (53), is monotone increasing—in other words, that

$$\theta(y)\ge\theta^\star(y) \implies \mathcal{T}(\theta)(y)\ge\mathcal{T}(\theta^\star)(y), \quad \forall y\in[y^\star,\infty).$$

We already know from the proof of Lemma 9 that $\zeta(y):=\mathcal{T}(\theta)(y)$ and $\zeta^\star(y):=\mathcal{T}(\theta^\star)(y)$ are the limits of the monotone increasing sequences $\{\zeta^k(y)\}_k$ and $\{(\zeta^\star)^k(y)\}_k$. Hence we shall just focus on showing that $\zeta^k(y)\ge(\zeta^*)^k(y)$. We establish the latter relationship by way of induction.

The induction property holds true for $k=1$ because $\zeta^1(y)=(\zeta^\star)^1(y)=0$ by definition of the sequences. Assume that $\zeta^k(y)\ge(\zeta^\star)^k(y)$ for $k\ge 1$. We now recall the notations $\chi^k(\cdot)=g(\zeta^k(\cdot))$ and $(\chi^\star)^k(\cdot)=g((\zeta^\star)^k(\cdot))$ and observe that $g'(u)\ge 0$ if $u\in[0,1/2]$; we thus have $\chi^k(y)\ge(\chi^\star)^k(y)$. Since $y\mapsto y\theta(y)$ increases (by the definition of $\theta(\cdot)$), it follows that $y\chi^k(y)\theta(y\chi^k(y))\ge y(\chi^\star)^k(y)\theta(y(\chi^\star)^k(y))$. Therefore, by (54), $f(\zeta^{k+1}(y))\ge f((\zeta^\star)^{k+1}(y))$; this inequality implies that $\zeta^{k+1}(y)\ge(\zeta^\star)^{k+1}(y)$ because $f'(\cdot)\ge 0$. Thus we have proved the induction and thereby completed the proof that map $\mathcal{T}$ is monotone increasing.

Next we consider the sequence $\{\zeta_n(y)\}_n$ defined by $\zeta_1(y)=0$ and the recursive equation $\zeta_{n+1}(y)=\mathcal{T}(\zeta_n)(y)$. Clearly $\zeta_2(y)\ge\zeta_1(y)$ and, because $\mathcal{T}$ is monotone increasing, $\zeta_{n+1}(y)\ge\zeta_n(y)$. Moreover, $\zeta_n(y)\in[0,1/2]$, $\zeta_n'(y)\le 0$, and $(y\zeta_n(y))'\ge 0$ as established in Lemma 9. Hence the sequence $\{\zeta_n(y)\}_n$ converges to a $\underline{\zeta}(y)$ that solves the fixed-point equation $\underline{\zeta}(y)=\mathcal{T}(\underline{\zeta}(y))$ and also satisfies $\underline{\zeta}(y)\in[0,1/2]$, $\underline{\zeta}'(y)\le 0$, and $(y\underline{\zeta}(y))'\ge 0$.

We now want to construct the minimum and maximum solutions. If $\zeta(\cdot)$ is another solution of the fixed-point equation, then $\zeta(y) \geq \zeta_1(y)$, where $\zeta_1(y)$ is the first term of the sequence $\{\zeta_n(y)\}_n$ that we just contructed. Because $\mathcal{T}$ is monotone increasing, we immediately obtain that $\zeta(y) \geq \zeta_n(y)$. Passing to the limit for the sequence on the RHS, it follows that $\zeta(y) \geq \underline{\zeta}(y)$—proving that $\underline{\zeta}(y)$ is the minimum solution. Instead of the previous sequence starting at $\zeta_1(y) = 0$, consider a sequence $\{\zeta^n(y)\}_n$ that starts at $\zeta^1(y) = 1/2$. We can prove using similar arguments that this sequence is monotone decreasing and that it converges to a $\bar{\zeta}(y)$; the function $\bar{\zeta}(\cdot)$ defined pointwise in this fashion is the maximum solution. The proof of Lemma 10 is now complete. □

The function $z(\cdot)$ can be expressed explicitly as in (28a) provided we can find functions $\psi(\cdot), \varphi(\cdot)$ and $\rho(\cdot)$ that satisfy the conditions stated in Lemma 8. The explicit expression for $\bar{x}(\cdot)$ in (29c) is a consequence of (25c), (27a), (28c), and (47a). Because the LHS of (29c) must be positive, we have the following inequality:

$$\psi(y) \leq \sqrt{1-\rho^2(y)}. \tag{62}$$

Lemma 11 helps refine our search of the functions $\psi(\cdot)$, $\varphi(\cdot)$ and $\rho(\cdot)$ by leveraging on the condition (62).

