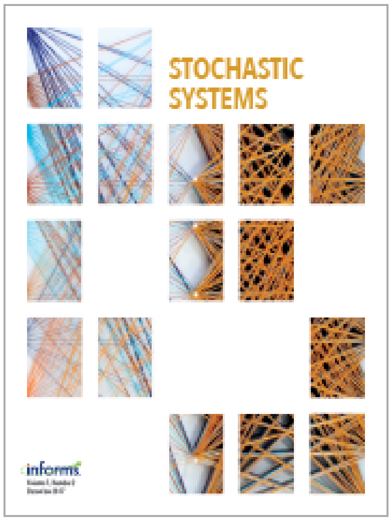


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Average Cost Brownian Drift Control with Proportional Changeover Costs

John H. Vande Vate

To cite this article:
John H. Vande Vate (2021) Average Cost Brownian Drift Control with Proportional Changeover Costs. Stochastic Systems 11(13):218-263. https://doi.org/10.1287/stsy.2021.0071

informs
http://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 11, No. 3, September 2021, pp. 218–263
ISSN 1946-5238 (online)

Average Cost Brownian Drift Control with Proportional Changeover Costs

John H. Vande Vate[a]

[a] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205

Contact: jv3@gatech.edu, https://orcid.org/0000-0001-8629-5871 (JHVV)

Received: March 29, 2020
Revised: October 2, 2020
Accepted: December 3, 2020
Published Online in Articles in Advance: September 13, 2021

https://doi.org/10.1287/stsy.2021.0071

Abstract. This paper considers the problem of optimally controlling the drift of a Brownian motion with a finite set of possible drift rates so as to minimize the long-run average cost, consisting of fixed costs for changing the drift rate, processing costs for maintaining the drift rate, holding costs on the state of the process, and costs for instantaneous controls to keep the process within a prescribed range. We show that, under mild assumptions on the processing costs and the fixed costs for changing the drift rate, there is a strongly ordered optimal policy, that is, an optimal policy that limits the use of each drift rate to a single interval; when the process reaches the upper limit of that interval, the policy either changes to the next lower drift rate deterministically or resorts to instantaneous controls to keep the process within the prescribed range, and when the process reaches the lower limit of the interval, the policy either changes to the next higher drift rate deterministically or again resorts to instantaneous controls to keep the process within the prescribed range. We prove the optimality of such a policy by constructing smooth relative value functions satisfying the associated simplified optimality criteria. This paper shows that, under the proportional changeover cost assumption, each drift rate is active in at most one contiguous range and that the transitions between drift rates are strongly ordered. The results reduce the complexity of proving the optimality of such a policy by proving the existence of optimal relative value functions that constitute a nondecreasing sequence of functions. As a consequence, the constructive arguments lead to a practical procedure for solving the problem that is tens of thousands of times faster than previously reported methods.

Funding: This work has been supported in part by the National Science Foundation [Grant 0825840].

Keywords: diffusion process • drift control • algorithm

1. Introduction

Consider the problem of managing capacity in an on-demand environment, such as an internet service center, build-to-order manufacturing facility, or order-fulfillment center, modeled as a Brownian drift control problem with the objective of minimizing the long-term average cost. The controller can, at some cost, shift the processing rate among a finite set of alternatives by adding or removing capacity. The controller incurs a cost for capacity per unit time and a delay cost that reflects the opportunity cost of revenue waiting to be recognized or the customer service impacts of delaying the completion of orders. Furthermore, the controller incurs a cost per unit to reject orders or idle resources as necessary to keep the process within a prescribed range. Ormeci Matoglu and Vande Vate (2011) address this problem by introducing a practical restriction that allows the controller to change the capacity only when the process state takes a value in a given finite set $\mathcal{S}$. They call this the $\mathcal{S}$-restricted Brownian control problem and formulate a linear programming model to solve it. In a subsequent paper Ormeci Matoglu et al. (2015) show that an optimal solution to the $\mathcal{S}$-restricted problem can be found among a special class of policies called deterministic nonoverlapping control band policies and employ a column-generation approach to refine the set $\mathcal{S}$ to construct an ϵ-optimal policy.

The $\mathcal{S}$-restricted problem essentially precludes the possibility of finding smooth relative value functions, and so solving it requires Tanaka's formula and local time (see Chung and Williams 1990). In this paper, we address the unrestricted problem by adopting a different approach: rather than relying on Tanaka's formula and local time, we construct a deterministic control band policy that is optimal for a problem with modified parameters and whose relative value functions are smooth. Parametrically adjusting the problem parameters yields a policy for the given problem and smooth relative value functions, proving the optimality of this policy. Not restricting the points to a fixed finite set allows us to prove that—under the proportional changeover cost assumption—there is a strongly ordered optimal policy, that is, an optimal policy that limits the use of each drift rate to a single interval; when the process reaches the upper limit of that interval, the policy either changes to the next lower drift rate deterministically or resorts to instantaneous controls to keep the process within the prescribed range, and when the process reaches the lower limit of the interval, the policy either changes to the next higher drift rate deterministically or again resorts to instantaneous controls to keep the process within the prescribed range. We reduce the complexity of proving the optimality of such a policy by proving the existence of optimal relative value functions that form a nondecreasing sequence of functions. As a consequence, our constructive arguments lead to a practical procedure for solving the problem that is tens of thousands times faster than those reported in Ormeci Matoglu et al. (2015).

The problem of controlling a Brownian motion by changing its drift rate has been studied at least since Bather (1968) posed the problem of controlling the output of a dam as one of adjusting the drift rate of a Brownian motion. Since that time, many authors have explored the problem under a variety of constraints and cost functions. See, for example, Ata et al. (2005), Avram and Karaesmen (1996), Chernoff and Petkau (1978), Ghosh and Weerasinghe (2007), Perry and Bar-Lev (1989), Rath (1977), Ormeci Matoglu and Vande Vate (2011), and Wu and Chao (2014).

Earlier studies, including Avram and Karaesmen (1996), Perry and Bar-Lev (1989), and Rath (1977) restrict the controller to only two drift rates, and among those works that allow more than two drift rates, many, including Ata et al. (2005), Bather (1968), and Ghosh and Weerasinghe (2007), do not consider a changeover cost or cost for changing the drift rate. Changeover costs represent a fixed investment required to realize the operating advantages of a more appropriate drift rate. They force the controller to look beyond the immediate advantages to ensure that the change will accrue sufficient benefit over time to merit the fixed investment.

Chernoff and Petkau (1978, p. 717) introduce the problem of minimizing the long-run average cost in which the controller can change the drift rate (and the diffusion coefficient) among the members of a finite set in the face of changeover costs as well as processing and delay costs (but no costs for idling capacity or rejecting orders). They observe that, for problems involving more than two drift rates, "the analytic approach becomes cumbersome." In this paper, we introduce an analytic approach that is significantly less cumbersome and reveals an optimal policy with a simple and practical structure under two assumptions:

- The cost of changing the drift rate is positive and proportional to the difference in the rates when the constant of proportionality depends only on whether the change increases or decreases the drift rate.
- The processing cost is proportional to the drift rate. It is more expensive to reject work than to process it and more expensive to idle capacity than to process work.

Under these two assumptions, we (i) simplify the optimality criteria while maintaining the sharpness of the bounds they provide; (ii) show that the search for optimal policies can be restricted to the class of strongly ordered policies that limit the use of each drift rate to a single control band and either rely on instantaneous controls or deterministically change to the next lower drift rate at the upper limit of that band or either rely on instantaneous controls or deterministically change to the next higher drift rate at the lower limit of that band; (iii) show that the search for an optimal policy can be conducted in such a way as to avoid the need for Tanaka's formula or local time by maintaining smooth relative value functions; and (iv) develop constructive arguments leading to practical procedures for constructing optimal policies.

Section 2 introduces the average cost Brownian drift control problem and our assumptions and establishes optimality criteria under those assumptions. Section 3 describes key properties of various components of optimal policies and the relative value functions used to prove their optimality. In Section 4, we prove Theorem 1, which states that, under our two assumptions, a strongly ordered policy is optimal. A Wolfram computable document format application provides an interactive tool for solving problems with up to 10 drift rates. The algorithm behind this tool is the subject of a forthcoming paper. Computational studies in Section 5 suggest the approach is on the order of tens of thousands of times faster than the approach in Ormeci Matoglu et al. (2015).

2. The Brownian Drift Control Problem

Let

$$W(T) = W(0) + \int_0^T \mu(t)\, dt + \int_0^T \sigma\, dB(t), \quad T \ge 0$$

be a diffusion process with drift rate $\mu(t)$ in some fixed finite set Λ for each $t \ge 0$, variance $\sigma^2 > 0$, and initial level $W(0)$ on some filtered space $\{\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t, t \ge 0\}$. The process $W(T)$ describes the difference between the cumulative work to have arrived and the cumulative work processed by time T, that is, the netput process (Harrison 1985). The drift rate $\{\mu(t) : t \ge 0\}$, which is adapted to the Brownian motion $\{B(t) : t \ge 0\}$, is the difference between the average arrival rate and the rate at which work is completed. We assume the arrival process is time homogeneous and that the controller can, at some cost, shift the processing rate among the finite set Λ of alternatives. The controller must exert the minimal instantaneous control required to keep the process within the prescribed range $[\theta, \Theta]$, where $0 \leq \theta < \Theta$ and Θ is finite. (For an extension of this approach to problems in which Θ may be infinite, but the controller has the freedom to invoke instantaneous controls at any time, see Ormeci Matoglu et al. (2020)). We let $L(T)$ denote the cumulative increases and $R(T)$ the cumulative decreases up to time T exerted by the controller at θ and Θ, respectively. The resulting controlled process is

$$X(T) = X(0) + \int_0^T \mu(t)\,dt + \sigma B(T) + L(T) - R(T), \quad T \geq 0, \tag{1}$$

where $X(0) = W(0)$. We assume, without loss of generality, that $W(0) \in [\theta, \Theta]$. To avoid tedious special cases, we also assume that $0 \notin \Lambda$. Analogous results hold when $0 \in \Lambda$.

A policy defines the times at which and the amounts by which to adjust the drift rate. We restrict attention to the space $\mathcal{P}$ of all nonanticipating policies $\Psi = \{(T_i, \mu_i) : i \geq 0\}$, where (i) $0 = T_0 < T_1 < T_2 < \ldots < T_i < T_{i+1}, \ldots$ is a sequence of stopping times and (ii) each $\mu_i \in \Lambda$ is a random variable adapted to $\mathcal{F}_{T_i}$, indicating the drift rate to change to at time T_i. Under the policy $\Psi = \{(T_i, \mu_i) : i \geq 0\}$, the drift rate $\mu(t) = \mu_i$ for $T_i \leq t < T_{i+1}$.

For each policy $\{(T_i, \mu_i) : i \geq 0\} = \{\mu(t) : t \geq 0\}$, the associated Skorokhod problem (a) $X(t) \in [\theta, \Theta]$, $t \geq 0$; (b) $L(\cdot)$ and $R(\cdot)$ are nondecreasing and continuous with $L(0) = R(0) = 0$; and (c) $\int_0^T 1_{\{X(t) > \theta\}}\, dL(t) = \int_0^T 1_{\{X(t) < \Theta\}}\, dR(t) = 0$, $T \geq 0$, where the continuous process $\{X(t) : t \geq 0\}$ is defined by (1) uniquely defines $\{X, L, R\}$ (see section 2.4 of Harrison 1985). Note that, because the drift rate controls uniquely determine the instantaneous controls exerted at the boundaries, we do not include the latter in our specification of a policy.

To change the drift rate from a given rate $u \in \Lambda$ to a different rate $v \in \Lambda$, the controller must pay a fixed cost $K(u, v)$ that is proportional to the difference in the drift rates, and the constant of proportionality depends only on whether the change increases or decreases the drift rate. We only require that κ, the sum of these two constants, be strictly positive.

Assumption 1. *The cost $K(u, v)$ to change the drift rate from $u \in \Lambda$ to $v \in \Lambda$ is of the form*

$$K(u, v) = \begin{cases} \kappa_+(v - u) & \text{if } v \geq u \\ \kappa_-(u - v) & \text{if } u > v \end{cases},$$

where $\kappa = \kappa_+ + \kappa_- > 0$.

Note that Assumption 1 allows the controller to, for example, enjoy an incentive to reduce the drift rate, that is, $\kappa_- < 0$, and face a cost to increase the drift rate, that is, $\kappa_+ > 0$, as long as κ, the sum of the two costs, is positive. The requirement that κ be positive ensures there is a cost to changing the drift rate. It is sufficient to impose this restriction on the sum of the two costs essentially because, as detailed in the proof of Proposition 1, the controller must increase the drift rate to be able to later decrease the drift rate. Ormeci Matoglu et al. (2019) show that there is an optimal policy with a very simple structure when changeovers are free.

When $X(t) > 0$, there is a backlog of orders, which incurs a linear delay cost at rate h per unit. We allow h to be negative as would be the case when scarcity or time waiting to obtain the service confers value on that service. Instantaneous control exerted at θ (Θ) to keep the process in the interval $[\theta, \Theta]$ incurs a cost per unit of U (M). The controller must also pay a fixed cost p per unit of processing capacity to maintain the drift rate; that is when the drift rate is μ, the controller incurs processing costs at a rate of $p\mu$ per unit time.

We consider the *average cost Brownian control problem*, which is to find a nonanticipating policy Ψ that minimizes the long-run average cost:

$$AC(\Psi) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\int_0^T p\mu(t) + hX(t)\,dt + UL(T) + MR(T) + \sum_{i=1}^{N(T)} K(\mu_{i-1}, \mu_i) \right],$$

where, for each $T \geq 0$, $N(T) = \sup\{n \geq 0 : T_n \leq T\}$ denotes the number of changes in the drift rate by time T. Our proof of Theorem 1, characterizing an optimal policy with a simple exploitable structure, is constructive. We prove the theorem by induction, parameterizing the bounds, costs of instantaneous controls, and changeover costs of a problem on a subset of the drift rates to produce a policy that can be extended to an optimal policy on the full set of drift rates. To help avoid confusion about which problem we are addressing at any point, we use ACBC$(\Lambda, \theta, \Theta, U, M, K)$ to denote the problem with the specified parameters and $\boldsymbol{\gamma}(\Lambda, \theta, \Theta, U, M, K)$ to denote its optimal average cost. When the changeover cost function K satisfies Assumption 1, we write ACBC$(\Lambda, \theta, \Theta,$

U, M, κ) and reserve the more general notation for cases with more complicated changeover cost functions. We primarily focus on changes to the set Λ, the lower limit θ, and the cost U of instantaneous controls at this lower limit. To streamline notation, we often use expressions such as $\boldsymbol{\gamma}(\Lambda', o, V)$ in place of the longer $\boldsymbol{\gamma}(\Lambda', o, \Theta, V, M, K)$ with the understanding that the omitted parameters retain their original values.

Assumption 2. *The cost of capacity is proportional to the drift rate at a cost p per unit per unit time that satisfies $U > p > -M$. When $\boldsymbol{\gamma}(\Lambda, \theta, \Theta, p, M, K) < h\theta$, we also allow $U = p$, and when $\boldsymbol{\gamma}(\Lambda, \theta, \Theta, U, p, K) < h\Theta$, we also allow $M = -p$.*

Assumption 2 requires idling capacity to be more expensive than operating it and rejecting work to cost more than it saves in avoided processing. Ormeci Matoglu et al. (2019) show that, when this assumption fails to hold, the closely related economic average cost Brownian control problem, in which the controller can invoke instantaneous controls at any time, either admits no lower bound or has an optimal policy that relies on a single drift rate. Although the controller enjoys more latitude in deploying instantaneous controls in the economic average cost problem, we adopt Assumption 2 in the classic average cost setting. Arguments in the economic average cost setting are complicated by having to address additional special cases, and our primary purpose is to introduce a theoretically and computationally effective strategy for addressing the family of drift control problems. It is interesting to note that the exceptional boundary conditions in Assumption 2 are needed to address the cases in which the classic formulation's restrictions on the use of instantaneous controls lead to a higher average cost. See Ormeci Matoglu et al. (2020) for an extension of the results in this paper to the economic average cost Brownian control problem.

A deterministic control band $\psi = (u, s, S, \beta, \tau)$ is defined by a drift rate $u \in \Lambda$; a proper interval $(s, S) \subseteq [\theta, \Theta]$; and a pair β, $\tau \in \Lambda$, where $\beta = u$ only if $s = \theta$ and $\tau = u$ only if $S = \Theta$. Given $\mu(t) = u$ and $X(t) \in (s, S)$, a policy implementing the control band ψ maintains the drift rate u until X first hits $\{s, S\}$, at which point it changes the drift rate to β if X first hits s and to τ if X first hits S. If $\beta = u$ (and so $s = \theta$), a policy implementing the control band ψ maintains the drift rate u until X first hits S and relies on instantaneous controls at θ to keep the process from falling below the lower limit θ. Similarly, if $\tau = u$ (and so $S = \Theta$), a policy implementing the control band ψ maintains the drift rate u until X first hits s and relies on instantaneous controls at Θ to keep the process from exceeding that bound. A policy implementing the control band $\psi = (u, \theta, \Theta, u, u)$ maintains the drift rate u indefinitely.

A control band policy $\Psi = \{\psi(u) : u \in \Lambda\}$ is defined by a single control band $\psi(u) = (u, s_u, S_u, \beta(u), \tau(u))$ for each $u \in \Lambda$ with the property that $\cup_{u \in \Lambda}(s_u, S_u) = (\theta, \Theta)$. See Figure 1. For each control band policy $\Psi = \{\psi(u) : u \in \Lambda\}$, we construct an associated directed graph $G(\Psi)$ with a node for each drift rate $u \in \Lambda$ and edges $(u, \beta(u))$ and $(u, \tau(u))$ for each $u \in \Lambda$. Following Puterman (2005), the control band policy Ψ is said to be *unichain* if $G(\Psi)$ contains exactly one strongly connected component with no outgoing edges.

For each unichain policy Ψ, we distinguish between the *recurrent* drift rates, $\mathcal{R}(\Psi)$, defined as the nodes of the unique strongly connected component of $G(\Psi)$ with no outgoing edges, and the remaining drift rates, which we call *transient*. Note that, when Ψ is unichain, there is a path in $G(\Psi)$ from each drift rate to each recurrent drift rate. Consequently, the process eventually reaches a recurrent drift rate, and once it does, it continues to use only recurrent drift rates thereafter.

We say a unichain policy $\Psi = \{\psi(u) : u \in \Lambda\}$ is *ordered* if, for each $u \in \Lambda$, $\beta(u) \geq u \geq \tau(u)$ and *strongly ordered* if $\beta(u)$ is either u or the next greater drift rate and $\tau(u)$ is either u or the next smaller drift rate. Theorem 1 states that, under Assumptions 1 and 2, we may restrict the search for an optimal policy to the family of strongly ordered policies.

To facilitate discussions involving ordered policies, we reserve $u(j)$ to denote the j^{th} smallest drift rate in Λ so that $u(1) < u(2) < \ldots < u(n)$, where $n = |\Lambda|$. To avoid cumbersome notation, we often use $j \in \Lambda$ in place of $u(j) \in \Lambda$ and, for example, refer to a control band $\psi(u(j)) = (u(j), s_{u(j)}, S_{u(j)}, \beta(u(j)), \tau(u(j)))$ with the more streamlined notation $\psi(j) = (j, s_j, S_j, \beta(j), \tau(j))$ and write $K(i, j)$ in place of the more cumbersome $K(u(i), u(j))$. With this notation, a strongly ordered policy satisfies $\beta(j) \in \{j, j+1\} \cap \Lambda$ and $\tau(j) \in \{j, j-1\} \cap \Lambda$ for each $j \in \Lambda$.

For each ordered policy $\Psi = \{\psi(u) : u \in \Lambda\}$, define $\overline{\mu}(\Psi)$ to be the maximum recurrent rate and $\underline{\mu}(\Psi)$ to be the minimum recurrent rate, that is, $\overline{\mu}(\Psi) = \max \mathcal{R}(\Psi)$ and $\underline{\mu}(\Psi) = \min \mathcal{R}(\Psi)$.

Definition 1. For each nonnegative integer k, define $\mathbb{C}_k$ to be the set of all real-valued functions f on $[\theta, \Theta]$ such that the derivatives $f', f'', \ldots, f^{(k)}$ exist and are continuous except at a finite set of points.

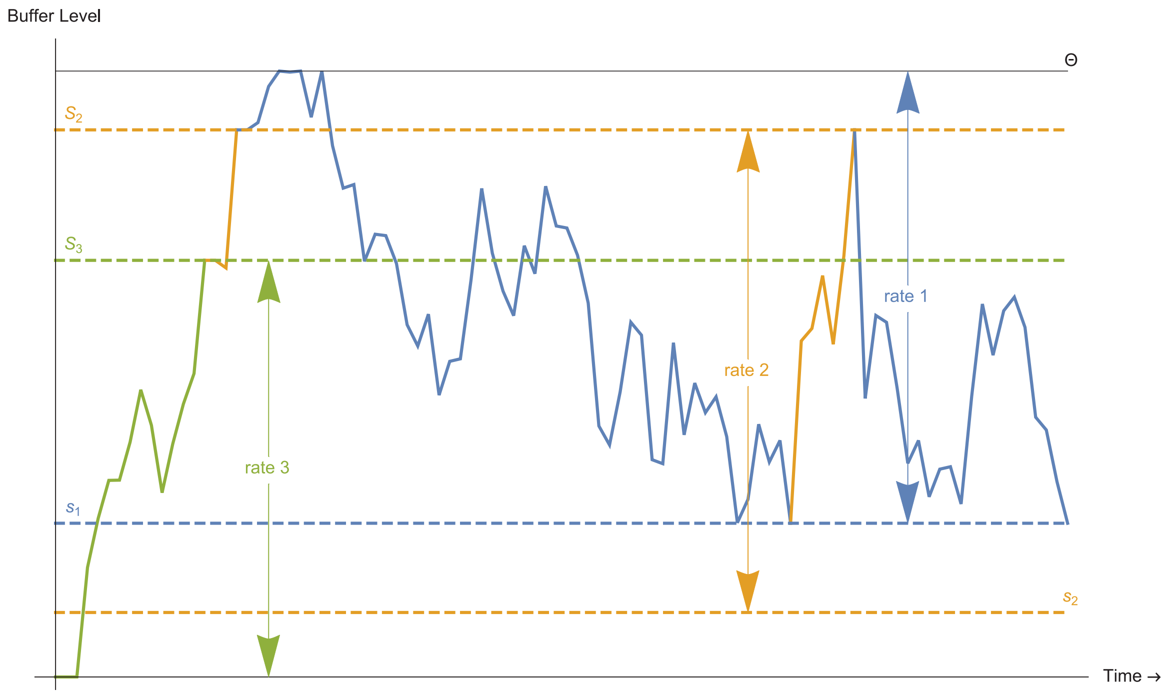
Proposition 1 appears in Ormeci Matoglu and Vande Vate (2011).

Proposition 1. *Suppose that, for each $u \in \Lambda$, $f(\cdot, u) \in \mathbb{C}_2$ and, together with the scalar γ, satisfies,*

$$\Gamma f(x, u) + pu + hx \geq \gamma \text{ for almost all } x \in (\theta, \Theta), \tag{2}$$

$$f(x, u) - f(x, v) \geq -K(v, u) \text{ for all } x \in [\theta, \Theta] \text{ and } v \in \Lambda, \tag{3}$$

Figure 1. A Strongly Ordered Policy with Three Rates



Notes. Invokes Instantaneous Controls at Zero and Θ, Maintains the Fastest Rate, $u(3)$, in the Range from Zero to S_3, Maintains the Slowest Rate, $u(1)$ in the Range from s_1 to Θ, and Maintains $u(2)$ in the Range from s_2 to S_2. The arrows indicate the bands for each drift rate. The color indicates the drift rate employed: green for $u(3)$, orange for $u(2)$, and blue for $u(1)$.

$$\partial_x f(\theta, u) \geq -U \text{ and} \tag{4}$$

$$\partial_x f(\Theta, u) \leq M, \tag{5}$$

where

$$\Gamma f(x,u) \equiv \frac{\sigma^2}{2}\partial_{x,x} f(x,u) + u\,\partial_x f(x,u). \tag{6}$$

Then, $\gamma \leq \mathrm{AC}(\Psi)$ *for each policy* $\Psi \in \mathcal{P}$ *and each initial state* $(X(0), \mu(0))$.

When the partial derivatives of the $\mathbb{C}_2$ functions $\{f(\cdot,u) : u \in \Lambda\}$ form a nondecreasing sequence of real-valued functions on $[\theta,\Theta]$, that is, when $v > u$, $\partial_x f(x,v) \geq \partial_x f(x,u)$ for all $x \in [\theta,\Theta]$ at which both $\partial_x f(x,u)$ and $\partial_x f(x,v)$ are defined, we restate Proposition 1 in terms of these partial derivatives as we do in Proposition 2 using the following notation: for each $i \in \Lambda$, let

$$K_i = \begin{cases} K(i,i-1) + K(i-1,i) & \text{if } i-1 \in \Lambda \\ 0 & \text{otherwise} \end{cases},$$

and given a nondecreasing sequence of real-valued functions $\delta = \{\delta(\cdot,i) : I \in \Lambda\}$, define

$$\Delta_i(\delta) = \begin{cases} \displaystyle\int_\theta^\Theta \delta(x,i) - \delta(x,i-1)\,dx & \text{if } i-1 \in \Lambda \\ 0 & \text{otherwise} \end{cases}$$

for each $i \in \Lambda$.

Proposition 2. *Under Assumption 1, suppose the scalar* γ *and each member of the nondecreasing sequence* $\delta = \{\delta(\cdot,u) : u \in \Lambda\}$ *of functions in* $\mathbb{C}_1$ *satisfies*

$$\frac{\sigma^2}{2}\partial_x \delta(x,u) + u\delta(x,u) + pu + hx \geq \gamma \text{ for almost all } x \in [\theta,\Theta], \tag{7}$$

$$\delta(\theta,u) \geq -U, \tag{8}$$
$$\delta(\Theta,u) \leq M, \text{ and} \tag{9}$$
$$\Delta_u(\delta) \leq K_u. \tag{10}$$

Then, $\gamma \leq \mathrm{AC}(\Psi)$ *for each policy* $\Psi \in \mathcal{P}$ *and each initial state* $(X(0),\mu(0))$.

Proof. Each of Conditions (7)–(9) is a simple translation from the corresponding condition in Proposition 1. To show that (10) implies (3), we argue that, if the nondecreasing sequence of $\mathbb{C}_1$ functions $\{\delta(\cdot,u): u \in \Lambda\}$ satisfies (10) for each $u \in \Lambda$, then there exist scalars $f(\theta,u)$ for $u \in \Lambda$ such that the $\mathbb{C}_2$ functions $f(x,u) = \int_\theta^x \delta(y,u)\,dy + f(\theta,u)$ for $u \in \Lambda$ satisfy (3). To see this, rewrite (3) as

$$f(x,u) - f(x,v) \geq -\kappa_-(v-u) \text{ for all } x \in [\theta,\Theta] \text{ and } v,u \in \Lambda \text{ with } v > u, \tag{11}$$
$$f(x,u) - f(x,v) \geq -\kappa_+(u-v) \text{ for all } x \in [\theta,\Theta] \text{ and } v,u \in \Lambda \text{ with } u > v. \tag{12}$$

Because $\{\delta(\cdot,u): u \in \Lambda\}$ is a nondecreasing sequence by assumption, we can rewrite (11) and (12) as

$$-\kappa_+(v-u) \leq f(\theta,v) - f(\theta,u) \leq \kappa_-(v-u) - \int_\theta^\Theta \delta(y,v) - \delta(y,u)\,dy \text{ for } u,v \in \Lambda \text{ with } v > u. \tag{13}$$

The system of linear constraints (13) on the variables $f(\theta,\mu)$ for $\mu \in \Lambda$ has a solution (see, for example, section 5.2 of Ahuja et al. 1993) if and only if there is no negative length cycle in the complete directed network on the set of nodes Λ under the arc costs

$$c_{i,j} = \begin{cases} \kappa_-(u(j)-u(i)) - \displaystyle\int_\theta^\Theta \delta(y,j) - \delta(y,i)\,dy & \text{if } j > i \\ \kappa_+(u(i)-u(j)) & \text{if } j < i \end{cases} \ldots \tag{14}$$

Consider a cycle $u^1, u^2, \ldots, u^k, u^1$ of drift rates in Λ, beginning and ending with the same drift rate. Because the cycle begins at u^1 and returns to u^1,

$$\sum_{u^{i+1}<u^i} (u^i - u^{i+1}) = \sum_{u^{i+1}>u^i} (u^{i+1} - u^i),$$

and the length of this cycle under the costs (14) is

$$\begin{aligned} &\sum_{u^{i+1}>u^i} \left(\kappa_-(u^{i+1}-u^i) - \int_\theta^\Theta \delta\left(y,u^{i+1}\right) - \delta\left(y,u^i\right)dy\right) + \sum_{u^{i+1}<u^i} \kappa_+(u^i - u^{i+1}) \\ &= \sum_{u^{i+1}>u^i} \left((\kappa_- + \kappa_+)(u^{i+1}-u^i) - \int_\theta^\Theta \delta\left(y,u^{i+1}\right) - \delta\left(y,u^i\right)dy\right), \end{aligned}$$

which is nonnegative for every cycle if and only if (10) is satisfied.

Corollary 1 is a modest generalization of Proposition 2 that allows us to incorporate an additional fixed factor into the costs for changing to and from the fastest drift rate in Λ. This simple extension serves a key role in the proof of Theorem 1. Analogous arguments apply for the slowest drift rate as well.

Corollary 1. *Suppose*

$$K(u,v) = \begin{cases} \kappa_+(v-u) & \text{if } v \geq u \\ \kappa_-(u-v) & \text{if } u > v \end{cases} + \begin{cases} k_0 & \text{if } \max\{u,v\} = \max\Lambda \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

for each pair $u, v \in \Lambda$ *and some scalar* $k_0 \geq 0$, *where* $\kappa_+ + \kappa_- > 0$. *Suppose further that the scalar* γ *and each member of the nondecreasing sequence* $\{\delta(\cdot,u): u \in \Lambda\}$ *of functions in* $\mathbb{C}_1$ *satisfies* (7)–(10). *Then,* $\gamma \leq \mathrm{AC}(\Psi)$ *for each policy* $\Psi \in \mathcal{P}$ *and each initial state* $(X(0),\mu(0))$.

Given a pair (Ψ,δ), where Ψ is a unichain policy and δ is a nondecreasing sequence of $\mathbb{C}_1$ functions, let $r_1 < r_2 < \ldots < r_{|\mathcal{R}(\Psi)|}$ be the members of $\mathcal{R}(\Psi)$ and define

$$\overline{K}_i(\Psi) = \begin{cases} K(r_i,r_{i-1}) + K(r_{i-1},r_i) & \text{if } i > 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \overline{\Delta}_i(\Psi) = \begin{cases} \displaystyle\int_\theta^\Theta \delta(x,r_i) - \delta(x,r_{i-1})\,dx & \text{if } i > 1 \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1,2,\ldots,|\mathcal{R}(\Psi)|$.

Corollary 2 highlights the fact that, when all of the "active" bounds of Proposition 2 are satisfied with equality, the lower bound is "tight" and so yields the average cost of the given policy. Corollary 3 extends this observation to the more general changeover costs of the form (15).

Corollary 2. *Under Assumption 1, let $\Psi = \{\psi(u) : u \in \Lambda\}$ be a unichain policy and suppose that the scalar γ and each member of the nondecreasing sequence $\delta = \{\delta(\cdot, u) : u \in \Lambda\}$ of functions in $\mathbb{C}_1$ satisfies*

$$\frac{\sigma^2}{2}\partial_x\delta(x,u) + u\delta(x,u) + pu + hx = \gamma \text{ for almost all } x \in (s_u, S_u), \tag{16}$$

$$\delta(\theta, u) = -U, \tag{17}$$

$$\delta(\Theta, u) = M \text{ and} \tag{18}$$

$$\overline{\Delta}_i(\Psi) = \overline{K}_i(\Psi) \text{ for } i = 1, 2, \ldots |\mathcal{R}(\Psi)|, \tag{19}$$

and then $\gamma = AC(\Psi)$.

Corollary 3. *When the costs K to change the drift rate are defined by (15), let $\Psi = \{\psi(u) : u \in \Lambda\}$ be a unichain policy and suppose that the scalar γ and each member of the nondecreasing sequence $\delta = \{\delta(\cdot, u) : u \in \Lambda\}$ of functions in $\mathbb{C}_1$ satisfies (16)–(19); then, $\gamma = AC(\Psi)$.*

A pair (γ, δ), where γ is a scalar and $\delta = \{\delta(\cdot, u) : u \in \Lambda\}$ is a nondecreasing sequence of functions in $\mathbb{C}_1$ that satisfies (7)–(10) is said to satisfy *the lower bound conditions*. A triple (Ψ, γ, δ), where Ψ is a unichain policy, γ is a scalar, and $\delta = \{\delta(\cdot, u) : u \in \Lambda\}$ is a nondecreasing sequence of functions in $\mathbb{C}_1$ that satisfies (16)–(19) is said to be a *complementary triple*. When

$$\delta(s_u, u) < -p < \delta(S_u, u) \tag{20}$$

for each $u \in \Lambda$, we say the complementary triple (Ψ, γ, δ) is a *canonical complementary triple*. To accommodate the boundary conditions of Assumption 2, a canonical complementary triple also allows $\delta(s_u, u) = -p$ if $s_u = \theta$ and $\gamma < h\theta$. Condition (20) highlights a relationship between the processing cost p and the partial derivatives δ of the relative value functions f that is both a key observation and an important tool in the proof of Theorem 1. Lemma 1, which originally appeared in Ormeci Matoglu et al. (2015), shows that an optimal policy can only transition to a faster rate at a point at which the partial derivative of the relative value function is less than $-p$ and can only transition to a slower rate at point at which the partial derivative of the relative value function is greater than $-p$. As we show in Lemma 5, this property is key to ensuring that the partial derivatives of the relative value functions satisfy (7) of Proposition 2.

Observe that, if (Ψ, γ, δ) is a complementary triple and (γ, δ) satisfies the lower bound conditions, then γ is a lower bound on the average cost of any nonanticipating policy and Ψ is a unichain policy that achieves the lower bound γ, proving that Ψ is an optimal policy. We call such a triple an *optimal triple*, and when it is also a canonical complementary triple, we say (Ψ, γ, δ) is a *canonical optimal triple*. Finally, when (Ψ, γ, δ) is a canonical optimal triple and Ψ is strongly ordered, we say (Ψ, γ, δ) is a *strongly ordered canonical optimal triple*. Theorem 1 states that, under Assumptions 1 and 2, a strongly ordered canonical optimal triple exists and, hence, that the lower bounds of Proposition 2 are tight and there is an optimal policy that is not just ordered, but strongly ordered.

Theorem 1. *Under Assumptions 1 and 2, the infimum of $AC(\Psi)$ over all policies $\Psi \in \mathcal{P}$ is equal to the supremum of γ over all pairs (γ, δ) satisfying the conditions of Proposition 2. In particular, there is a canonical optimal triple (Ψ, γ, δ) such that Ψ is strongly ordered.*

Before proving Theorem 1, we observe an important consequence: if (γ, δ) satisfies the lower bound conditions, then (19) implies

$$\Delta_u = K_u \text{ for each } u \in \Lambda \text{ with } \underline{\mu}(\Psi) < u \le \overline{\mu}(\Psi). \tag{21}$$

Thus, we may restate (19) more simply as (21). One feature of a strongly ordered policy Ψ is that $\mathcal{R}(\Psi) = \{u \in \Lambda : \underline{\mu}(\Psi) \le u \le \overline{\mu}(\Psi)\}$.