Lemma 11. *There exists a unique value $\bar{y}_0 > y^\star$ such that $\psi(y) \lesseqqgtr \sqrt{1-\rho(y)^2}$ for all $y \gtreqqless \bar{y}_0$. Furthermore, $\psi(\infty) = 0$.*

Since $\psi(y) \leq 1/2$, we know that condition (62) is satisfied whenever $\sqrt{1-\rho^2(y)} \geq 1/2$ or $y \geq 2\sqrt{3}(r-\mu)\sqrt{ck}$. Therefore, $y^\star < \bar{y}_0 < 2\sqrt{3}(r-\mu)\sqrt{ck}$ must also hold.

Proof of Lemma 11. Define $\Lambda(y) := \sqrt{y^2 - 9(r-\mu)^2 kc} - y\psi(y)$. Then

$$\Lambda'(y) = y[y^2 - 9(r-\mu)^2 kc]^{-1/2} - (y\psi(y))' \geq 1/2$$

and so $\Lambda(\cdot)$ is monotone increasing. Hence, from the definition of $y^\star$ in (27b), it follows that $\Lambda(y) \leq -y^\star\psi(y^\star) < 0$ for all $y < y^\star$. In addition,

$$\Lambda(y) \geq \sqrt{y^2 - 9(r-\mu)^2 kc} - y/2 \geq y - 3(r-\mu)\sqrt{kc} - y/2 = y/2 - 3(r-\mu)\sqrt{kc} \to \infty$$

as $y \to \infty$. Thus there exists a unique $\bar{y}_0 > y^\star$ such that $\Lambda(\bar{y}_0) = 0$, $\lambda(y) > 0$ if $y > \bar{y}_0$ and $\Lambda(y) < 0$ if $y < y^\star$. As a result, $\psi(y) \lesseqqgtr \sqrt{1-\rho^2(y)}$ if $y \gtreqqless \bar{y}_0$.

Next, we deduce from Lemma 10 that $\psi(\cdot)$ (as given in (28a)) is positive and monotone decreasing; hence it converges to a $\psi_\infty \in \mathbb{R}_+$ as $y \to \infty$. From (29a) and the definition of f in (52), we have

$$f(\psi_\infty) = \lim_{y\to\infty} \frac{2}{3}\sqrt{\rho(y)^2 + \varphi(y)^2\psi(y\varphi(y))^2}. \tag{63}$$

By the definition of $\rho(\cdot)$ in (28c), $\rho(y) \to 0$ as $y \to \infty$. Also, $y \mapsto y\varphi(y)$ is monotone increasing by Lemma 10; therefore, $y\varphi(y) \to \infty$ as $y \to \infty$. Moreover, from the definition of g in (52) and equation (29b), $\varphi(y) = g(\psi(y))$. Equation (63) thus simplifies to $f(\psi_\infty) = \frac{2}{3}g(\psi_\infty)\psi_\infty$. Using the functions $f(\cdot)$ and $g(\cdot)$ defined in (52) and then re-arranging yields

$$\sqrt{\frac{1+2\psi_\infty/3}{1-2\psi_\infty}} = \frac{1+\psi_\infty/3}{1-\psi_\infty}.$$

From this equality we obtain $\psi_\infty(1+\psi_\infty) = 0$, which implies that $\psi_\infty = 0$ because $\psi_\infty \geq 0$. This completes the proof of Lemma 11. □

All that remains for the proof of Proposition 4 is to establish the monotonicity of $z(\cdot)$ and $\bar{x}(\cdot)$. Since $y \mapsto y + y\psi(y)$ increases, we obtain immediately that $z(\cdot)$ is monotone increasing. Next, $y \mapsto \psi^2(y) + \rho^2(y)$ decreases and so we have that $y \mapsto 1 - \sqrt{\psi^2(y) + \rho^2(y)}$ is both monotone increasing and positive. Hence $y \mapsto y(1 - \sqrt{\psi^2(y) + \rho^2(y)})$ is monotone increasing, from which it follows that $\bar{x}(\cdot)$ is monotone increasing.

Property (30) obtains because the functions $\psi(\cdot)$ and $\rho(\cdot)$ vanish as $y \to \infty$.

Appendix H: Proof of Proposition 5

Here we consider only the particular case in (26). We prove the proposition via a series of lemmas, addressing in turn the free boundary problem, the quasi-variational inequality, and the verification theorem.

Free-boundary problem

Lemma 12. *The function v given in* (31a) *satisfies the regularity* (8) *and solves the FBP in* (9a)-(9d).

Proof. The function $A(\cdot)$ is defined in general by (19d). We can use the particular expression for G to obtain (31b). Since $\bar{y}(x) \geq 2c(r-\mu)x^{1/2}$, it follows that $A(x) \geq 0$. The function v as defined by (31a) is continuous. From (31b) at $\bar{x}(y)$, we have

$$\frac{1}{2}\frac{y\bar{x}(y)^{-1/2}}{r-\mu} + A'(\bar{x}(y))y^3 = c. \tag{64}$$

We observe the continuity of $x \mapsto v_x(y,x)$ at the boundary $x = \bar{x}(y)$, which implies that

$$v_y(y,x) = \frac{\bar{x}(y)^{1/2}}{r-\mu} + 3y^2 A(\bar{x}(y))$$

for $x < \bar{x}(y)$. Hence $y \mapsto v_y(y,x)$ is also continuous at $\bar{y}(x)$, which proves v is C^1. If there is a set of points at which a second-order derivative is not defined, this set has measure 0. □

Lemma 13. *The function v given in* (31a) *majorizes the perpetuity value of capacity, v_L.*

Proof. By (31a), $v(y,x) \geq v_L(y,x)$ holds trivially when $y \leq \bar{y}(x)$. When $y \geq \bar{y}(x)$, $v(y,x) \geq c[x - \bar{x}(y)] + y\sqrt{\bar{x}(y)}/[r-\mu]$. Hence $v(y,x) \geq v_L(y,x)$ holds if we can prove that

$$c[x - \bar{x}(y)] + \frac{y\sqrt{\bar{x}(y)}}{r-\mu} \geq \frac{y\sqrt{x}}{r-\mu} \tag{65}$$

when $\bar{x}(y) \geq x$—or, equivalently if we prove that $x \mapsto y\sqrt{x}/[r-\mu] - cx$ is monotone increasing on $(0, \bar{x}(y)]$ or that $\frac{1}{2}\frac{yx^{-1/2}}{r-\mu} \geq c$ when $y \geq \bar{y}(x)$. The latter property holds because $\frac{1}{2}\frac{\bar{y}(x)x^{-1/2}}{r-\mu} \geq c$, which proves that the inequality $v(y,x) \geq v_L(y,x)$ in the case where $y \geq \bar{y}(x)$, as desired. □

LEMMA 14. *The function v given in* (31a) *minorizes the function v_U given in* (14).

Proof. We start by demonstrating the property $A(\cdot) \leq A_U(\cdot)$. Using (31b) and (14b), we may write

$$A_U(x) = -\int_x^\infty \left(c - \frac{1}{2}\frac{\bar{y}_U(\xi)\xi^{-1/2}}{r-\mu}\right)\bar{y}_U(\xi)^{-3}\,\mathrm{d}\xi;$$

proving this property thus amounts to showing that $(c - \frac{1}{2}\bar{y}(x)x^{-1/2}/[r-\mu])\bar{y}(x)^{-3} \geq (c - \frac{1}{2}\bar{y}_U(x)x^{-1/2}/[r-\mu])\bar{y}_U(x)^{-3}$. Because $\bar{y}(x) \geq \bar{y}_U(x)$, the property is established if we can show that $y \mapsto (c - \frac{1}{2}yx^{-1/2}/[r-\mu])y^{-3}$ is monotone increasing on $[\bar{y}_U(x), \infty)$. The monotonicity property is immediate once we observe that the derivative in y remains positive on this interval.