3. Preliminaries

Each ordered optimal triple (Ψ, γ, δ) is a complementary triple, and so, for each $u \in \Lambda$,

$$\delta(x,u) = \begin{cases} \delta(x, \beta(u)) & \text{for } x \le s_u \\ g(x, u, \gamma) & \text{for } s_u < x < S_u\,, \\ \delta(x, \tau(u)) & \text{for } x \ge S_u \end{cases} \tag{22}$$

where

$$g(x,u,\gamma) = C_u e^{-\frac{2u}{\sigma^2}x} - \frac{h}{u}x + \frac{\gamma}{u} + \frac{h\sigma^2}{2u^2} - p \tag{23}$$

is the general solution to the differential Equation (16) and C_u is an arbitrary constant satisfying

$$g(s_u,u,\gamma) = \begin{cases} g(s_u,\beta(u),\gamma) & \text{if } \beta(u) > u \\ -U & \text{if } \beta(u) = u \end{cases} \quad \text{and} \tag{24}$$

$$g(S_u,u,\gamma) = \begin{cases} g(S_u,\tau(u),\gamma) & \text{if } \tau(u) < u \\ M & \text{if } \tau(u) = u \end{cases} \tag{25}$$

so that

$$\delta(x,u) = \begin{cases} \dots \\ g(x,\beta(\beta(u)),\gamma) & \text{for } s_{\beta(\beta(u))} < x \le s_{\beta(u)} \\ g(x,\beta(u),\gamma) & \text{for } s_{\beta(u)} < x \le s_u \\ g(x,u,\gamma) & \text{for } s_u < x < S_u \\ g(x,\tau(u),\gamma) & \text{for } S_{\tau(u)} > x \ge S_u \\ g(x,\tau(\tau(u)),\gamma) & \text{for } S_{\tau(\tau(u))} > x \ge S_{\tau(u)} \\ \dots \end{cases} \tag{26}$$

is continuous and satisfies (16)–(18). Note that

$$\partial_x g(x,u,\gamma) = -\frac{2u}{\sigma^2} C_u e^{-\frac{2u}{\sigma^2}x} - \frac{h}{u} \tag{27}$$

and

$$\partial_{x,x} g(x,u,\gamma) = \frac{4u^2}{\sigma^4} C_u e^{-\frac{2u}{\sigma^2}x}, \tag{28}$$

from which we see that $g(\cdot,u,\gamma)$ is either convex or concave, depending on the sign of C_u.

Lemma 1, which appears in Ormeci Matoglu et al. (2015), proves remarkably powerful. It shows that an optimal policy can only increase the drift rate at a point at which the partial derivative of the relative value function is less than $-p$ and decrease the drift rate at point at which the partial derivative of the relative value function is greater than $-p$. A fundamental step in proving Theorem 1 is proving the existence of optimal relative value functions whose partial derivatives cross this threshold once, lie below $-p$ before that point, and remain above $-p$ beyond that point.

Lemma 1. *For each pair $u, v \in \Lambda$ and each choice of C_u, C_v, and $\gamma \in \mathbb{R}$,*

$$\frac{\sigma^2}{2}\partial_x g(x,v,\gamma) + ug(x,v,\gamma) + pu + hx - \gamma = (u-v)(g(x,v,\gamma)+p) \text{ for all } x \in \mathbb{R}. \tag{29}$$

Furthermore, if $g(z,u,\gamma) = g(z,v,\gamma)$ for some point $z \in \mathbb{R}$, then

$$(u-v)(g(z,v,\gamma)+p) = \frac{\sigma^2}{2}\Big(\partial_x g(z,v,\gamma) - \partial_x g(z,u,\gamma)\Big). \tag{30}$$

When (Ψ,γ,δ) is a canonical complimentary triple, Lemma 2 establishes, for each $u \in \Lambda$, the existence and uniqueness of a point $x_p(u) \in (s_u, S_u)$ such that $\delta(x_p(u),u) = -p$ and $\partial_x\delta(x_p(u),u) > 0$. These points play a pivotal role in the proof of Theorem 1. Lemma 2 ensures the existence and uniqueness of points at which the components of the partial derivatives of the relative value functions cross the critical value $-p$ from below.

Lemma 2. *Given real constants u, γ, and C_u, suppose the function $g(\cdot,u,\gamma)$ defined by (23) satisfies $g(s,u,\gamma) < -p < g(S,u,\gamma)$ for two real values s and S with $s < S$. Then, there is a unique real value $x_p \in (s,S)$ such that $g(x_p,u,\gamma) = -p$ and $\partial_x g(x_p,u,\gamma) > 0$. Similarly, if $g(s,u,\gamma) = -p < g(S,u,\gamma)$ and $\gamma < hs$ or if $g(s,u,\gamma) < -p = g(S,u,\gamma)$ and $\gamma < hS$, there is a unique real value $x_p \in (s,S)$ such that $g(x_p,u,\gamma) = -p$ and $\partial_x g(x_p,u,\gamma) > 0$.*

Proof. We first consider the case $g(s,u,\gamma) < -p < g(S,u,\gamma)$. Because $g(\cdot,u,\gamma)$ is continuous, there must be a point $x_p \in (s,S)$ such that $g(x_p,u,\gamma) = -p$ and $\partial_x g(x_p,u,\gamma) \ge 0$. That x_p is unique follows from the fact that $g(\cdot,u,\gamma)$ must be either convex or concave. To see that the inequality is strict, observe that, if $\partial_x g(x_p, u, \gamma) = 0$, then either $g(\cdot, u, \gamma)$ is convex and $g(x, u, \gamma) \geq -p$ for all $x \in \mathbb{R}$, contradicting the fact that $g(s, u, \gamma) < -p$, or $g(\cdot, u, \gamma)$ is concave and $g(x, u, \gamma) \leq -p$ for all $x \in \mathbb{R}$, contradicting the fact that $g(S, u, \gamma) > -p$.

We next consider the case $g(s, u, \gamma) = -p < g(S, u, \gamma)$ and $\gamma < hs$. Because $g(s, u, \gamma) = -p$, $C_u = e^{2u/\sigma^2}(hs/u - h\sigma^2/2u^2 - \gamma/u)$ and $\partial_x g(s, u, \gamma) = 2/\sigma^2(\gamma - hs) < 0$ from which we see, by the preceding arguments, that x_p exists and is unique. Similarly, when $g(s, u, \gamma) < -p = g(S, u, \gamma)$ and $\gamma < hS$, the fact that $g(S, u, \gamma) = -p$ implies $C_u = e^{2u/\sigma^2 S}(hS/u - h\sigma^2/2u^2 - \gamma/u)$ and $\partial_x g(S, u, \gamma) = 2/\sigma^2(\gamma - hS) < 0$. We conclude by observing that, if $g(s, u, \gamma) = -p = g(S, u, \gamma)$, then

$$\gamma = hs + h(S - s)\left(\frac{1}{1 - e^{-\frac{2u}{\sigma^2}(S-s)}} - \frac{\sigma^2}{2u(S - s)}\right).$$

And because $0 < 1/(1 - e^{-x}) - 1/x < 1$ for all finite values of x, γ lies between hs and hS.

Lemma 3 characterizes the unique real value at which a component of the partial derivative of a relative value function crosses the critical value $-p$ from below. We frequently invoke Lemma 3 with the points s and S being either the lower and upper limits of a control band or the lower and upper limits of the process with the corresponding function values V and ω, respectively, at those points. The conditions relating these values to the critical value $-p$ reflect the conditions of Assumption 2. That characterization relies on the Lambert-W functions $W_k(\cdot)$, where $W_k(\cdot)$ is the k^{th} branch of the inverse relation for the function $f(z) = ze^z$ defined on the complex field, to characterize the value of x_p. For each real value z with $-e^{-1} < z < 0$, there are two possible real values for the inverse relation of the function $f(z) = ze^z$, one on the branch $W_{-1}(\cdot)$ with $W_{-1}(z) < -1$ and the other on the branch $W_0(\cdot)$ with $W_0(z) > -1$ (see, for example, Lambert 1758, 1772; Euler 1921; Corless et al. 1996). The proof is technical and appears in the appendix.

Lemma 3. *Suppose that, for some $S > s \geq 0$, $u \neq 0$, and γ, $g(s, u, \gamma) = -V$, where either $V < -p$ or $V = -p$ and $\gamma < hs$, and $g(S, u, \gamma) = \omega$, where either $\omega > -p$ or $\omega = -p$ and $\gamma < hS$, then*

$$x_p(u) = \begin{cases} s + \dfrac{\sigma^2}{2u}\log\left(\dfrac{\gamma - u(p - V)}{\gamma}\right) & \text{if } h = 0 \\ \dfrac{\gamma}{h} + \dfrac{\sigma^2}{2u}\left(1 + W_{\ell(hu)}\left(e^{-1-\frac{2u}{\sigma^2}\frac{\gamma}{h}+\frac{2u}{\sigma^2}s}\left(-1 - \dfrac{2u}{\sigma^2}\dfrac{\gamma}{h} + \dfrac{2u}{\sigma^2}s + \dfrac{2u^2}{h\sigma^2}(p - V)\right)\right)\right) & \text{if } h \neq 0 \end{cases}, \quad (31)$$

where

$$\ell(x) = \begin{cases} 0 & \text{if } x < 0 \\ -1 & \text{if } x > 0 \end{cases},$$

is the unique real value satisfying $g(x_p(u), u, \gamma) = -p$ and $\partial_x g(x_p(u), u, \gamma) \geq 0$.

Lemma 4 further emphasizes the critical value $-p$, initially highlighted in Lemma 1, and the points $x_p(u)$ for $u \in \Lambda$. To streamline the proof of Lemma 4 and subsequent lemmas, we introduce the following notation: given a function $\delta(\cdot, u)$ of the form (26), we define $\rho_u(x)$ for each point $x \in (\theta, \Theta)$ and drift rate $u \in \Lambda$ to be the drift rate $v \in \Lambda$ such that $\delta(x, u) = g(x, v, \gamma)$ and $s_v < x < S_v$.

We occasionally employ (30) in the context of a point $z \in (\theta, \Theta)$ at which $g(z, v, \gamma) = \delta(z, u)$, where $g(\cdot, v, \gamma)$ is a function of the form (23) and $\delta(\cdot, u)$ is a $\mathbb{C}_1$ function of the form (26). When $\delta(\cdot, u)$ is differentiable at z, the fact that

$$(\rho_u(z) - v)(g(z, v, \gamma) + p) = \frac{\sigma^2}{2}\left(\partial_x g(z, v, \gamma) - \partial_x \delta(z, u)\right) \quad (32)$$

is an immediate consequence of Lemma 1. When $\delta(\cdot, u)$ is not differentiable at z, that is, $\delta(y, u) = g(y, \rho_u(z-), \gamma)$ for $y \in (z - \epsilon, z)$ and $\delta(y, u) = g(y, \rho_u(z+), \gamma)$ for $y \in (z, z + \epsilon)$ for some $\epsilon > 0$, where $\rho_u(z-) \neq \rho_u(z+)$, we see from Lemma 1 that

$$(w - v)(g(z, v, \gamma) + p) = \frac{\sigma^2}{2}\left(\partial_x g(z, v, \gamma) - \partial_x g(z, w, \gamma)\right) \text{ for } w \in \{\rho_u(z-), \rho_u(z+)\}. \quad (33)$$

We often write (32) without explicitly addressing the special case in which $\delta(\cdot, u)$ is not differentiable at z with the understanding that (33) extends the result to the subdifferential.

In the proofs of Lemmas 4 and 5, we employ the notation $\beta^i(u)$ and $\tau^i(u)$ for a drift rate $u \in \Lambda$ and an integer $i > 0$ to refer to the iterated composition of the maps β and τ, respectively, for example, $\beta^2(u) = \beta(\beta(u))$.

Lemma 4 extends the observations in Lemmas 1 and 2 about crossings of the critical value $-p$ from the individual components $g(\cdot,u,\gamma)$ defining δ to the partial derivatives of the relative value functions themselves.

Lemma 4. *Under Assumption 1, suppose that (Ψ,γ,δ) is a canonical complementary triple. Then, for each $u\in\Lambda$, $\delta(x,u)>-p$ for all $x\in(x_p(u),\Theta)$ and $\delta(x,u)<-p$ for all $x\in(\theta,x_p(u))$. Further, for each pair u, $v\in\Lambda$ with $u<v$, $x_p(v)\le x_p(u)$.*

Proof. By Lemma 2, the assumption that (Ψ,γ,δ) is a canonical complimentary triple ensures that $x_p(u)$ exists and is unique for each $u\in\Lambda$.

We first argue that, for each $u\in\Lambda$, $\delta(x,u)>-p$ for all $x\in(S_u,\Theta)$. Suppose that, for some $u\in\Lambda$, $\delta(x,u)\le -p$ for some point $x\in(S_u,\Theta)$ and let $v=\rho_u(x)$. Because $x>S_u$, $v\neq u$, and so it must be the case that $v=\tau^i(u)$ for some integer $i\ge 1$ and, in particular, $s_v<S_{\tau^{i-1}(u)}<x<S_v$. Now, because (Ψ,γ,δ) is a canonical complementary triple by assumption, $g(s_v,v,\gamma)<-p$, $g(S_{\tau^{i-1}(u)},v,\gamma)>-p$, and $g(S_v,v,\gamma)>-p$. So $g(x,v,\gamma)\le -p$, contradicts the fact that $g(\cdot,v,\gamma)$ is either convex or concave. Note that, when $\gamma<h\Theta$, $g(\Theta,u,\gamma)$ can be equal to $-p$.

We next observe that, because $\delta(x,u)=g(x,u,\gamma)$ for all $x\in(s_u,S_u)$, $g(s_u,u,\gamma)<-p=g(x_p(u),u,\gamma)<g(S_u,u,\gamma)$ if $\delta(x^*,u)\le -p$ for some $x^*\in(x_p(u),S_u)$ and then $g(\cdot,u,\gamma)$ can be neither convex nor concave. Similar arguments show that $\delta(x,u)<-p$ for all $x\in(\theta,x_p(u))$. When $u<v$, the fact that $x_p(v)\le x_p(u)$ follows immediately from the fact that δ is a nondecreasing sequence, and so, in particular, $\delta(x_p(u),v)\ge\delta(x_p(u),u)=-p$.

Lemma 5 is an immediate consequence of Lemma 4 and highlights the fact that a canonical complimentary triple satisfies all the optimality conditions except possibly (10). Thus, to prove that the canonical complementary triples we construct in our inductive argument are indeed optimal triples, we need only show they satisfy (10).

Lemma 5. *Under Assumption 1, suppose that (Ψ,γ,δ) is an ordered canonical complementary triple. Then, δ and γ satisfy (7)–(9).*

Proof. Because (Ψ,γ,δ) is a complementary triple, $\delta(\cdot,u)$ satisfies (8) and (9) with equality for each $u\in\Lambda$. To see that the pair γ and δ satisfy (7) as well, consider a drift rate $u\in\Lambda$. For $x\in(s_u,S_u)$, $\delta(x,u)=g(x,u,\gamma)$ and satisfies (7) with equality. For $x\in(\theta,\Theta)$ such that $\delta(\cdot,u)$ is differentiable at x,

$$\begin{aligned}\frac{\sigma^2}{2}\partial_x\delta(x,u)+u\delta(x,u)+pu+hx-\gamma&=\frac{\sigma^2}{2}\partial_x g(x,\rho_u(x),\gamma)+ug(x,\rho_u(x),\gamma)+pu+hx-\gamma\\&=(u-\rho_u(x))(g(x,\rho_u(x),\gamma)+p)\end{aligned}\tag{34}$$

by Lemma 1. Because the policy is ordered, when $x<s_u$, $\rho_u(x)=\beta^i(u)>u$ for some integer $i\ge1$, and by Lemma 4, $g(x,\rho_u(x),\gamma)=\delta(x,u)<-p$ because $x<s_u<x_p(u)$. Thus, the right-hand side of (34) is positive.

Similarly, when $x>S_u$, $\rho_u(x)=\tau^i(u)<u$ for some integer $i\ge1$, and by Lemma 4, $g(x,\rho_u(x),\gamma)=\delta(x,u)>-p$ because $x>S_u>x_p(u)$. Thus, the right-hand side of (34) is again positive, proving that γ and δ satisfy (7).

When Λ consists of a single drift rate u, the only possible policy is $\Psi=\{(u,\theta,\Theta,u,u)\}$, and (16)–(19) reduce to

$$\frac{\sigma^2}{2}\partial_x g(x,u)+ug(x,u)+pu+hx=\gamma \text{ for almost all } x\in[\theta,\Theta],\tag{35}$$

$$g(\theta,u)=-U \text{ and}\tag{36}$$

$$g(\Theta,u)=M.\tag{37}$$

Conditions (36) and (37) imply

$$\gamma=\gamma(u,\theta,\Theta,U,M)\equiv\frac{\left(M+Ue^{-\frac{2u}{\sigma^2}(\Theta-\theta)}\right)u+h(\Theta-\theta)}{1-e^{-\frac{2u}{\sigma^2}(\Theta-\theta)}}+h\theta+pu-\frac{h\sigma^2}{2u} \text{ and}\tag{38}$$

$$C_u=e^{\frac{2u}{\sigma^2}\theta}\left(p-U-\frac{\gamma}{u}+\frac{h\theta}{u}-\frac{h\sigma^2}{2u^2}\right)=e^{\frac{2u}{\sigma^2}\theta}\left(\frac{u(M+U)+h(\Theta-\theta)}{\left(e^{-\frac{2u}{\sigma^2}(\Theta-\theta)}-1\right)u}\right).\tag{39}$$

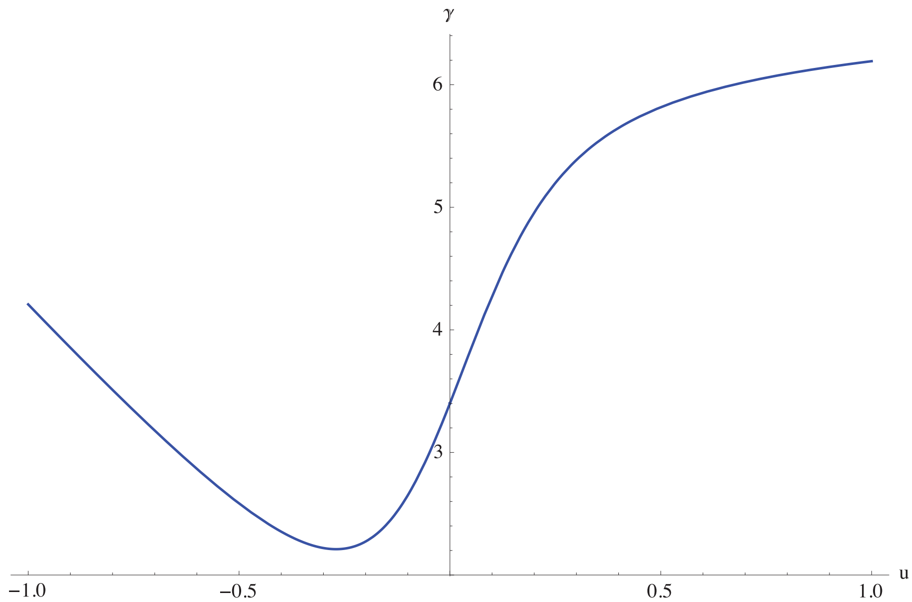
Thus, we have Lemma 6, which forms the basis for our inductive argument.