If we now recall the expressions for $v(y,x)$ when $y \leq \bar{y}(x)$ and for $v_U(y,x)$ when $y \leq \bar{y}_U(x)$ and also note that $\bar{y}_U(x) \leq \bar{y}(x)$, then we obtain immediately from the property $A(\cdot) \leq A_U(\cdot)$ that $v(y,x) \leq v_U(y,x)$ when $y \leq \bar{y}_U(x)$. Next we show that $v(y,x) \leq v_U(y,x)$ when $y \geq \bar{y}(x)$. Given (31a) for $y \geq \bar{y}(x)$ and (14a) for $y \geq \bar{y}_U(x)$, it remains only to check that

$$-c\bar{x}(y) + \frac{y\sqrt{\bar{x}(y)}}{r-\mu} + A(\bar{x}(y))y^3 \leq -c\bar{x}_U(y) + \frac{y\sqrt{\bar{x}_U(y)}}{r-\mu} + A_U(\bar{x}_U(y))y^3. \tag{66}$$

From (44a) this above can be re-written as

$$-cz(y) + \frac{y\sqrt{z(y)}}{r-\mu} + A(z(y))y^3 - k \leq -c\bar{x}_U(y) + \frac{y\sqrt{\bar{x}_U(y)}}{r-\mu} + A_U(\bar{x}_U(y))y^3. \tag{67}$$

It therefore suffices to show that

$$-cz(y) + \frac{y\sqrt{z(y)}}{r-\mu} + A_U(z(y))y^3 < -c\bar{x}_U(y) + \frac{y\sqrt{\bar{x}_U(y)}}{r-\mu} + A_U(\bar{x}_U(y))y^3. \tag{68}$$

Given that $z(y) \geq \bar{x}_U(y)$, property (68) will be satisfied if we can show that $x \mapsto -cx + y\sqrt{x}/[r-\mu] + A_U(x)y^3$ is decreasing on $[\bar{x}_U(y), \infty)$. In that case,

$$-c + \frac{y}{2\sqrt{x}(r-\mu)} + A_U'(x)y^3 \leq 0 \quad \forall x \geq \bar{x}_U(y).$$

From the value of $A_U'(x)$ we conclude that this amounts to proving

$$\left(c - \frac{1}{2}\frac{\bar{y}_U(x)x^{-1/2}}{r-\mu}\right)\bar{y}_U(x)^{-3} \leq \left(c - \frac{1}{2}\frac{yx^{-1/2}}{r-\mu}\right)y^{-3} \quad \forall y \leq \bar{y}_U(x),$$

which is true if $y \mapsto \left(c - \frac{1}{2}\frac{yx^{-1/2}}{r-\mu}\right)y^{-3}$ is decreasing on $[0, \bar{y}_U(x)]$. After computing the derivative in y, we can easily see that the function is decreasing on that interval. Thus we have shown that $v(y,x) \le v_U(y,x)$ when $y \le \bar{y}_U(x)$ or $y \ge \bar{y}(x)$. It remains to consider the interval $[\bar{y}_U(x), \bar{y}(x)]$. On this interval we have $\mathcal{L}v(y,x) = y\sqrt{x}$ and $\mathcal{L}v_U(y,x) \ge y\sqrt{x}$. Therefore,

$$\begin{aligned}
(v_U - v)(\bar{y}_U(x), x) &\ge 0, \\
\mathcal{L}(v_U - v)(y,x) &\ge 0, \quad y \in (\bar{y}_U(x), \bar{y}(x)), \\
(v_U - v)(\bar{y}(x), x) &\ge 0,
\end{aligned}$$

from which it immediately follows that $(v_U - v)(y,x) \ge 0$ on $[\bar{y}_U(x), \bar{y}(x)]$. The next step is to write $v(y,x)$ in (31a) as in (43) for $y \ge \bar{y}(x)$:

$$v(y,x) = cx + v(y, z(y)) - cz(y) - k. \tag{69}$$

Recall that $\bar{y}(x) \ge \bar{y}(0) = \bar{y}_0$. We have to prove that $z(y)$ is the maximum point of $z \mapsto v(y,z) - cz$ on (x, ∞) for any $x \le \bar{x}(y)$. Since the expression for $v(y,x) - cx$ does not depend on x for $x \le \bar{x}(y)$, it follows that $\max_{z > x}\{v(y,z) - cz\} = \max_{z \ge \bar{x}(y)}\{v(y,z) - cz\}$ for all $x \le \bar{x}(y)$. Now we must check that

$$v(y, z(y)) - cz(y) = \max_{z \ge \bar{x}(y)} \{v(y,z) - cz\}. \tag{70}$$

For $x \ge \bar{x}(y)$ we have $v(y,x) = y\sqrt{x}/[r-\mu] + A(x)y^3$, and we know from (42) that $v_x(y, z(y)) = c$. Property (70) is now a consequence of

$$v_x(y,x) \begin{cases} < c & \forall x > z(y), \\ > c & \forall x \in (\bar{x}(y), z(y)). \end{cases} \tag{71}$$

If we introduce the inverse $\bar{z}(\cdot)$ of $z(\cdot)$ defined on $[z(\bar{y}), \infty)$, then (71) translates into $x \ge z(\bar{y}_0)$ and

$$v_x(y,x) \begin{cases} < c & \forall y \in (\bar{y}_0, \bar{z}(x)), \\ > c & \forall y \in (\bar{z}(x), \bar{y}(x)). \end{cases} \tag{72}$$

To check (72), we compute v_x using (31b). After re-arranging terms, we obtain

$$(v_x(y,x) - c)y^{-3} = \left[\left(c - \frac{1}{2}\frac{\bar{y}(x)x^{-1/2}}{r-\mu}\right)\bar{y}(x)^{-3} - \left(c - \frac{1}{2}\frac{yx^{-1/2}}{r-\mu}\right)y^{-3}\right] \tag{73}$$

For $x \ge z(\bar{y}_0)$, we have $\bar{y}_0 \le \bar{z}(x) \le \bar{y}_U(x) \le \bar{y}(x)$. From (73) we see easily that $y \mapsto (v_x(y,x) - c)y^{-3}$ increases on $[\bar{y}, \bar{y}_U(x)]$ and decreases on $[\bar{y}_U(x), \bar{y}(x)]$; this function attains at $\bar{y}_U(x)$ a maximum that is positive and is zero at $\bar{y}(x)$. That zero value is necessarily unique and is attained in $[\bar{y}_0, \bar{y}_U(x)]$—in fact, at $\bar{z}(x)$. Hence both (72) and (71) hold. Therefore, (31a) can be interpreted as (9a). □

Quasi-variational inequality

LEMMA 15. *The function v given in* (31a) *satisfies the regularity* (8) *and also solves the QVI* (6a)-(6e).

Proof Given the previous results, we need only prove that the function v satisfies the following inequalities:

$$\mathcal{L}v(y,x) - y\sqrt{x} \geq 0, \qquad y > \bar{y}(x), \tag{74a}$$

$$v(y,x) \geq \mathcal{M}v(y,x), \qquad y \leq \bar{y}(x). \tag{74b}$$

We first prove (74a). When $y > \bar{y}(x)$, we have by (31a)

$$\begin{aligned} v(y,x) &= \frac{y\sqrt{\bar{x}(y)}}{r-\mu} + A(\bar{x}(y))y^3 - c[\bar{x}(y) - x], \\ v_y(y,x) &= \frac{\sqrt{\bar{x}(y)}}{r-\mu} + 3y^2 A(\bar{x}(y)) + \bar{x}'(y)\left[\frac{1}{2}\frac{y\bar{x}(y)^{-1/2}}{r-\mu} + A'(\bar{x}(y))y^3 - c\right]. \end{aligned}$$

It now follows from (64) that

$$\begin{aligned} v_y(y,x) &= \frac{\sqrt{\bar{x}(y)}}{r-\mu} + 3y^2 A(\bar{x}(y)), \\ v_{yy}(y,x) &= \frac{1}{2}\frac{\bar{x}(y)^{-1/2}}{r-\mu} + 6yA(\bar{x}(y)) + 3y^2 A'(\bar{x}(y))\,\bar{x}'(y). \end{aligned}$$

By (26) we can write

$$\mathcal{L}v(y,x) - y\sqrt{x} = y\big(\sqrt{\bar{x}(y)} - \sqrt{x}\big) + rc[x - \bar{x}(y)] - \frac{r-3\mu}{2}\bar{x}'(y)\frac{c}{y}\left(1 - \frac{\sqrt{\bar{x}_U(y)}}{\sqrt{\bar{x}(y)}}\right). \tag{75}$$