Lemma 6. *Under Assumption 2, when Λ consists of a single drift rate u, (Ψ,γ,δ), where*

$$\Psi=\{(u,\theta,\Theta,u,u)\}$$
$$\gamma=\gamma(\{u\},\theta,\Theta,U,M) \text{ and}$$
$$\delta(x,u)=g(x,u,\gamma) \text{ for } \theta\le x\le\Theta,$$

where C_u is defined by (39), is a canonical optimal policy.

Figure 2. The Function $\boldsymbol{\gamma}(\cdot)$ with $\sigma = h = p = 1$, $M = 1.5$, $U = 5$, $\theta = 0$, and $\Theta = 10$



Note. The function $\boldsymbol{\gamma}(\cdot)$ is not convex but, under Assumption 2, is unimodal.

To avoid tedious notation, we use $\boldsymbol{\gamma}(u)$ in place of $\gamma(\{u\}, \theta, \Theta, U, M)$. From Figure 2, it is evident that $\boldsymbol{\gamma}(\cdot)$ is not convex. Lemma 7 states that, under Assumption 2, $-\boldsymbol{\gamma}(\cdot)$ is unimodal, and so $\boldsymbol{\gamma}(\cdot)$ achieves a minimum value on $\mathbb{R}$ at a unique point u^*. The proof of Lemma 7 is technical and appears in the appendix.

Whereas Lemmas 1–4 are key to the argument that an optimal policy can only transition to a faster rate on one side of the critical value $-p$ and to a slower rate on the other side of that value, Lemmas 7 and 8 provide key tools for ensuring that the partial derivatives of the corresponding relative value functions do cross and so that there are points at which the policy can change the drift rate.

Lemma 7. *If $U \geq p \geq -M$, the function $-\boldsymbol{\gamma}(\cdot): \mathbb{R} \to \mathbb{R}$ is unimodal.*

With γ and C_u defined as in (38) and (39), (Ψ, γ, g), where $\Psi = \{(u, \theta, \Theta, u, u)\}$, is a canonical complementary triple and the pair (γ, g) satisfies the lower bound conditions (7)–(9), and so (Ψ, γ, g) is a canonical optimal triple for ACBC$(\{u\}, \theta, \Theta, U, M, K)$ (the changeover costs are irrelevant for a single drift rate).

We observe without proof that $\boldsymbol{\gamma}(\Lambda, \theta, \Theta, U, M, \kappa)$ is increasing in U and M and nondecreasing (increasing if no single-rate policy is optimal) in κ. Lemma 8 appears in Ormeci Matoglu et al. (2015). We offer a shorter proof here.

Lemma 8. *If $\gamma \leq \boldsymbol{\gamma}(\mu)$ and $g(\theta, \mu, \gamma) = -U$, then $g(\Theta, \mu, \gamma) \leq M$. Similarly, if $g(\Theta, \mu, \gamma) = M$, then $g(\theta, \mu, \gamma) \geq -U$.*

Proof. This is an immediate consequence of (38) and the fact that $\gamma(u, \theta, \Theta, U, M, K)$ is increasing in U and M.

Our inductive argument relies primarily on adjusting the costs of instantaneous controls of a problem on a subset of the drift rates. But that capability is limited by Assumption 2. Once the cost of idling capacity, for example, falls to p, the cost of processing, the controller no longer has an incentive to process work. In some cases, as illustrated in Example 2, reducing the cost of instantaneous controls to this limit is not sufficient to ensure we can find a policy on the smaller problem that can be extended to an optimal policy on the full problem. In this case, our argument shifts to adjusting the bounds of the process for the smaller problem while holding the cost of idling capacity, for example, at its limiting value p. For this process to work, we must ensure that the optimal average cost for the smaller problem is decreasing as we change the bound. Lemma 9 provides conditions ensuring the optimal average cost reduces.

Lemma 9. *Suppose $\boldsymbol{\gamma}(\Lambda, \theta, \Theta, p, M, \kappa) < h\theta$ and (Ψ, γ, δ) is a strongly ordered canonical optimal triple for* ACBC$(\Lambda, \theta, \Theta, p, M, \kappa)$ *with $s_{n-1} > \theta$, where $n = |\Lambda|$. Then, $\partial_\theta \boldsymbol{\gamma}(\Lambda, \theta, \Theta, p, M, \kappa) < 0$.*

Proof. Observe that $\partial_x g(\theta, n, \gamma) = \sigma^2/2(\gamma - h\theta) < 0$, and so $g(\theta + \epsilon, n, \gamma) < -p$ for all ϵ sufficiently small. Simply adjusting s_n to $\theta + \epsilon$ transforms (Ψ, γ, δ) to a strongly ordered canonical optimal triple for ACBC$(\Lambda, \theta + \epsilon, \Theta, -g(\theta + \epsilon, n, \gamma), M, \kappa)$, proving that $\boldsymbol{\gamma}(\Lambda, \theta + \epsilon, \Theta, -g(\theta + \epsilon, n, \gamma), M, \kappa) = \gamma$. But the fact that $g(\theta + \epsilon, n, \gamma) < -p$ and the average cost is increasing in the cost of instantaneous controls at the lower bound proves that

$\boldsymbol{\gamma}(\Lambda, \theta+\epsilon, \Theta, p, M, \kappa) < \boldsymbol{\gamma}(\Lambda, \theta+\epsilon, \Theta, -g(\theta+\epsilon, n, \gamma), M, \kappa) = \gamma$. As this is true for all sufficiently small ϵ, we see that $\partial_\theta \boldsymbol{\gamma}(\Lambda, \theta, \Theta, p, M, \kappa) < 0$.

4. Proof of Theorem 1

Our proof of Theorem 1 is by induction on n, the cardinality of Λ. As we observed in Lemma 6, Theorem 1 is trivially true when $n = 1$ and Λ consists of a single drift rate. Lemmas 10–12 provide the inductive steps by showing how, assuming Theorem 1 is true for a given set Λ of drift rates, to construct a strongly ordered canonical optimal triple for a problem satisfying Assumptions 1 and 2 on a larger set $\Lambda \cup \{v\}$ of drift rates, on which $v > u(n) = \max \Lambda$. The idea is to construct a strongly ordered canonical optimal triple for $\Lambda \cup \{v\}$ from a strongly ordered canonical optimal triple for a problem on the smaller set Λ with an adjusted cost V for instantaneous controls at the lower bound and, potentially, an adjusted lower bound o, where $p \le V \le U$ and $\theta \le o$ satisfy Assumption 2.

To facilitate these arguments, we first introduce the parameter V into our notation so that $(\Psi_V, \gamma_V, \delta_V)$ with $\Psi_V = \{\psi_V(u) : u \in \Lambda\}$, where, for each $u \in \Lambda$, $\psi_V(u) = (u, s_u(V), S_u(V), \beta_V(u), \tau_V(u))$, and $\delta_V = \{\delta_V(\cdot, u) : u \in \Lambda\}$, is a canonical optimal triple for the problem ACBC$(\Lambda, \theta, \Theta, V, M, \kappa)$. To avoid confusion, we also introduce parameters Z, z, and γ into our notation for $x_p(\cdot)$ so that $x_p(u, Z, z, \gamma)$ is defined as the unique solution x to $g(x, u, \gamma) = -p$ with $\partial_x g(x, u, \gamma) \ge 0$ when $g(z, u, \gamma) = -Z$.

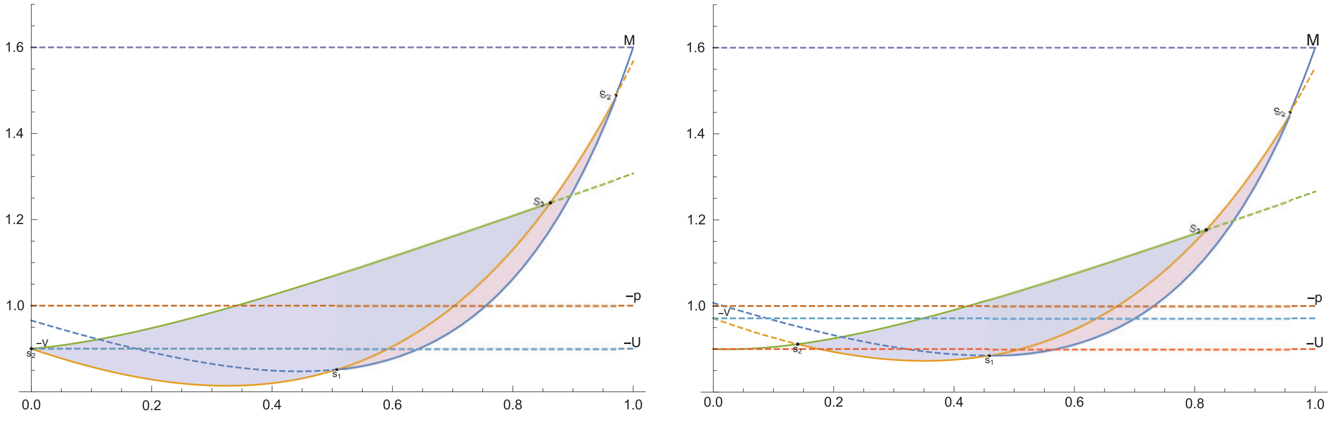
Lemma 10 establishes, for each choice of V satisfying Assumption 2 and each strongly ordered canonical optimal triple (Ψ, γ, δ) for ACBC$(\Lambda, \theta, \Theta, V, M, \kappa)$, conditions under which simple adjustments to Ψ and δ yield a strongly ordered triple $(\tilde{\Psi}, \gamma, \tilde{\delta})$ satisfying all the conditions of Proposition 2 except possibly the condition $\Delta_v(\tilde{\delta}) \le K_v$.

Example 1. To illustrate the arguments and the construction, we consider the problem with Λ consisting of the two drift rates $u(1) = -2$ and $u(2) = -1$ and the new, faster rate $v = u(3) = 2$. The problem parameters are $\sigma = \Theta = -h = -p = 1$, $U = -0.9$, $M = 1.6$, and $\kappa = 0.02$. Figure 3 illustrates the situations addressed in Lemma 10, in which we fix C_v so that $g(0, v, \gamma) = -U$ and consider strongly ordered optimal triples for the problem on $\Lambda = \{u(1), u(2)\}$ under different values for V, the cost of idling. In each case, the policy maintains drift rate $u(2)$ in the range from zero to S_2 and maintains the drift rate $u(1)$ in the range from s_1 to $\Theta = 1$. The partial derivatives of the relative value functions proving the optimality of the policy have $\delta(x, u(1))$ equal to $g(x, u(1), \gamma)$ (in blue) for $x > s_1$ and to $g(x, u(2), \gamma)$ (in orange) for $x \le s_1$ and $\delta(x, u(2))$ equal to $g(x, u(2), \gamma)$ for $x < S_2$ and to $g(x, u(1), \gamma)$ for $x \ge S_2$. The area between these functions, Δ_2 in red, is equal to $\kappa(u(2) - u(1))$.

Changing V changes the average cost γ, the value of $g(0, u(2), \gamma)$, and the positions of the points s_1 and S_2. Fixing C_v so that $g(0, v, \gamma) = -U$ defines $g(\cdot, v, \gamma)$ (in green). The point $S_v = S_3$ is the unique point of intersection of $g(\cdot, v, \gamma)$ and $\delta(\cdot, u(2))$ above the critical value $-p$. The point $s = s_2$ is the unique point of intersection of these two functions below the critical value $-p$.

Figure 3(a) illustrates the case initially, when V coincides with the original cost of idling. In this case, $s_2 = 0$, and Lemma 10 shows that, if $s_2 < s_1$ and $S_3 < S_2$, then defining $\delta(x, u(3))$ equal to $g(x, u(3), \gamma)$ (in green) for $x < S_3$, equal to $g(x, u(2), \gamma)$ (in orange) for $S_3 \le x < S_2$, and equal to $g(x, u(1), \gamma)$ (in blue) for $S_2 \le x$ yields a nondecreasing

Figure 3. Canonical Optimal Solutions for the Problem of Example 1 on $\Lambda = \{u(1), u(2)\}$ for Different Values of V, Illustrating the Constructions of Lemma 10



sequence of functions $\{\delta(\cdot,u(i)): i=1,2,3\}$ satisfying all the conditions of Proposition 2 except possibly that Δ_3, the area in blue, may be larger than $\kappa(u(3)-u(2))$. If $\Delta_3 \le \kappa(u(3)-u(2))$, Lemma 11 shows that these functions, together with an optimal policy for the problem on Λ, provide a canonical optimal solution for the problem on $\Lambda\cup\{v\}$. If $\Delta_3 > \kappa(u(3)-u(2))$, as is the case in Figure 3(a), we reduce V as illustrated in Figure 3(b).

Figure 3(b) illustrates the situation for a choice of V strictly between the limits p and U. Note that decreasing V, the adjusted cost of idling for drift rates $u(1)$ and $u(2)$, reduced γ, the average cost, and Δ_3, the area in blue. At the value of V shown in Figure 3(b), $\Delta_v = \kappa(u(3)-u(2))$, and Lemma 11 shows that defining $\delta(x,u(3))$ equal to $g(x,u(3),\gamma)$ (in green) for $x<S_3$, equal to $g(x,u(2),\gamma)$ (in orange) for $S_3 \le x < S_2$, and equal to $g(x,u(1),\gamma)$ (in blue) for $S_2 \le x$ and adjusting $\delta(x,u(2))$ and $\delta(x,u(1))$ to equal $g(x,u(3),\gamma)$ for $x<s_2$, yields a nondecreasing sequence $\tilde\delta = \{\delta(\cdot,u(i)): i=1,2,3\}$ of functions satisfying all the conditions of Proposition 2. Indeed, Lemma 11 shows that adjusting the policy for Λ to change from $u(2)$ to $u(3)$ at s_2, maintaining $u(3)$ in the range from zero to S_3, and changing to $u(2)$ at that point, yields an optimal policy, which, together with $\tilde\delta$, defines a canonical optimal solution on $\Lambda\cup\{v\}$.

Lemmas 11 and 12 ensure that there are values of the parameter V that satisfy the requirements of Lemma 10 and that either we can choose a value of V such that $\Delta_v(\tilde\delta)\le K_v$ and so $(\tilde\Psi,\gamma,\tilde\delta)$ is a strongly ordered canonical optimal triple for $\Lambda\cup\{v\}$ or, failing that, show how to extend the arguments of Lemma 10 to problems of the form ACBC$(\Lambda,o,\Theta,p,M,\kappa)$, where V is fixed at the minimum value p and the lower bound is increased to a value $o>\theta$. In this case, we show that there is a value of o such that similar adjustments to the associated strongly ordered optimal triple satisfy all the conditions of Proposition 2, including the condition on Δ_v, and so yield a strongly ordered canonical optimal triple for $\Lambda\cup\{v\}$.

Lemma 10. *Under Assumptions 1 and 2, given a drift rate $v\ne 0$ such that $\gamma\le\boldsymbol{\gamma}(v)$ and $v>u(n)=\max\Lambda$ and a strongly ordered canonical optimal triple (Ψ,γ,δ) for ACBC$(\Lambda,\theta,\Theta,V,M,\kappa)$, where $p\le V\le U$ satisfies Assumption 2, define*

$$C_v = e^{\frac{2v}{\sigma^2}\theta}\left(p-U+\frac{h\theta}{v}-\frac{h\sigma^2}{2v^2}-\frac{\gamma}{v}\right)$$

so that $g(\theta,v,\gamma)=-U$.

If $x_p(v,U,\theta,\gamma)$ exists and $x_p(v,U,\theta,\gamma)<x_p(n,V,\theta,\gamma)$, then (i) *there is a unique point $S_v\in(x_p(v,U,\theta,\gamma),\Theta]$ such that $g(S_v,v,\gamma)=\delta(S_v,n)$ and* (ii) *there is a unique point $s\in[\theta,x_p(v,U,\theta,\gamma))$ such that $g(s,v,\gamma)=\delta(s,1)$ and $g(x,v,\gamma)>\delta(x,n)$ for $x\in(s,S_v)$. Moreover, if $s<s_{n-1}$ and $S_v<S_n$, then* (iii) *$\tilde\Psi=\{\tilde\psi_u : u\in\Lambda\cup\{v\}\}$, where $\tilde\psi_v=(v,\theta,S_v,v,n)$, $\tilde\psi_n=(n,s,S_n,v,\tau(n))$, and $\tilde\psi_i=\psi_i$ for $i=1,2,\ldots,n-1$, is strongly ordered and $(\tilde\Psi,\gamma,\tilde\delta)$, where*

$$\tilde\delta(x,v)=\begin{cases} g(x,v,\gamma) & x<S_v\\ \delta(x,n) & S_v\le x\end{cases} \text{ and } \tilde\delta(x,i)=\begin{cases} g(x,v,\gamma) & x<s\\ \delta(x,i) & s\le x\end{cases} \text{ for } i=1,2,\ldots,n \tag{40}$$

satisfies all the conditions of Proposition 2 except possibly the condition $\Delta_v(\tilde\delta)\le K_v$. If $x_p(v,U,\theta,\gamma)\ge x_p(n,V,\theta,\gamma)$, then (iv) *$g(x,v,\gamma)\le\delta(x,n)$ for all $x\in[\theta,\Theta]$ and the inequality is strict except possibly at $x_p(n,V,\theta,\gamma)$.*

Proof. To simplify notation, write $x_p(v)$ in place of $x_p(v,U,\theta,\gamma)$, $x_p(n)$ in place of $x_p(n,V,\theta,\gamma)$, and $x_p(1)$ in place of $x_p(1,-M,\Theta,\gamma)$ (note that $x_p(1)$ is the solution to $g(x,1,\gamma)=-p$ with $\partial_x g(x,1,\gamma)\ge 0$, where $g(\Theta,1,\gamma)=M$). We first consider the case in which $x_p(v)$ exists and $x_p(v)<x_p(n)$.

To prove (i), observe that because (Ψ,γ,δ) is a strongly ordered optimal triple for ACBC$(\Lambda,\theta,\Theta,V,M,\kappa)$, $\delta(\Theta,n)=M$ and because $\gamma\le\boldsymbol{\gamma}(v)$, $g(\Theta,v,\gamma)\le M$ by Lemma 8. By assumption $x_p(v)<x_p(n)$, and so, by Lemma 4, $g(x_p(v),v,\gamma)=-p>\delta(x_p(v),n)$. Thus, we see that there is at least one point $x^*\in(x_p(v),\Theta]$ such that $g(x^*,v,\gamma)=\delta(x^*,n)$. Choose S_v to be the smallest such point so that $\partial_x g(S_v,v,\gamma)\le\partial_x\delta(S_v,n)$ and $g(S_v,v,\gamma)\ge -p$ by Lemma 1.