Since $\bar{x}_U(y) > \bar{x}(y)$, the last term is positive. Let us check that $y\big[\sqrt{\bar{x}(y)} - \sqrt{x}\big] + rc(x - \bar{x}(y)) \geq 0$ if $\bar{x}(y) > x$. It suffices to show that $y - rc\big(\sqrt{x} + \sqrt{\bar{x}(y)}\big) \geq 0$ for $x \leq \bar{x}(y)$ or $y - 2rc\sqrt{\bar{x}(y)} \geq 0$. This is equivalent to $\sqrt{\bar{x}(y)}/\sqrt{\bar{x}_U(y)} \leq \frac{3}{2}(r-\mu)/2$. But the LHS is less than 1, and $1 \leq \frac{3}{2}(r-\mu)/\mu$ by (26). Thus we have verified (74a).

We now verify (74b). For $y \leq \bar{y}(x)$ we have $\bar{x}(y) \leq x$. Therefore,

$$\mathcal{M}v(y,x) = \sup_{z>x}\{v(y,z) - cz\} - k + cx = v(y,\bar{x}(y)) - c[\bar{x}(y) - x]$$

if $\bar{x}(y) \leq x \leq z(y)$. So we must check that

$$v(y,x) - cx \geq \begin{cases} v(y,\bar{x}(y)) - c\bar{x}(y), & x \in [\bar{x}(y), z(y)], \\ \sup_{z>x}\{v(y,z) - cz\} - k, & x > z(y). \end{cases} \tag{76}$$

On the one hand, we saw in (71) that $x \mapsto v(y,x) - cx$ is decreasing for $x > z(y)$. Hence $\sup_{z>x}\{v(y,z) - cz\} = v(y,x) - cx$ if $x > z(y)$, from which it follows that (74b) is trivial. On the other hand, $x \mapsto v(y,x) - cx$ is increasing on $(\bar{x}(y), z(y))$, which implies the first inequality (76). This completes the proof of Lemma 15.

Verification theorem

It now remains only to prove that the regular solution of the QVI v, as given in (31a), coincides with the value function in (1). First, as an upper solution of (1), v majorizes the value function (see Lemma 5 for the general case). Our next lemma establishes that v also minorizes the value function, which allows us to conclude that the function v given in (31a) coincides with the value function; v is also the smallest solution as well as the smallest upper solution of (6a)-(6e).

LEMMA 16. *The function v given in* (31a) *minorizes the value function in* (1).

Proof. We construct $\hat{\nu}$ as in Section 5.2. First we prove that $\hat{\tau}_n \uparrow \infty$ a.s. as $n \uparrow \infty$. Put $\eta(y) := \bar{y}(z(y))$ and note that $\eta(y) > 1$ for any $y < \infty$. We shall apply Itô's lemma between two immediately subsequent intervention times to obtain

$$\eta(Y_{\hat{\tau}_n})^3 e^{-r\hat{\tau}_{n+1}} - Y_{\hat{\tau}_n}^3 e^{-r\hat{\tau}_n} = 3\sigma \int_{\hat{\tau}_n}^{\hat{\tau}_{n+1}} Y_t^3 e^{-rt}\,\mathrm{d}W_t,$$
$$e^{-r\hat{\tau}_{n+1}} - \left(\frac{Y_{\hat{\tau}_n}}{\eta(Y_{\hat{\tau}_n})}\right)^3 e^{-r\hat{\tau}_n} = \frac{3\sigma}{\eta(Y_{\hat{\tau}_n})^3} \int_{\hat{\tau}_n}^{\hat{\tau}_{n+1}} Y_t^3 e^{-rt}\,\mathrm{d}W_t,$$

from which it follows that $\mathbb{E}[e^{-\hat{\tau}_{n+1}}] = \mathbb{E}[(Y_{\hat{\tau}_n}/\eta(Y_{\hat{\tau}_n}))^3 e^{-r\hat{\tau}_n}]$ in light of the standard properties of stochastic integrals. Since $\{\hat{\tau}_n\}_{n\geq 1}$ is an increasing sequence, we have $\hat{\tau}_n \uparrow \hat{\tau}$ a.s. By Lebesgue's theorem, we can pass to the limit and obtain $\mathbb{E}\big[e^{-r\hat{\tau}}\big(1-(Y_{\hat{\tau}}/\eta(Y_{\hat{\tau}}))^3\big)\big] = 0$. Because (i) $\bar{y}(\cdot)$ is monotone increasing, (ii) $z(\cdot) > \bar{x}(\cdot)$, and (iii) $\bar{x}(\cdot)$ is the inverse of $\bar{y}(\cdot)$, we have $\eta(y) := \bar{y}(z(y)) > \bar{y}(\bar{x}(y)) = y$ for all $y > 0$. Also, since $1-(y/\eta(y))^3 > 0$ for all $y > 0$, it follows that $e^{-r\hat{\tau}}[1-(Y_{\hat{\tau}}/\eta(Y_{\hat{\tau}}))^3] = 0$ a.s. and $\hat{\tau}_n \to \infty$ a.s. as $n \to \infty$.

We now proceed as in the proof of Proposition 2 and write

$$v(y,x) = \mathbb{E}[v(Y_{\hat{\tau}_N \wedge T}, X_{\hat{\tau}_N \wedge T}) e^{-r(\hat{\tau}_N \wedge T)}] + \mathbb{E}\Big[\int_0^{\hat{\tau}_N \wedge T} Y_t \sqrt{\hat{X}_t} e^{-rt}\,\mathrm{d}t\Big] - \mathbb{E}\Big[\sum_{n=1}^{N} (k + c\hat{\xi}_n) e^{-r\hat{\tau}_n} \mathbb{I}_{\{\hat{\tau}_n < T\}}\Big]. \tag{77}$$

For $\hat{\tau}_n \wedge T < t < \hat{\tau}_{n+1} \wedge T$, we have $\hat{X}_t = X_{\hat{\tau}_n} = z(Y_{\hat{\tau}_n}) = z(Y_{\hat{\tau}_n \wedge T})$. Recalling that $z^{1/2}(y) \leq y/[2c(r-\mu)]$, we obtain $\hat{X}_t^{1/2} \leq Y_{\hat{\tau}_n}/[2c(r-\mu)] \leq [\sup_{0\leq s\leq t} Y_s]/[2c(r-\mu)]$. This inequality holds for any t, so

$$\mathbb{E}\Big[\int_0^{\hat{\tau}_N \wedge T} Y_t \sqrt{\hat{X}_t} e^{-rt}\,\mathrm{d}t\Big] \leq \frac{1}{2c(r-\mu)} \mathbb{E}\Big[\int_0^T Y_t \sup_{0\leq s\leq t} Y_s e^{-rt}\,\mathrm{d}t\Big],$$
$$\leq \frac{1}{2c(r-\mu)} \int_0^T \mathbb{E}\Big[\sup_{0\leq s\leq t} Y_s^2 e^{-rt}\Big]\,\mathrm{d}t.$$

Let us now check that

$$\int_0^\infty \mathbb{E}\Big[\sup_{0\leq s\leq t} Y_s^2 e^{-rt}\,\mathrm{d}t\Big] < \infty. \tag{78}$$

We have $\mathbb{E}\Big[Y_t^2 e^{-rt}\Big] = y^2 e^{(\mu-2r/3)t}$ and, from standard estimates on stochastic integrals,

$$\mathbb{E}\Big[\sup_{0\le s\le t} Y_s^2 e^{-rt}\Big] \le ce^{-rt}\Big(y^2 + (t+1)\int_0^t \mathbb{E}[Y_s^2]\,\mathrm{d}s\Big).$$

Hence (78) follows because $-2r+3\mu<0$. We next consider the term

$$\mathbb{E}\big[v(Y_{\hat\tau_N\wedge T}, X_{\hat\tau_N\wedge T})e^{-r(\hat\tau_N\wedge T)}\big] \le \mathbb{E}\big[v_U(Y_{\hat\tau_N\wedge T}, X_{\hat\tau_N\wedge T})e^{-r(\hat\tau_N\wedge T)}\big].$$