We first argue that $g(S_v,v,\gamma)>-p$, and so $\partial_x g(S_v,v,\gamma)<\partial_x\delta(S_v,n)$. To see this, assume that $g(S_v,v,\gamma)=-p$. Then, $\delta(S_v,n)=g(S_v,v,\gamma)=-p$, and by Lemma 4, S_v must equal $x_p(n)$ and so be smaller than S_n, which implies that $\rho_n(S_v)=n$. Thus, $-p=g(x_p(v),v,\gamma)=g(x_p(n),n,\gamma)=g(S_v,v,\gamma)$, and because $x_p(v)<x_p(n)=S_v$ and $\partial_x g(x_p(v),v,\gamma)\ge 0$, we see that $g(\cdot,v,\gamma)$ is concave and $\partial_x g(S_v,v,\gamma)<0$. Finally, by Lemma 1, $\partial_x g(S_v,n,\gamma)=\partial_x g(S_v,v,\gamma)=\partial_x g(x_p(n),n,\gamma)<0$, contradicting the fact that $\partial_x g(x_p(n),n,\gamma)>0$ by Lemma 2. Thus, we see that $-p<g(S_v,v,\gamma)=\delta(S_v,n)$, and so $\partial_x g(S_v,v,\gamma)<\partial_x\delta(S_v,n)$ by Lemma 1 and $S_v>x_p(n)$ by Lemma 4.

To see that S_v is unique, supposed to the contrary that $g(\cdot,v,\gamma)-\delta(\cdot,n)$ admits roots S_v, T_1, T_2, …, with $x_p(v)<S_v<T_1<T_2\ldots\le\Theta$. Observe that, because $\partial_x g(S_v,v,\gamma)<\partial_x\delta(S_v,n)$, it must be the case that $\partial_x g(T_1,v,\gamma)\ge\partial_x\delta(T_1,n)$, and so, by Lemma 1, $\delta(T_1,n)\le -p$ and, by Lemma 4, $T_1\le x_p(n)<S_v<T_1$, a contradiction.

To prove (ii), observe that, because (Ψ,γ,δ) is a strongly ordered optimal triple for ACBC$(\Lambda,\theta,\Theta,V,M,\kappa)$, $\delta(\theta,1)=-V$, and by construction, $g(\theta,v,\gamma)=-U\le -V=\delta(\theta,1)$. Because $x_p(v)<x_p(n)$, by assumption,

$g(x_p(v),\ v,\gamma) > \delta(x_p(v),n) \geq \delta(x_p(v),1)$, proving that there is a point $x^* \in [\theta, x_p(v))$ such that $g(x^*,v,\gamma) = \delta(x^*,1)$. For any such point s, the fact that $s < x_p(v) < x_p(n) \leq x_p(1)$ implies, by Lemma 4, that $g(s,v,\gamma) < -p$, and so, by Lemma 1, $\partial_x g(s,v,\gamma) > \partial_x \delta(s,1)$.

To see that s is unique, supposed to the contrary that $g(\cdot,v,\gamma) - \delta(\cdot,1)$ admits roots s, t_1, t_2, $\ldots$, with $x_p(v) > s > t_1 > t_2 \ldots \geq \theta$. Observe that, because $\partial_x g(s,v,\gamma) > \partial_x \delta(s,1)$, it must be the case that $\partial_x g(t_1,v,\gamma) \leq \partial_x \delta(t_1,1)$, and so, by Lemma 1, $g(t_1,v,\gamma) \geq -p$, contradicting the fact that any such root must satisfy $g(t_1,v,\gamma) < -p$.

To prove (iii), observe that, if $s < s_{n-1}$ and $S_v < S_n$, then $\rho_n(S_v) = n$ and the policy $\tilde{\Psi}$ is strongly ordered. Moreover, the triple $(\tilde{\Psi},\gamma,\tilde{\delta})$ satisfies all the conditions of Proposition 2 except possibly the condition $\Delta_v(\tilde{\delta}) \leq K_v$.

To prove (iv), we first consider the case $x_p(v) > x_p(n)$ and argue that $g(x,v,\gamma) < \delta(x,n)$ for all $x \in [\theta,\Theta]$. Let $x^* = x_p(n)$ and observe that $g(x^*,v,\gamma) < \delta(x^*,n) = -p$. To see that $g(x,v,\gamma) < \delta(x,n)$ for all $x \in [\theta,x^*)$, assume there is a point $z \in [\theta,x^*)$ such that $g(z,v,\gamma) = \delta(z,n)$ and choose z to be the largest such point. Then, $\partial_x g(z,v,\gamma) \leq \partial_x \delta(z,n)$, which, by Lemma 1, implies $\delta(z,n) \geq -p$, contradicting Lemma 4. Similarly, assume there is a point $S \in (x^*,\Theta]$ such that $g(S,v,\gamma) = \delta(S,n)$ and choose S to be the smallest such point. Then $\partial_x g(S,v,\gamma) \geq \partial_x \delta(S,n)$, which, by Lemma 1, implies $\delta(S,n) \leq -p$, contradicting Lemma 4.

We finally consider the case when $x_p(v) = x_p(n)$. Let $x^* = x_p(n)$ and observe that $\partial_{x,x}(g(x,v,\gamma) - g(x,n,\gamma)) = -2v/\sigma^2 \partial_x g(x,v,\gamma) + 2u(n)/\sigma^2 \partial_x g(x,n,\gamma)$, and because $\partial_x g(x^*,v,\gamma) = \partial_x g(x^*,n,\gamma) \geq 0$, by Lemma 1, $\partial_{x,x}(g(x^*,v,\gamma) - g(x^*,n,\gamma)) = -2(v-u(n))/\sigma^2 \partial_x g(x^*,n,\gamma) \leq 0$. Thus, we see that x^* is a local maximum of $g(\cdot,v,\gamma) - g(\cdot,n,\gamma)$ and the preceding arguments apply.

To argue the existence of a value for the parameter V that satisfies the requirements of Lemma 10 that $s < S_{n-1}$ and $S_v < S_n$ and enjoys the property that $\Delta_v(\tilde{\delta}) \leq K_v$, we require additional insights into the relationships between the parameter V and the integral Δ_v and, importantly, among the integrals Δ_i for different drift rates $u(i)$. Uncovering these relationships is not straightforward because $\Delta_v(\cdot)$ is a complicated function when $S_v > S_n$. We consider the simpler, but related function $D(v,u,s,z)$, where, for each pair of points s and z with $s < z$ and pair of drift rates u and v with $v > u$,

$$D(v,u,s,z) = \int_s^z g(x,v,\gamma) - g(x,u,\gamma)\, dx.$$

Note that, when (Ψ,γ,δ) is a strongly ordered canonical optimal triple for Λ, $D(i,i-1,s_{i-1},S_i) = \Delta_i(\delta)$ for each drift rate $i = 2,3,\ldots,n$. Note too that the stipulation in Lemma 10 that $x_p(v,U,\theta,\gamma) < x_p(n,V,\theta,\gamma)$ is sharp in the sense that, if $x_p(v,U,\theta,\gamma) \geq x_p(n,V,\theta,\gamma)$, $g(x,v,\gamma) \leq \delta(x,n)$ for all $x \in [\theta,\Theta]$ and Δ_v cannot be positive.

We show that, under certain conditions, $D(\cdot,u,s,z)$ exhibits a diseconomy of scale that proves useful in addressing Theorem 1. As the proof of Proposition 3 makes clear, $D(\cdot,u,s,z)$ does not exhibit incremental diseconomy; that is, it is not generally true that $\partial_v D(v,u,s,z) \leq 0$ for $v > u$. But the function does exhibit a specific kind of average diseconomy; that is, $D(v,u,s,z)/(v-u)$ is decreasing in v for $v > u$. We say that a function $f: \mathbb{R} \to \mathbb{R}$ such that $f(y)/(y-x)$ is decreasing in y for $y > x$ enjoys the *Tovey property* at x. We argue that, for each drift rate u, the function $D(\cdot,u,s,z)$, enjoys a weak form of the Tovey property at u. In particular, we prove that, when $s < z$ and $g(s,u,\gamma) = g(s,v,\gamma) < -p < g(x,u,\gamma)$ for some point $x \geq z$, $D(v,u,s,z)/(v-u)$ is decreasing in v for $v > u$ whenever $D(v,u,s,z) > 0$. The restriction that $g(s,u,\gamma) = g(s,v,\gamma) < -p < g(x,u,\gamma)$ for some point $x \geq z$ reflects the fact that, although we have no explicit expression for S_v, the point in (s,∞) at which $g(\cdot,v,\gamma)$ crosses $g(\cdot,u,\gamma)$, we do know that $g(S_v,u,\gamma) > -p$. Thus, Proposition 3 proves that, if there is a candidate $x \geq z$ for the point S_v, $D(v,u,s,z)/(v-u)$, as a function of v, enjoys this weak form of the Tovey property. The proof is long and technical and can be found in the appendix.

Proposition 3. *For each $u \neq 0$ and pair (s,z) with $s < z$, the function $D(v,u,s,z)/(v-u)$, where*

$$D(v,u,s,z) = \int_s^z g(x,v,\gamma) - g(x,u,\gamma)\, dx \tag{41}$$

is strictly decreasing in v so long as (i) *$v > u$,* (ii) *$g(s,u,\gamma) = g(s,v,\gamma) < -p$,* (iii) *$g(x,u,\gamma) \geq -p$ for some point $x \geq z$, and* (iv) *$D(v,u,s,z) > 0$. Similarly, for each pair (z,S) with $z < S$, the function $D(v,u,z,S)/(v-u)$ is strictly increasing in u so long as* (i) *$v > u$,* (ii) *$g(S,u,\gamma) = g(S,v,\gamma) > -p$,* (iii) *$g(x,v,\gamma) \leq -p$ for some point $x \leq z$, and* (iv) *$D(v,u,z,S) > 0$.*

Lemmas 11 and 12 constitute the heart of the proof of Theorem 1. They complete the induction argument by proving Theorem 1 is true for problems with $n+1$ drift rates, assuming it is true for problems with n drift rates. Lemma 11 considers, for each value of V with $p \leq V \leq U$ and starting with $V = U$, the construction described in Lemma 10. We argue that, if $\Delta_v \leq K_v$ when $V = U$, then v is transient, and adding it does not reduce the average cost. See Figure 3(a) for an example. Indeed, we exploit Proposition 3 to show that, if v is transient, all faster drift rates are transient as well. If $\Delta_v > K_v$ when $V = U$, we argue that Δ_v decreases as we decrease the parameter V. In this case, we again exploit Proposition 3 to show that, if $\boldsymbol{\gamma}(\Lambda, \theta, p) \geq h\theta$, there is a value V^* with $p < V^* < U$ such that $\Delta_v = K_v$, and at that point, $s < s_{n-1}$ and $S_v < S_n$. Thus, the construction of Lemma 10 yields a strongly ordered canonical optimal triple for $\Lambda \cup \{v\}$. In some cases, even though $\boldsymbol{\gamma}(\Lambda, \theta, p) < h\theta$, there is a value V^* with $p < V^* < U$ such that $\Delta_v = K_v$. This situation occurs, for example, if $x_p(v) \geq x_p(n)$ and so $\Delta_v \leq 0$ when $V = p$, from which we see that there must be a value V^* between U and p at which $\Delta_v = K_v$. See Figure 3(b) for an example. Thus, Lemma 11 addresses the cases in which either

- $\boldsymbol{\gamma}(\Lambda, \theta, p) \geq h\theta$, and so only adjustments to the cost of instantaneous controls at the lower bound θ are required, or
- $\boldsymbol{\gamma}(\Lambda, \theta, p) < h\theta$, and so adjustments to both the cost of instantaneous controls at the lower bound θ and, potentially, to the lower bound itself are possible, but because $x_p(v, U, \theta, \boldsymbol{\gamma}(\Lambda, \theta, p)) \geq x_p(n, p, \theta, \boldsymbol{\gamma}(\Lambda, \theta, p))$, it is not necessary to adjust the lower bound.

Lemma 12 completes the argument by addressing the situation in which $\boldsymbol{\gamma}(\Lambda, \theta, p) < h\theta$ and $x_p(v, U, \theta, \boldsymbol{\gamma}(\Lambda, \theta, p)) < x_p(n, p, \theta, \boldsymbol{\gamma}(\Lambda, \theta, p))$. In this case, adjustments to the lower bound θ may be required as well.

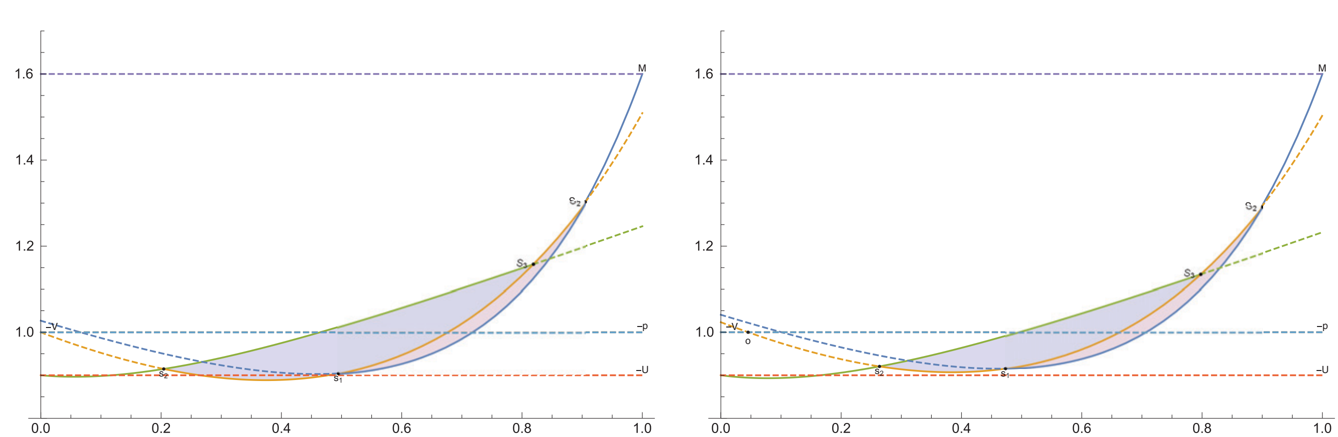
Example 2. To illustrate the cases addressed in Lemmas 11 and 12, we consider the same three-rate problem of Example 1 but with κ reduced to 0.01. Figure 4(a) illustrates the circumstances that require Lemma 12, and Figure 4(b) illustrates how Lemma 12 resolves those cases by increasing o, the lower limit on the process for the problem on $\Lambda = \{u(1), u(2)\}$.

Figure 4(a) illustrates the situation when $V = p$, its lower limit, and $o = \theta = 0$. Decreasing V reduces γ, the average cost, and Δ_3, the area in light gray. But even with $V = p$, its lowest value, Δ_3 is greater than $\kappa(u(3) - u(2))$. In this case, to observe the conditions of Assumption 2, we fix V at p and increase the lower limit o for the problem on $\Lambda = \{u(1), u(2)\}$. Note that, because $\gamma < 0$ and $\theta = 0$, $\gamma < h\theta$ and increasing o further reduces the average cost γ and the area Δ_3. Lemma 12 addresses this situation as illustrated in Figure 4(b).

Figure 4(b) illustrates the situation when $V = p$ and $o > \theta = 0$. Note that increasing o, the lower bound for the problem on $\Lambda = \{u(1), u(2)\}$, further reduces γ, the average cost, and Δ_3, the area in blue. At the value of o shown here, $\Delta_3 = \kappa(u(3) - u(2))$, and Lemma 12 shows that defining $\delta(x, u(3))$ equal to $g(x, u(3), \gamma)$ (in green) for $x < S_3$, equal to $g(x, u(2), \gamma)$ (in orange) for $S_3 \leq x < S_2$, and equal to $g(x, u(1), \gamma)$ (in blue) for $S_2 \leq x$ and adjusting $\delta(x, u(2))$ and $\delta(x, u(1))$ to equal $g(x, u(3), \gamma)$ for $x < s_2$ yields a nondecreasing sequence $\tilde{\delta} = \{\delta(\cdot, u(i)) : i = 1, 2, 3\}$ of functions satisfying all the conditions of Proposition 2. Indeed, Lemma 12 shows that adjusting the policy for Λ to change from $u(2)$ to $u(3)$ at s_2, maintaining $u(3)$ in the range from zero to S_3, and changing to $u(2)$ at that point yields an optimal policy, which, together with $\tilde{\delta}$, defines a canonical optimal solution on $\Lambda \cup \{v\}$.

Lemmas 11 and 12 require both that the new rate v be larger than the members of Λ and that $\boldsymbol{\gamma}(v)$, the average rate of the single-rate policy using only v, be at least as great as the optimal average cost using the rates in Λ. To ensure these conditions, especially as we add the first few rates, we start the process with a drift rate having the lowest single-rate average cost, that is, with a rate u such that $\boldsymbol{\gamma}(u)$ is minimum among all candidate drift rates. The arguments of Lemmas 11 and 12 do not describe how to add drift rates smaller than this initial drift rate. Rather than repeat these arguments for the case of $v < u(1)$, we observe that we can apply these arguments to the transformed problem on the set of drift rates $-\Lambda = \{-u : u \in \Lambda\}$ with holding cost $-h$, processing cost $-p$, and

Figure 4. Canonical Optimal Solutions for the Problem of Example 2 on $\Lambda = \{u(1), u(2)\}$ for Different Values of V, Illustrating the Constructions of Lemmas 11 and 12



costs of instantaneous controls interchanged, that is, equal to M at θ and U at Θ. Each control band policy $\Psi = \{(u, s_u, S_u, \beta(u), \tau(u)) : u \in -\Lambda\}$ with average cost γ for the transformed problem translates into a corresponding policy $\{(-u, \Theta - S_u, \Theta - s_u, -\tau(u), -\beta(u)) : u \in -\Lambda\}$ with average cost $\gamma + h\Theta$ for the original problem.

Lemma 11. *Under Assumptions 1 and 2, given a strongly ordered canonical optimal triple $(\Psi_V, \gamma_V, \delta_V)$ for $\mathrm{ACBC}(\Lambda, \theta, \Theta, V, M, \kappa)$ for each V with $p \le V \le U$ satisfying Assumption 2 and a drift rate $v \ne 0$ such that $\gamma_U \le \boldsymbol{\gamma}(v)$ and $v > u(n) = \max \Lambda$, define*

$$C_v(\gamma) = e^{\frac{2v}{\sigma^2}\theta}\left(p - U + \frac{h\theta}{v} - \frac{h\sigma^2}{2v^2} - \frac{\gamma}{v}\right)$$

so that $g(\theta, v, \gamma) = -U$. Then, (i) *$x_p(v, U, \theta, \gamma_U)$ exists and $x_p(v, U, \theta, \gamma_U) < x_p(n, U, \theta, \gamma_U)$,* (ii) *there is a unique point $S_v(U) \in (x_p(v, U, \theta, \gamma_U), \Theta]$ such that $g(S_v(U), v, \gamma_U) = \delta_U(S_v(U), n)$, and* (iii) *the point $s(U) = 0$ is the unique point in $[\theta, x_p(v, U, \theta, \gamma_U))$ such that $g(s(U), v, \gamma_U) = \delta_U(s(U), 1)$. Furthermore,* (iv) *$g(x, v, \gamma_U) > \delta_U(x, n)$ for $x \in (s(U), S_v(U))$.*

Let

$$\Delta_v(U) = \int_\theta^{S_v(U)} (g(x, v, \gamma_U) - \delta_U(x, n))\, dx.$$

If $\Delta_v(U) \le \kappa(v - u(n))$, then (v) *$S_v(U) < S_n(U)$ and $(\Psi_U \cup \psi(v), \gamma_U, \delta_U \cup \{\delta_v(\cdot, v)\})$, where*

$$\psi(v) = (v, \theta, S_v(U), v, n) \text{ and}$$
$$\delta(x, v) = \begin{cases} g(x, v, \gamma_U) & x < S_v(U) \\ \delta_U(x, n) & S_v(U) \le x \end{cases},$$

is a strongly ordered canonical optimal triple for $\mathrm{ACBC}(\Lambda \cup \{v\}, \theta, \Theta, U, M, \kappa)$, v is transient, and adding the new drift rate v does not decrease the average cost.

If $\Delta_v(U) > \kappa(v - u(n))$, then for each value V such that $p < V \le U$ and $x_p(v, U, \theta, \gamma_V) < x_p(n, V, \theta, \gamma_V)$, define $S_v(V)$ to be the unique point in $(x_p(v, U, \theta, \gamma_V), \Theta]$ such that $g(S_v(V), v, \gamma_V) = \delta_V(S_v(V), n)$ and define $s(V)$ to be the unique point in $[\theta, x_p(v, U, \theta, \gamma_V))$ such that $g(s(V), v, \gamma_V) = \delta_V(s(V), 1)$. Then, (vi) *either $\gamma_p < h\theta$ and $x_p(v, U, \theta, \gamma_p) < x_p(n, p, \theta, \gamma_p)$ or there is a value V^* with $p < V^* < U$ such that the strongly ordered canonical optimal triple $(\psi_{V^*}, \gamma_{V^*}, \delta_{V^*})$ for $\mathrm{ACBC}(\Lambda, \theta, \Theta, V^*, M, \kappa)$ satisfies*

$$s(V^*) < s_{n-1}(V^*)$$
$$S_v(V^*) < S_n(V^*) \text{ and}$$
$$\Delta_v(V^*) = \int_{s(V^*)}^{S_v(V^*)} (g(x, v, \gamma_{V^*}) - g(x, n, \gamma_{V^*}))\, dx = \kappa(v - u(n))$$

and so $(\Psi^*, \gamma_{V^*}, \delta^*)$ is a strongly ordered canonical optimal triple for $\mathrm{ACBC}(\Lambda \cup \{v\}, \theta, \Theta, U, M, \kappa)$, where $\Psi^* = \{\psi^*(u) : u \in \Lambda \cup \{v\}\}$ for all $x \in [\theta, \Theta]$ and

$$\psi^*(v) = (v, \theta, S_v(V^*), v, n),$$
$$\psi^*(n) = (n, s(V^*), S_n(V^*), v, \tau(n))$$
$$\psi^*(i) = \psi_{V^*}(i) \text{ for } i = 1, 2, \ldots, n-1$$

and $\delta^* = \{\delta^*(\cdot, u) : u \in \Lambda \cup \{v\}\}$, where

$$\delta^*(x, v) = \begin{cases} g(x, v, \gamma_{V^*}) & x < S_v(V^*) \\ \delta_{V^*}(x, n) & S_v(V^*) \le x \end{cases} \text{ and } \delta^*(x, i) = \begin{cases} g(x, v, \gamma_{V^*}) & x < s(V^*) \\ \delta_{V^*}(x, i) & s(V^*) \le x \end{cases}$$

for all $x \in [\theta, \Theta]$ and $i = 1, 2, \ldots, n$.

Proof. To simplify notation, write γ in place of γ_U, δ in place of δ_U, $x_p(v)$ in place of $x_p(v, U, \theta, \gamma_U)$, $x_p(n)$ in place of $x_p(n, U, \theta, \gamma_U)$ and S_n in place of $S_n(U)$, etc. We first argue that $x_p(v) < x_p(n)$. Because $-U = g(\theta, v, \gamma) = \delta(\theta, n) = g(\theta, n, \gamma)$, we have, by Lemma 1, that $\partial_x g(\theta, v, \gamma) - \partial_x g(\theta, n, \gamma) = 2/\sigma^2(p - U)(u(n) - v) > 0$. Because $\gamma \le \boldsymbol{\gamma}(v)$, we see, by Lemma 8, that $g(\Theta, v, \gamma) \le M = \delta(\Theta, n)$. So there is a point $S \in (\theta, \Theta]$ such that $g(S, v, \gamma) = \delta(S, n)$. Let S_v be the smallest such point, and let $u = \rho_n(S_v)$. Then, $\partial_x g(S_v, v, \gamma) \le \partial_x g(S_v, u, \gamma)$, which, by Lemma 1, implies that $g(S_v, v, \gamma) \ge -p$. We argue that $\partial_x g(S_v, v, \gamma) < \partial_x g(S_v, u, \gamma)$, and so $g(S_v, v, \gamma) > -p$. To see this, assume to the contrary that $\partial_x g(S_v, v, \gamma) = \partial_x g(S_v, u, \gamma)$, and so $g(S_v, v, \gamma) = -p$. Then, $S_n > S_v$ by Lemma 4, and so $u = u(n)$, $S_v = x_p(n)$, and $\partial_x g(S_v,n,\gamma) > 0$. But, by Lemma A.23, $\partial_x g(S_v,n,\gamma) < 0$, a contradiction. Thus, we see that $\partial_x g(S_v,v,\gamma) < \partial_x g(S_v,u,\gamma)$, $g(S_v,v,\gamma) > -p$, and $S_v > x_p(n)$. The fact that $g(\theta,v,\gamma) = g(\theta,n,\gamma) < -p$ and $g(x,v,\gamma) > g(x,n,\gamma)$ for $x \in (\theta, x_p(n)]$ implies that $x_p(v) < x_p(n)$, which, together with Lemma 10, proves (ii), (iii), and (iv).

To prove (v) and (vi), we consider the two cases separately: $u(n) \notin \mathcal{R}(\Psi_U)$ and $u(n) \in \mathcal{R}(\Psi_U)$.

For case $u(n) \notin \mathcal{R}(\Psi_U)$, when $u(n) \notin \mathcal{R}(\Psi_U)$, $u(n)$ is transient, and we argue that v is transient as well. In particular, the fact that $u(n)$ is transient implies $s_n = s_{n-1} = \theta$. We argue that $S_v < S_n$ and $\Delta_v(U) \le \kappa(v - u(n))$, and so the construction of Lemma 10 yields a strongly ordered canonical optimal triple proving that v is transient. Let $\omega = g(S_n,n,\gamma) = g(S_n,n-1,\gamma) > -p$ and observe that $\gamma(n,\theta,S_n,-U,\omega) = \gamma(n-1,\theta,S_n,-U,\omega) = \gamma$. Because $-\gamma(\cdot,\theta,\ S_n,-U,\omega)$ is unimodal by Lemma 7, $\gamma(n,\theta,S_n,-U,\omega) = \gamma(n-1,\theta,S_n,-U,\omega)$ and $v > u(n) > u(n-1)$, it follows that $\gamma(v,\theta,S_n,-U,\omega) > \gamma$, and so, by Lemma 8, $g(S_n,v,\gamma) < \omega$, proving that $S_v < S_n$. To see that $\Delta_v(U) \le \kappa(v-u(n))$, observe that, by Proposition 3, so long as it is positive, $D(r,n-1,\theta,S_v)/(r-u(n-1))$ is decreasing in r for $r > u(n-1)$ when we define C_r so that $g(\theta,r,\gamma) = -U$. In particular, because $v > u(n) > u$,

$$\begin{aligned}\frac{D(v,n-1,\theta,S_v)}{v-u(n-1)} &< \frac{D(n,n-1,\theta,S_v)}{u(n)-u(n-1)} = \frac{D(n,n-1,\theta,S_n)}{u(n)-u(n-1)} - \frac{D(n,n-1,S_v,S_n)}{u(n)-u(n-1)}\\ &\le \kappa - \frac{D(n,n-1,S_v,S_n)}{u(n)-u(n-1)} < \kappa,\end{aligned}$$

where the final inequality follows from the fact that $\theta = s_{n-1} < S_v < S_n$ and $g(x,n,\gamma) > g(x,n-1,\gamma)$ for all $x \in (s_{n-1},S_n)$. Thus, the construction of Lemma 10 yields a strongly ordered canonical optimal triple proving that v is transient.

For case, $u(n) \in \mathcal{R}(\Psi_U)$, we argue that, as we decrease V from U toward p, $S_v(V)$ falls below $S_n(V)$ before $s(V)$ reaches $s_{n-1}(V)$, and so between these two points, which we refer to as V_S and V_s, respectively, $s(V) < s_{n-1}(V)$ and $S_v(V) < S_n(V)$ as required for the construction of Lemma 10. We further argue that, for some value of $V \in (V_s,V_S)$, $\Delta_v(V) \le K_v$, and so the construction leads to a strongly ordered canonical optimal triple for $\Lambda \cup \{v\}$.

Clearly, γ_V is increasing in V. If $\gamma_p < h\theta$ and $x_p(v,U,\theta,\gamma_p) \ge x_p(n,p,\theta,\gamma_p)$, the fact that $x_p(v,U,\theta,\gamma_p) \ge x_p(n,p,\theta,\gamma_p)$ ensures that there is V_0 with $p \le V_0 < U$ such that $x_p(v,U,\theta,\gamma_{V_0}) = x_p(n,V_0,\theta,\gamma_{V_0})$ and $x_p(v,U,\theta,\gamma_V) < x_p(n,V,\theta,\gamma_V)$ for all $V \in (V_0,U]$. If $\gamma \ge h\theta$, we argue that there is V_0 with $p < V_0 < U$ such that $x_p(v,U,\theta,\gamma_{V_0}) = x_p(n,V_0,\theta,\gamma_{V_0})$ and $x_p(v,U,\theta,\gamma_V) < x_p(n,V,\theta,\gamma_V)$ for all $V \in (V_0,U]$. To see this, observe that, when $V = p$, $\gamma = \gamma_p$, $C_u(\gamma_p) = e^{2u/\sigma^2\theta}(h\theta/u - h\sigma^2/2u^2 - \gamma_p/u)$, $g(\theta,u,\gamma_p) = -p$, and $\partial_x g(\theta,u,\gamma_p) = 2/\sigma^2(\gamma_p - h\theta)$. Thus, if $\gamma_p \ge h\theta$, $x_p(u,p,\theta,\gamma_p) = \theta < x_p(v,U,\theta,\gamma_p)$, proving that there is $p < V_0 < U$ such that $x_p(v,U,\theta,\gamma_{V_0}) = x_p(n,V_0,\theta,\gamma_{V_0})$ and $x_p(v,U,\theta,\gamma_V) < x_p(n,V,\theta,\gamma_V)$ for all $V \in (V_0,U]$. Further, by Lemma 10 (the final paragraph in which we consider the case $x_p(v) = x_p(n)$), $g(x,v,\gamma_{V_0}) < g(x,n,\gamma_{V_0})$ for $x \in [\theta, x_p(n,V_0,\theta,\gamma_{V_0}))$. In particular, $g(s_{n-1}(V_0),v,\gamma_{V_0}) < g(s_{n-1}(V_0),n,\gamma_{V_0})$, and so $s(V_0) > s_{n-1}(V_0)$. It follows that there is $V_s \in (V_0,U]$ such that $s(V_s) = s_{n-1}(V_s)$ and $s(V) < s_{n-1}(V)$ for all $V \in (V_s,U]$. Similarly, if $S_v(U) \ge S_n(U)$, there is $V_S \in (V_0,U]$ such that $S_v(V_S) = S_n(V_S)$ and $S_v(V) > S_n(V)$ for all $V \in (V_S,U]$. Note that, if $S_v(U) < S_n(U)$, we define V_S to be U.

We argue that, if $S_v(U) \ge S_n(U)$, then $S_v(V_s) < S_n(V_s)$, and so $V_S > V_s$. Let $s = s(V_s) = s_{n-1}(V_s)$, and let $S = S_n(V_s)$, $\gamma = \gamma_{V_s}$, $\alpha = g(s,v,\gamma) = g(s,n,\gamma) = g(s,n-1,\gamma) < -p$, and $\omega = g(S,n,\gamma) = g(S,n-1,\gamma) > -p$. Because the two functions $g(\cdot,n,\gamma)$ and $g(\cdot,n-1,\gamma)$ cross at the points s and S, $\gamma = \gamma(n,s,S,\alpha,\omega) = \gamma(n-1,s,S,\alpha,\omega)$. But, by Lemma 7, $-\gamma(\cdot,s,S,\alpha,\omega)$ is unimodal, and so the fact that $\gamma(n,s,S,\alpha,\omega) = \gamma(n-1,s,S,\alpha,\omega)$ and $v > u(n) > u(n-1)$ implies that $\gamma(v,s,S,\alpha,\omega) > \gamma$. By Lemma 8, it follows that $g(S,v,\gamma) < \omega$, and so $S_v(V_s) < S = S_n(V_s)$.

We next argue that, for $V \in (V_s,U]$,

$$\Delta_v(V) = \int_{s(V)}^{S_v(V)} g(x,v,\gamma_V) - \delta_V(x,n)\,dx$$

is increasing in V. To prove this, we construct a second, related family of problems $\mathrm{ACBC}(\Lambda \cup \{v\},\theta,\Theta,U,M,\tilde{K}_V)$ and corresponding optimal policies. In particular, for each value V and pair $\mu,\eta \in \Lambda \cup \{v\}$, define

$$\tilde{K}_V(\mu,\eta) = \begin{cases} K(\mu,\eta) + \frac{1}{2}(\Delta_v(V) - \kappa(v-u(n))) & \text{if } \mu \ne \eta \text{ and } v \in \{\mu,\eta\} \\ K(\mu,\eta) & \text{otherwise}\end{cases}.$$

We show that γ_V is the optimal average cost for $\mathrm{ACBC}(\Lambda \cup \{v\},\theta,\Theta,U,M,\tilde{K}_V)$, and because γ_V is increasing in V, it must be the case that $1/2(\Delta_v(V) - \kappa(v-u(n)))$ is increasing in V, and so $\Delta_v(V)$ is increasing in V.

A modest generalization, the construction of Lemma 10 yields a canonical optimal triple $(\tilde{\psi}_V, \gamma_V, \tilde{\delta}_V)$ for ACBC$(\Lambda \cup \{v\}, \theta, \Theta, U, M, \tilde{K}_V)$; that is,

$$\tilde{\psi}_V(v) = \big(v, \theta, S_v(V), v, \rho_n(S_v(V))\big),$$
$$\tilde{\psi}_V(n) = (n, s(V), S_n(V), v, n-1), \text{ and}$$
$$\tilde{\psi}_V(i) = \psi_V(i) \text{ for } i = 1, 2, \ldots, n-1$$

and

$$\tilde{\delta}_V(x,v) = \begin{cases} g(x,v,\gamma_V) & \theta \le x < S_v(V) \\ \delta_V(x,n) & S_v(V) \le x \le \Theta \end{cases} \text{ and } \tilde{\delta}_V(x,i) = \begin{cases} g(x,v,\gamma_V) & \theta \le x < s(V) \\ \delta_V(x,i) & s(V) \le x \le \Theta \end{cases}$$

for $i = 1, 2, \ldots, n$. The generalization here is that, for $V > V_S$, $S_v(V) > S_n(V)$, and so $\rho_n(S_v(V)) < n$, so the triple is not strongly ordered because $\tau(v) < n$. Thus, for each value of the parameter V, the problem ACBC$(\Lambda \cup \{v\}, \theta, \Theta, U, M, \tilde{K}_V)$ shares the same average cost γ_V. Because only the changeover cost $\tilde{K}_V$ changes with V, we conclude that $\tilde{K}_V$ must be increasing in V.

We next argue that, if $S_v(V) \ge S_n(V)$, then $\Delta_v(V) > \kappa(v - u(n))$, and so, the contrapositive, if $\Delta_v(V) \le \kappa(v - u(n))$, then $S_v(V) < S_n(V)$, ensuring that the process ultimately yields a strongly ordered canonical optimal triple.