From (14a) we obtain $v_U(y,x) \le \frac{yx^{1/2}}{r-\mu} + \frac{4}{9}\frac{y^2}{c(r-\mu)^2}$. Therefore,

$$\begin{aligned}\mathbb{E}\big[v(Y_{\hat\tau_N\wedge T}, X_{\hat\tau_N\wedge T})e^{-r(\hat\tau_N\wedge T)}\big] \le{}& \frac{4}{9c(r-\mu)^2}\mathbb{E}\big[Y^2_{\hat\tau_N\wedge T}e^{-r(\hat\tau_N\wedge T)}\big]\\ &+\frac{1}{r-\mu}\mathbb{E}\big[Y_{\hat\tau_N\wedge T}X^{1/2}_{\hat\tau_N\wedge T}e^{-r(\hat\tau_N\wedge T)}\big].\end{aligned}$$

It now follows from the preceding estimate of $\hat X_t^{1/2}$ that

$$\mathbb{E}\big[v(Y_{\hat\tau_N\wedge T}, X_{\hat\tau_N\wedge T})e^{-r(\hat\tau_N\wedge T)}\big] \le \frac{17}{18c(r-\mu)^2}\mathbb{E}\Big[\sup_{0\le t\le T} Y_t^2 e^{-r(\hat\tau_N\wedge T)}\Big].$$

Therefore, by (77),

$$\begin{aligned}v(y,x) \le{}& \frac{17}{18c(r-\mu)^2}\mathbb{E}\Big[\sup_{0\le t\le T} Y_t^2 e^{-r(\hat\tau_N\wedge T)}\Big]\\ &+\mathbb{E}\Big[\int_0^\infty Y_t\sqrt{\hat X_t}e^{-rt}\,\mathrm{d}t\Big] - \mathbb{E}\Big[\sum_{n=1}^N (k+c\hat\xi_n)e^{-r\hat\tau_n}\mathbb{1}_{\{\hat\tau_n<T\}}\Big].\end{aligned}$$

If we now let $N\to\infty$, then

$$\begin{aligned}v(y,x) \le{}& \frac{17}{18c(r-\mu)^2}\mathbb{E}\Big[\sup_{0\le t\le T} Y_t^2 e^{-rT}\Big]\\ &+\mathbb{E}\Big[\int_0^\infty Y_t\sqrt{\hat X_t}e^{-rt}\,\mathrm{d}t\Big] - \mathbb{E}\Big[\sum_{n=1}^\infty (k+c\hat\xi_n)e^{-r\hat\tau_n}\mathbb{1}_{\{\hat\tau_n<T\}}\Big].\end{aligned}$$

Finally, from (78) and the continuity of $t\mapsto \mathbb{E}[\sup_{0\le s\le t} Y_s^2 e^{-rt}]$, we can state that

$$\mathbb{E}\Big[\sup_{0\le t\le T} Y_t^2 e^{-rT}\Big] \to 0$$

as $T\to\infty$. From the last inequality it follows that $\mathbb{E}\big[\sum_{n=1}^\infty (k+c\hat\xi_n)e^{-r\hat\tau_n}\big]<\infty$. Hence $\hat\nu$ is admissible and so $v(y,x)\le J_{yx}(\hat\nu)$, which completes the proof of the lemma. □

Short author biographies

Alain Bensoussan

Dr. Alain Bensoussan is Lars Magnus Ericsson Chair at the Naveen Jindal School of Management, University of Texas at Dallas and Director of the International Center for Decision and Risk Analysis which develops risk management research as it pertains to large-investment industrial projects that involve new technologies, applications and markets. He is also Chair Professor at City University of Hong Kong and Professor Emeritus at the University of Paris Dauphine.

He is the co-editor of *Risk and Decision Analysis*, a IEEE, SIAM and AMS fellow and a recipient of the Von Humboldt Research Award and of the W.T. & Ida Reid Prize for his contributions to the fields of differential equations and control theory. He is a member of the French Academy of Sciences.

Benoît Chevalier-Roignant

Dr. Benoît Chevalier-Roignant is at King's Business School, King's College London. His research is at the intersection of corporate finance and industrial organization, investigating how market uncertainty and strategic reactions affect a firm's investment behaviours. Benoît has done research on capacity expansion decisions, innovation investments, production flexibility and market entry in uncertain competitive environments. He co-authored "Competitive Strategy: Options and Games" (with Lenos Trigeorgis) published at *The MIT Press.*

Beside his academic track record, Benoît has year-long professional experience in professional services firms, with an extensive expertise on corporate transactions.