Because $V_S \ge V_s, s(V_S) \le s_{n-1}(V_S)$. To simplify notation, let $s = s(V_S)$, $s_{n-1} = s_{n-1}(V_S)$, $S_v = S_v(V_S)$, and $S_n = S_n(V_S)$ and fix $\gamma = \gamma_{V_S}$. Because $s \le s_{n-1}$, $D(v,n,s,S_v) \ge D(v,n,s_{n-1},S_v)$, and by Proposition 3, $D(v,r,s_{n-1},S_v)/(v-r)$ is strictly increasing in r for $r < v$. In particular, because $v > u(n) > u(n-1)$,

$$\frac{D(v,n,s_{n-1},S_v)}{v-u(n)} > \frac{D(v,n-1,s_{n-1},S_v)}{v-u(n-1)} = \frac{D(v,n,s_{n-1},S_v)}{v-u(n-1)} + \frac{D(n,n-1,s_{n-1},S_n)}{v-u(n-1)}$$
$$\ge \frac{D(v,n,s_{n-1},S_v)}{v-u(n-1)} + \kappa\frac{u(n)-u(n-1)}{v-u(n-1)},$$

and so

$$D(v,n,s_{n-1},S_v) > \kappa(v-u(n)).$$

It follows that, for $V \ge V_S$,

$$\Delta_v(V) \ge \Delta_v(V_S) \ge D(v,n,s,S_v(V_S)) \ge D(v,n,s_{n-1},S_v(V_S)) > \kappa(v-u(n)).$$

We finally argue that $\Delta_v(V_s) \le \kappa(v-u(n))$, and so, if $\Delta_v(U) > \kappa(v-u(n))$, there is a value $V^* \in (V_s, V_S)$ such that $\Delta_v(V^*) = \kappa(v-u(n))$, $s(V^*) < s_{n-1}(V^*)$, and $S_v(V^*) < S_n(V)$. To simplify notation, let $s = s(V_s)$, $S_v = S_v(V_s)$, and $S_n = S_n(V_s)$ and fix $\gamma = \gamma_{V_s}$. By Proposition 3,

$$\frac{D(v,n-1,s,S_v)}{v-u(n-1)} < \frac{D(n,n-1,s,S_v)}{u(n)-u(n-1)} \le \kappa - \frac{D(n,n-1,S_v,S_n)}{u(n)-u(n-1)},$$

and so

$$D(v,n-1,s,S_v) < \kappa(v-u(n-1)) - D(n,n-1,S_v,S_n)\frac{v-u(n-1)}{u(n)-u(n-1)}$$
$$\le \kappa(v-u(n-1)) - D(n,n-1,S_v,S_n).$$

Then,

$$D(v,n,s,S_v) < \kappa(v-u(n-1)) - D(n,n-1,S_v,S_n) - D(n,n-1,s,S_v)$$
$$= \kappa(v-u(n-1)) - D(n,n-1,s,S_n) = \kappa(v-u(n)).$$

Lemma 12 addresses the case in which $\gamma_p < h\theta$, and even reducing the cost of instantaneous controls at θ for drift rate $u(n)$ to the limit p leaves $x_p(v) < x_p(n)$ and $\Delta_v(p) > \kappa(v-u(n))$. We address this case via adjustments to the lower bound θ while holding the cost of pushing at the lower bound equal to the processing cost p. The arguments are analogous to those in Lemma 11, but we include them here for completeness.

Lemma 12. *Under Assumptions 1 and 2, suppose the drift rate $v \neq 0$ satisfies $\boldsymbol{\gamma}(\Lambda,\theta,U) \le \boldsymbol{\gamma}(v)$, $v > u(n) = \max\Lambda$, $\Delta_v(U) > \kappa(v-u(n))$, $\boldsymbol{\gamma}(\Lambda,\theta,p) < h\theta$, and $x_p(v,U,\theta,\boldsymbol{\gamma}(\Lambda,\theta,p)) < x_p(n,p,\theta,\boldsymbol{\gamma}(\Lambda,\theta,p))$. Define*

$$C_v(\gamma) = e^{\frac{2v\theta}{\sigma^2}}\left(p - U + \frac{h\theta}{v} - \frac{h\sigma^2}{2v^2} - \frac{\gamma}{v}\right)$$

so that $g(\theta,v,\gamma) = -U$. For each value $o \ge \theta$ such that $\boldsymbol{\gamma}(\Lambda,o,p) < hV$, let $(\Psi_o,\gamma_o,\delta_o)$ be a strongly ordered canonical optimal triple for $ACBC(\Lambda,o,\Theta,p,M,\kappa)$ and define $S_v(o)$ to be the unique point in $(x_p(v,U,\theta,\gamma_o),\Theta]$ such that $g(S_v(o),v,\gamma_o) = \delta_o(S_v(o),n)$ and $s(o)$ to be the unique point in $[\theta,x_p(v,U,\theta,\gamma_o))$ such that $g(s(o),v,\gamma_o) = \delta_o(s(o),1)$. There is a value o^* with $\theta < o^*$ such that the strongly ordered canonical optimal triple $(\Psi_{o^*},\gamma_{o^*},\delta_{o^*})$ for $ACBC(\Lambda,o^*,\Theta,p,M,\kappa)$ satisfies

$$\begin{aligned}
&s(o^*) < s_{n-1}(o^*)\\
&S_v(o^*) < S_n(o^*) \text{ and}\\
&\Delta_v(o^*) = \int_{s(o^*)}^{S_v(o^*)} g(x,v,\gamma_{o^*}) - g(x,n,\gamma_{o^*})\,dx = \kappa(v-u(n)),
\end{aligned}$$

and so $(\Psi^*,\gamma_{o^*},\delta^*)$ is a strongly ordered canonical optimal triple for $ACBC(\Lambda\cup\{v\},\theta,\Theta,U,M,\kappa)$, where $\Psi^* = \{\psi^*(u) : u \in \Lambda\cup\{v\}\}$,

$$\begin{aligned}
\psi^*(v) &= (v,\theta,S_v(o^*),v,n),\\
\psi^*(n) &= (n,s(o^*),S_n(o^*),v,\tau(n)) \text{ and},\\
\psi^*(i) &= \psi_{o^*}(i) \text{ for } i = 1,2,\ldots,n-1
\end{aligned}$$

and $\delta^* = \{\delta^*(\cdot,u) : u \in \Lambda\cup\{v\}\}$, where

$$\delta^*(x,v) = \begin{cases} g(x,v,\gamma_{o^*}) & x < S_v(o^*)\\ \delta_{o^*}(x,n) & S_v(o^*) \le x\end{cases} \text{ and } \delta^*(x,i) = \begin{cases} g(x,v,\gamma_{o^*}) & x < s(o^*)\\ \delta_{o^*}(x,i) & s(o^*) \le x\end{cases}$$

for all $x \in [\theta,\Theta]$ and $i = 1,2,\ldots,n$.

Proof. To simplify notation, write γ in place of γ_θ, δ in place of δ_θ, $x_p(v)$ in place of $x_p(v,U,\theta,\gamma_U)$, $x_p(n)$ in place of $x_p(n,U,\theta,\gamma_U)$, and S_n in place of $S_n(\theta)$, etc. By assumption, $x_p(v) < x_p(n)$. By Lemma 9, γ_o is decreasing in o. To see that there is o_0 with $\theta < o_0$ such that $x_p(v,U,\theta,\gamma_{o_0}) = x_p(n,p,o_0,\gamma_{o_0})$ and $x_p(v,U,\theta,\gamma_o) < x_p(u,p,o,\gamma_o)$ for all $o \in [\theta,o_0)$, observe that $\partial_x g(V,n,\gamma_o) = 2/\sigma^2(\gamma_o - hV)$. Define $\bar{o} = \inf\{o > \theta : \gamma_o \ge ho \text{ and } o \ge x_p(v,U,\theta,\gamma_o)\}$. Because $x_p(v,U,\theta,\gamma)$ is bounded above by Θ, $\bar{o} \le \Theta$, and either $\bar{o} = x_p(v,U,\theta,\gamma_{\bar{o}})$ or $\gamma_{\bar{o}} = h\bar{o}$. If $\gamma_{\bar{o}} < h\bar{o}$, then $x_p(n,p,\bar{o},\gamma_{\bar{o}}) > \bar{o} = x_p(v,U,\theta,\gamma_{\bar{o}})$, proving there is $o_0 \in (\theta,\bar{o})$ such that $x_p(v,U,\theta,\gamma_{o_0}) = x_p(n,p,o_0,\gamma_{o_0})$. If $o < x_p(v,U,\theta,\gamma_o)$ for all $o \in (\theta,\bar{o}]$, then $\gamma_{\bar{o}} = h\bar{o}$, and so $\partial_x g(\bar{o},n,\gamma_{\bar{o}}) = 2/\sigma^2(\gamma_{\bar{o}} - h\bar{o}) = 0$ and $x_p(n,p,\bar{o},\gamma_{\bar{o}}) = \bar{o} < x_p(v,U,\theta,\gamma_{\bar{o}})$, again proving there is $o_0 \in (\theta,\bar{o})$ such that $x_p(v,U,\theta,\gamma_{o_0}) = x_p(n,p,o_0,\gamma_{o_0})$. Further, $S_v(o_0) = x_p(n,p,o_0,\gamma_{o_0}) > s_{n-1}(o_0)$, and by Lemma A.23, we see that $g(\cdot,n,\gamma_{o_0})$ and $g(\cdot,v,\gamma_{o_0})$ cannot intersect on $[o_0,x_p(v,U,\theta,\gamma_{o_0}))$, and so $g(x,v,\gamma_{o_0}) < g(x,n,\gamma_{o_0})$ for $x \in [o_0,x_p(v,U,\theta,\gamma_{o_0}))$. In particular, $g(s_{n-1}(o_0),v,\gamma_{o_0}) < g(s_{n-1}(o_0),n,\gamma_{o_0})$, and so $s(o_0) > s_{n-1}(o_0)$. If follows that there is $o_s < o_0$ such that $s(o_s) = s_{n-1}(o_s)$ and $s(o) < s_{n-1}(o)$ for all $o < o_s$. Similarly, if $S_v \ge S_n$, there is $o_S < o_0$ such that $S_v(o_S) = S_n(o_S)$ and $S_v(o) > S_n(o)$ for all $o < o_S$. Note that, if $S_v < S_n$, we define o_S to be θ.

We argue that, if $S_v \ge S_n$, then $S_v(o_s) < S_n(o_s)$, and so $o_S < o_s$. Let $s = s(o_s) = s_{n-1}(o_s)$, and let $S = S_n(o_s)$, $\gamma = \gamma_{o_s}$, $\alpha = g(s,v,\gamma) = g(s,n,\gamma) = g(s,n-1,\gamma) < -p$, and $\omega = g(S,n,\gamma) = g(S,n-1,\gamma) > -p$. Then, $\gamma = \gamma(n,s,S,\alpha,\omega) = \gamma(n-1,s,S,\alpha,\omega)$. But, by Lemma 7, $-\gamma(\cdot,s,S,\alpha,\omega)$ is unimodal, and so the fact that $\gamma(n,s,S,\alpha,\omega) = \gamma(n-1,s,S,\alpha,\omega)$ and $v > u(n) > u(n-1)$ implies that $\gamma(v,s,S,\alpha,\omega) > \gamma$. By Lemma 8, it follows that $g(S,v,\gamma) < \omega$, and so $S_v(o_s) < S = S_n(o_s)$.

We next argue that, for $o < o_s$,

$$\Delta_v(o) = \int_{s(o)}^{S_v(o)} g(x,v,\gamma_o) - \delta_o(x,n)\,dx$$

is decreasing in o. To see this, we construct a canonical optimal triple for the problem $ACBC(\Lambda\cup\{v\},\theta,\Theta,U,M,\tilde{K}_o)$, where, for each value o and pair μ, $\eta \in \Lambda\cup\{v\}$,

$$\tilde{K}_o(\mu,\eta) = \begin{cases} K(\mu,\eta) + 1/2(\Delta_v(o) - \kappa(v-u(n))) & \text{if } \mu \neq \eta \text{ and } v \in \mu,\eta\\ K(\mu,\eta) & \text{otherwise}\end{cases}$$

and argue that, because γ_o is decreasing in o, it must be the case that $1/2(\Delta_v(o) - \kappa(v - u(n)))$ is decreasing in o, and so $\Delta_v(o)$ is decreasing in o. To construct a canonical optimal triple $(\tilde{\Psi}_o, \gamma_o, \tilde{\delta}_o)$ for $\mathrm{ACBC}(\Lambda \cup \{v\}, \theta, \Theta, U, M, \tilde{K}_o)$, let

$$\tilde{\psi}_o(v) = (v, \theta, S_v(o), v, \rho_n(S_v(o))),$$
$$\tilde{\psi}_o(n) = (n, s(o), S_n(o), v, n-1), \text{ and}$$
$$\tilde{\psi}_o(i) = \psi_o(i)\, for\ i = 1, 2, \ldots, n-1,$$

and let

$$\tilde{\delta}_o(x, v) = \begin{cases} g(x, v, \gamma_o) & \theta \le x < S_v(o) \\ \delta_o(x, n) & S_v(o) \le x \le \Theta \end{cases} \text{ and } \tilde{\delta}_o(x, n) = \begin{cases} g(x, v, \gamma_o) & \theta \le x < s(o) \\ \delta_o(x, i) & s(o) \le x \le \Theta \end{cases}$$

or $x \in [\theta, \Theta]$ and $i = 1, 2, \ldots, n$.

We next argue that, if $S_v \ge S_n$, then $\Delta_v > \kappa(v - u(n))$, and so, the contrapositive, if $\Delta_v \le \kappa(v - u(n))$, then $S_v < S_n$. Because $o_S \le o_s$, $s(o_S) \le s_{n-1}(o_S)$. To simplify notation, let $s = s(o_S)$, $s_{n-1} = s_{n-1}(o_S)$, $S_v = S_v(o_S)$, and $S_n = S_n(o_S)$ and fix $\gamma = \gamma_{o_S}$. Because $s \le s_{n-1}$, $D(v, n, s, S_v) \ge D(v, n, s_{n-1}, S_v)$, and by Proposition 3,

$$\begin{aligned} \frac{D(v, n, s_{n-1}, S_v)}{v - u(n)} &> \frac{D(v, n-1, s_{n-1}, S_v)}{v - u(n-1)} = \frac{D(v, n, s_{n-1}, S_v)}{v - u(n-1)} + \frac{D(n, n-1, s_{n-1}, S_n)}{v - u(n-1)} \\ &\ge \frac{D(v, n, s_{n-1}, S_v)}{v - u(n-1)} + \kappa \frac{u(n) - u(n-1)}{v - u(n-1)}, \end{aligned}$$

and so

$$D(v, n, s_{n-1}, S_v) > \kappa(v - u(n)).$$

It follows that

$$\Delta_v(\theta) \ge \Delta_v(o_S) \ge D(v, n, s, S_v(o_S)) \ge D(v, n, s_{n-1}, S_v(o_S)) > \kappa(v - u(n)).$$

We next argue that $\Delta_v(o_s) \le \kappa(v - u(n))$, and so, because $\Delta_v(\theta) > \kappa(v - u(n))$, there is a value $o^* \in (o_S, o_s)$ such that $\Delta_v(o^*) = \kappa(v - u(n))$, $s(o^*) < s_{n-1}(o^*)$, and $S_v(o^*) < S_n(o^*)$. To simplify notation, let $s = s(o_s)$, $S_v = S_v(o_s)$, and $S_n = S_n(o_s)$ and fix $\gamma = \gamma_{o_s}$. By Proposition 3,

$$\frac{D(v, n-1, s, S_v)}{v - u(n-1)} < \frac{D(n, n-1, s, S_v)}{u(n) - u(n-1)} \le \kappa - \frac{D(n, n-1, S_v, S_n)}{u(n) - u(n-1)},$$

and so

$$\begin{aligned} D(v, n-1, s, S_v) &< \kappa(v - u(n-1)) - D(n, n-1, S_v, S_n) \frac{v - u(n-1)}{u(n) - u(n-1)} \\ &\le \kappa(v - u(n-1)) - D(n, n-1, S_v, S_n). \end{aligned}$$

Then,

$$\begin{aligned} D(v, n, s, S_v) &< \kappa(v - u(n-1)) - D(n, n-1, S_v, S_n) - D(n, n-1, s, S_v) \\ &= \kappa(v - u(n-1)) - D(n, n-1, s, S_n) = \kappa(v - u(n)). \end{aligned}$$

5. Conclusions

In this paper, we show that the n-rate average cost Brownian control problem, under the assumption of proportional changeover costs, admits an optimal policy that is a unichain, strongly ordered policy. This characterization refines that in Ormeci Matoglu and Vande Vate (2011) for the more general problem of changeover costs that satisfy a triangle inequality by showing that, under the proportional cost assumption, there is at most one band for each drift rate and indeed exactly one band for each drift rate between the slowest recurrent rate and the fastest recurrent rate. We further characterize the transitions among these bands as strongly ordered. Our proof is constructive and leads to a practical algorithm for constructing canonical optimal solutions.

We compare the performance of that approach with the performance of methods described in Ormeci Matoglu et al. (2015) for the example described in Ormeci Matoglu and Vande Vate (2020): the average arrival rate is 26,400 units per day, and the standard deviation of the arrival process, σ, is 7,348.47 units per day. Each line has a processing rate of 10,560 units per day, so the set of drift rates Λ includes the appropriate members of

Table 1. Runtime Comparisons: Results Reported in Ormeci Matoglu et al. (2015) (Previous) vs. the Methods Described in This Paper with Different Levels of Computational Precision (as Measured by the Maximum Difference Between Δ_i and $\kappa(u(i)-u(i-1))$ and the Maximum Discontinuity in $\delta(\cdot,i)$ for Any Drift Rate i)

$\lvert\Lambda\rvert$	Runtime in seconds, previous	Runtime, seconds	Max errorsin (Δ,δ)	Higher quality run time, seconds	Higher quality max errors in (Δ,δ)
3	99	0.315	$(0.002, 1.10\times10^{-6})$	0.543	$(3.90\times10^{-7}, 2.03\times10^{-10})$
4	555	0.689	$(0.005, 0.00002)$	1.255	$(1.01\times10^{-6}, 3.47\times10^{-9})$
6	21,552	1.93	$(0.025, 0.00002)$	2.560	$(6.56\times10^{-6}, 3.33\times10^{-9})$
8	263,309	7.485	$(0.029, 0.00002)$	8.568	$(4.58\times10^{-6}, 3.47\times10^{-9})$
10	NA	16.16	$(0.030, 0.00004)$	17.399	$(4.65\times10^{-6}, 3.56\times10^{-9})$

$\{15840, 5280, -5280, -15840, -26400, -36960, -47520, -58080, -68640, -79200\}$ corresponding to one, two, and three up to 10 lines operating. Starting up a line costs approximately \$20,000 (note this is the cost of starting an existing line and does not include the initial investment cost to build the line), and shutting one down temporarily costs approximately \$2,000; thus, $\kappa_+ = 20{,}000/10{,}560 = 125/66$, $\kappa_- = 2{,}000/10{,}560 = 25/132$, and $\kappa = 275/132 = 2.08333$. The delay cost, h, is \$2.88 per unit per day; p, the operating cost, is $-\$18.18$. The maximum buffer capacity $= 40{,}000$ units; U, the cost of idling, is \$400 per unit; and M, the cost of turning away work, is \$1,000 per unit. Table 1 reports

- The solution time of the highest quality solution reported in Ormeci Matoglu et al. (2015).
- The solution time for a comparable solution via the methods here. In each case, the problem was solved from scratch.
- The magnitudes of the largest error in any Δ_i (compares with \$22,000 cost to start and stop a single line) and the largest discontinuity in the functions δ_i in that solution.
- The solution time for a higher quality solution via the methods here.
- The magnitudes of the largest error in any Δ_i and the largest discontinuity in the functions δ_i in that higher quality solution.

The computations reported in Ormeci Matoglu et al. (2015) and those reported for the methods developed here were run on the same machine (iMac with 2.9 GHz Intel Core i5), both using *Mathematica 10*.

An interactive implementation of the methods developed here for up to 10 drift rates is available at https://sites.gatech.edu/avg-cost/.

Acknowledgments

The author is indebted to Melda Ormeci-Matoglu and two anonomous referees for valuable suggestions on earlier drafts and to Jim Dai for introducing the problem and for his patient and enlightening explanations of Brownian control and Itô calculus. The author is also grateful to Michael E2 at stackexchange for helpful advice and guidance into the intricacies of Mathematica. Without that remarkable tool, this work would not have been possible.

Appendix

A.1. Proof of Lemma 3

Lemma 3 (Repeated for Convenience). *Suppose that, for some $S > s \geq 0$, $u \neq 0$, and γ, $g(s,u,\gamma) = -V$, where either $V < -p$ or $V = -p$ and $\gamma < hs$, and $g(S,u,\gamma) = \omega$, where either $\omega > -p$ or $\omega = -p$ and $\gamma < hS$, then*

$$x_p(u) = \begin{cases} s + \dfrac{\sigma^2}{2u}\log\left(\dfrac{\gamma - u(p-V)}{\gamma}\right) & \text{if } h = 0 \\ \dfrac{\gamma}{h} + \dfrac{\sigma^2}{2u}\left(1 + W_{\ell(hu)}\left(e^{-1-\frac{2u}{\sigma^2}\frac{\gamma}{h}+\frac{2u}{\sigma^2}s}\left(-1 - \dfrac{2u}{\sigma^2}\dfrac{\gamma}{h} + \dfrac{2u}{\sigma^2}s + \dfrac{2u^2}{h\sigma^2}(p-V)\right)\right)\right) & \text{if } h \neq 0, \end{cases}$$

where

$$\ell(x) = \begin{cases} 0 & \text{if } x < 0 \\ -1 & \text{if } x > 0, \end{cases}$$

and this the unique real value satisfying $g(x_p(u),u,\gamma) = -p$ and $\partial_x g(x_p(u),u,\gamma) \geq 0$.

Proof. Lemma 2 ensures that the system $g(x,u,\gamma) = -p$ and $\partial_x g(x,u,\gamma) \geq 0$ has a unique real solution, and this solution lies in (s,S). We first consider the case when $h = 0$. In this case, the unique solution to $g(x,u,\gamma) = -p + \gamma/u + e^{-2u/\sigma^2 x}C_u = -p$

is $x = \sigma^2/2u\log(-uC_u/\gamma)$, and the unique solution to $g(s,u,\gamma) = -V$ is given by $C_u = e^{2u/\sigma^2 s}(p - V - \gamma/u)$. Because $g(s,u,\gamma) = -V$ and $g(S,u,\gamma) = -p + \epsilon$ for some $\epsilon > 0$,

$$\gamma = \frac{u}{e^{\frac{2u}{\sigma^2}(S-s)} - 1}\left(V - p + e^{\frac{2u}{\sigma^2}(S-s)}\epsilon\right) > 0 \text{ and}$$
$$-uC_u = \frac{e^{\frac{2u}{\sigma^2}S}u}{e^{\frac{2u}{\sigma^2}(S-s)} - 1}(V - p + \epsilon) > 0,$$

proving that $x_p(u)$ is real. Finally, $\partial_x g(x,u,\gamma) = -uC_u 2/\sigma^2 e^{-2u/\sigma^2 x} > 0$ for all real x, proving that $\partial_x g(x_p(u),u,\gamma) > 0$.

We next consider the case $h \neq 0$. Let

$$z = e^{-1-\frac{2u}{\sigma^2}\frac{\gamma}{h}+\frac{2u}{\sigma^2}s}\left(-1 - \frac{2u}{\sigma^2}\frac{\gamma}{h} + \frac{2u}{\sigma^2}s + \frac{2u^2}{h\sigma^2}(p - V)\right).$$

We first argue that $z \geq -e^{-1}$ and, when $hu > 0$, $z < 0$, proving that $1 + W_{\ell(hu)}(z)$ is real and has the same sign as $-hu$. To see that $z \geq -e^{-1}$, it is enough to show that

$$e^{-1-\frac{2u}{\sigma^2}\frac{\gamma}{h}+\frac{2u}{\sigma^2}s+\frac{2u^2}{h\sigma^2}(p-V)}\left(-1 - \frac{2u}{\sigma^2}\frac{\gamma}{h} + \frac{2u}{\sigma^2}s + \frac{2u^2}{h\sigma^2}(p - V)\right) \geq -e^{-1+\frac{2u^2}{h\sigma^2}(p-V)}. \quad \text{(A.1)}$$

When $h < 0$,

$$e^{-1-\frac{2u}{\sigma^2}\frac{\gamma}{h}+\frac{2u}{\sigma^2}s+\frac{2u^2}{h\sigma^2}(p-V)}\left(-1 - \frac{2u}{\sigma^2}\frac{\gamma}{h} + \frac{2u}{\sigma^2}s + \frac{2u^2}{h\sigma^2}(p - V)\right) \geq -e^{-1} \geq -e^{-1+\frac{2u^2}{h\sigma^2}(p-V)}.$$

Here, the first inequality follows from the fact that $ze^z > -e^{-1}$ for all $z \in \mathbb{R}$, and the second inequality follows from the fact that $2u^2/h\sigma^2(p - V) > 0$.

When $h > 0$, $-e^{-1+2u^2/h\sigma^2(p-V)} \geq -e^{-1}$. We consider the two cases $u > 0$ and $u < 0$ separately.

When $u > 0$, $u\gamma \geq u^2(p - V) + uhs$ by Lemma A.1, and so $-1 - 2u\gamma/(\sigma^2 h) + 2us/\sigma^2 + 2u^2/(h\sigma^2)(p - V) \leq -1$. Thus, it is sufficient to show that

$$-1 - \frac{2u}{\sigma^2}\frac{\gamma}{h} + \frac{2u}{\sigma^2}s + \frac{2u^2}{h\sigma^2}(p - V) \leq W_{\ell(hu)}\left(-e^{-1+\frac{2u^2}{h\sigma^2}(p-V)}\right), \quad \text{(A.2)}$$

where $\ell(hu) = -1$. Condition (A.2) is equivalent to

$$\gamma \geq hs + u(p - V) - \frac{h\sigma^2}{2u}\left(1 + W_{\ell(hu)}\left(-e^{-1+\frac{2u^2}{h\sigma^2}(p-V)}\right)\right). \quad \text{(A.3)}$$

Similarly, when $u < 0$, $u\gamma \leq u^2(p - V) + uhs$ by Lemma A.1, and so $-1 - 2u\gamma/(\sigma^2 h) + 2us/\sigma^2 + 2u^2/(h\sigma^2)(p - V) \geq -1$. Thus, it is sufficient to show that

$$-1 - \frac{2u}{\sigma^2}\frac{\gamma}{h} + \frac{2u}{\sigma^2}s + \frac{2u^2}{h\sigma^2}(p - V) \geq W_{\ell(hu)}\left(-e^{-1+\frac{2u^2}{h\sigma^2}(p-V)}\right), \quad \text{(A.4)}$$

where $\ell(hu) = 0$. Condition (A.4) is equivalent to (A.3). Lemma A.1 confirms (A.3), proving that $z \geq -e^{-1}$.

To see that $z < 0$ when h and u have the same signs, observe that the system $g(s,u,\gamma) = -V$ and $g(S,u,\gamma) = -p + \epsilon$, where $\epsilon > 0$, uniquely defines the values

$$C_u = e^{\frac{2u}{\sigma^2}s}\left(\frac{hs}{u} + p - V - \frac{h\sigma^2}{2u^2} - \frac{\gamma}{u}\right) \quad \text{(A.5)}$$

and

$$\gamma = \frac{h(S - s) + u\epsilon - u(p - V)e^{-\frac{2u}{\sigma^2}(S-s)}}{1 - e^{-\frac{2u}{\sigma^2}(S-s)}} + hs - \frac{h\sigma^2}{2u}.$$

When $h < 0$ and $u < 0$, $hS < 0 < u(p - V)$ and $1 < e^{-2Su/\sigma^2}$, and when $h > 0$ and $u > 0$, $hS > 0 > u(p - V)$ and $1 > e^{-2u/\sigma^2 S}$. In both cases,

$$\frac{h(S - s) - u(p - V)e^{-\frac{2u}{\sigma^2}(S-s)}}{1 - e^{-\frac{2u}{\sigma^2}(S-s)}} + hs \geq hs + u(p - V)$$

from which it follows that

$$\gamma = \frac{h(S - s) + u\epsilon - u(p - V)e^{-\frac{2u}{\sigma^2}(S-s)}}{1 - e^{-\frac{2u}{\sigma^2}(S-s)}} + hs - \frac{h\sigma^2}{2u}$$
$$> \frac{h(S - s) - u(p - V)e^{-\frac{2u}{\sigma^2}(S-s)}}{1 - e^{-\frac{2u}{\sigma^2}(S-s)}} + hs - \frac{h\sigma^2}{2u} \geq hs + u(p - V) - \frac{h\sigma^2}{2u},$$

proving that $z < 0$.

To see that $g(x_p(u), u, \gamma) = -p$, observe that

$$\begin{aligned} g(x_p(u), u, \gamma) &= -p + \frac{h\sigma^2}{2u^2} - \frac{h}{u}x_p(u) + \frac{\gamma}{u} + e^{-\frac{2u}{\sigma^2}x_p(u)}C_u \\ &= -p + \frac{h\sigma^2}{2u^2} - \frac{h}{u}\left(\frac{\gamma}{h} + \frac{\sigma^2}{2u}(1 + W_{\ell(hu)}(z))\right) + \frac{\gamma}{u} + e^{-\frac{2u}{\sigma^2}\left(\frac{\gamma}{h}+\frac{\sigma^2}{2u}(1+W_{\ell(hu)}(z))\right)}C_u \\ &= -p - e^{-W_{\ell(hu)}(z)}\left(\frac{h\sigma^2}{2u^2}e^{W_{\ell(hu)}(z)}W_{\ell(hu)}(z) - e^{-\frac{2u}{\sigma^2}\left(\frac{\gamma}{h}+\frac{\sigma^2}{2u}\right)}C_u\right) \\ &= -p - e^{-W_{\ell(hu)}(z)}\left(\frac{h\sigma^2}{2u^2}z - e^{-1-\frac{2u}{\sigma^2}\frac{\gamma}{h}}C_u\right). \end{aligned}$$

The fact that $g(s, u, \gamma) = -V$ ensures that C_u is defined by (A.5), from which we see that $g(x_p(u), u, \gamma) = -p$.

To see that $\partial_x g(\gamma/h + \sigma^2/2u(1 + W_{\ell(hu)}(z)), u, \gamma) \geq 0$, observe that

$$\begin{aligned} \partial_x g\left(\frac{\gamma}{h} + \frac{\sigma^2}{2u}(1 + W_{\ell(hu)}(z)), u, \gamma\right) &= \frac{h}{u}\left(-1 - e^{-\frac{2u}{\sigma^2}\left(\frac{\gamma}{h}+\frac{\sigma^2}{2u}(1+W_{\ell(hu)}(z))\right)}e^{1+\frac{2u\gamma}{\sigma^2 h}}z\right) \\ &= \frac{h}{u}(-1 - e^{-W_{\ell(hu)}(z)}z) \\ &= \frac{h}{u}\left(-1 - e^{-W_{\ell(hu)}(z)}e^{W_{\ell(hu)}(z)}W_{\ell(hu)}(z)\right) \\ &= \frac{h}{u}(-1 - W_{\ell(hu)}(z)) \geq 0. \end{aligned}$$

Lemma A.1. *Suppose that, for some $S > s \geq 0$ and $u \neq 0$, γ and C_u are such that $g(s, u, \gamma) = -V < -p < g(S, u, \gamma) = \omega$. If $h > 0$, then*

$$\gamma \geq hs + \max\left\{u(p - V), u(p - V) - \frac{h\sigma^2}{2u}\left(1 + W_{\ell(hu)}\left(-e^{-1+\frac{2u^2(p-V)}{h\sigma^2}}\right)\right)\right\}.$$

If $h < 0$, then

$$\gamma \geq hS + \max\left\{u(\omega + p), u(\omega + p) - \frac{h\sigma^2}{2u}\left(1 + W_{\ell(hu)}\left(-e^{-1+\frac{2u^2(\omega+p)}{h\sigma^2}}\right)\right)\right\}.$$

Proof. We first consider the case $h > 0$. To see that $\gamma \geq u(p - V) + hs$, consider the function $\delta(x) = -V$ for $x \in \mathbb{R}$. Note that $\delta(s) = \delta(S) = -V < -p$ and

$$\frac{\sigma^2}{2}\delta'(x) + u\delta(x) + pu + hx = u(p - V) + hx \geq u(p - V) + hs,$$

which, by Proposition 2, proves $\gamma \geq u(p - V) + hs$. Similarly, define

$$\underline{\gamma} = hs + u(p - V) - \frac{h\sigma^2}{2u}\left(1 + W_{\ell(hu)}\left(-e^{-1+\frac{2u^2(p-V)}{h\sigma^2}}\right)\right) \text{ and } C_u = e^{\frac{2u}{\sigma^2}s}\left(\frac{hs}{u} + p - V - \frac{h\sigma^2}{2u^2} - \frac{\underline{\gamma}}{u}\right)$$

so that $g(s, u, \underline{\gamma}) = -V$ and, for each $x \in \mathbb{R}$,

$$\frac{\sigma^2}{2}\partial_x g(x, u, \underline{\gamma}) + ug(x, u, \underline{\gamma}) + pu + hx = \underline{\gamma}.$$

By Proposition 2, to prove that $\underline{\gamma}$ is a lower bound on γ, it is enough to show that $g(S, u, \underline{\gamma}) \leq -p$, that is, that

$$g\left(S, u, \underline{\gamma}\right) = \left(p - V + \frac{hs}{u} - \frac{h\sigma^2}{2u^2} - \frac{\underline{\gamma}}{u}\right)e^{-\frac{2u}{\sigma^2}(S-s)} - \frac{hS}{u} + \frac{\underline{\gamma}}{u} + \frac{h\sigma^2}{2u^2} - p \leq -p. \tag{A.6}$$

Condition (A.6) is equivalent to

$$\frac{2u}{\sigma^2}(S - s) + (1 - e^{-\frac{2u}{\sigma^2}(S-s)})W_{\ell(hu)}\left(-e^{-1+\frac{2u^2(p-V)}{h\sigma^2}}\right) \geq \frac{2u^2(p - V)}{h\sigma^2}. \tag{A.7}$$

Lemma A.2 proves (A.7) is satisfied in each case.

We next consider the case $h < 0$. To see that $\gamma \geq u(\omega + p) + hS$, consider the function $\delta(x) = \omega$ for $x \in \mathbb{R}$. Note that $\delta(s) = \delta(S) = \omega > -p$ and

$$\frac{\sigma^2}{2}\delta'(x) + u\delta(x) + pu + hx = u(\omega + p) + hx \geq u(\omega + p) + hS,$$

which, by Proposition 2, proves $\gamma \geq u(\omega+p)+hS$. Similarly, define

$$\underline{\gamma} = hS + u(\omega+p) - \frac{h\sigma^2}{2u}\left(1 + W_{\ell(hu)}\left(-e^{-1+\frac{2u^2(\omega+p)}{h\sigma^2}}\right)\right) \text{ and}$$
$$C_u = e^{\frac{2u}{\sigma^2}S}\left(\frac{hS}{u} + \omega + p - \frac{h\sigma^2}{2u^2} - \frac{\underline{\gamma}}{u}\right)$$

so that $g(S,u,\underline{\gamma}) = \omega$ and, for each $x \in \mathbb{R}$,

$$\frac{\sigma^2}{2}\partial_x g(x,u,\underline{\gamma}) + ug(x,u,\underline{\gamma}) + pu + hx = \underline{\gamma}.$$

By Proposition 2, to prove that $\underline{\gamma}$ is a lower bound on γ, it is enough to show that $g(s,u,\underline{\gamma}) \geq -p$, that is, that

$$g(s,u,\underline{\gamma}) = \left(\omega + p + \frac{hS}{u} - \frac{h\sigma^2}{2u^2} - \frac{\underline{\gamma}}{u}\right)e^{\frac{2u}{\sigma^2}(S-s)} - \frac{hs}{u} + \frac{\underline{\gamma}}{u} + \frac{h\sigma^2}{2u^2} - p \geq -p. \tag{A.8}$$

Condition (A.8) is equivalent to

$$-\frac{2u}{\sigma^2}(S-s) + \left(1 - e^{\frac{2u}{\sigma^2}(S-s)}\right)W_{\ell(hu)}\left(-e^{-1+\frac{2u^2(\omega+p)}{h\sigma^2}}\right) \geq \frac{2u^2(\omega+p)}{h\sigma^2}. \tag{A.9}$$

Lemma A.2 proves (A.9) is satisfied in each case.

Lemma A.2. *For each a, $b \in \mathbb{R}$, $b < 0$, and $\ell \in \{0,-1\}$,*

$$a + (1-e^{-a})W_\ell(-e^{-1+b}) - b \geq 0.$$

Proof. For a, $b \in \mathbb{R}$, $b < 0$, and $\ell \in \{0,-1\}$, define

$$q(a,b) = a + (1-e^{-a})W_\ell(-e^{-1+b}) - b \text{ and}$$
$$a^*(b) = -1 + b - W_\ell(-e^{-1+b}).$$

Observe that, for each $b<0$, $q(a^*(b),b) = \partial_a q(a^*(b),b) = 0$, and because $\partial_{a,a}q(a,b) = -e^{-a}W_\ell(-e^{-1+b}) > 0$, $q(\cdot,b)$ is convex and achieves its minimum value of zero at $a^*(b)$, proving that $q(a,b) \geq 0$ for all a, $b \in \mathbb{R}$ with $b<0$.

A.2. Proof of Lemma 7

Lemma 7 (Repeated for Convenience). *If $U \geq p \geq -M$, the function $-\boldsymbol{\gamma}(\cdot): \mathbb{R} \to \mathbb{R}$ is unimodal.*

Proof. First observe that, because $\gamma(u,\theta,-U,\Theta,M) = \gamma(u,0,-U,\Theta-\theta,M) + h\theta$, it is sufficient to consider the case in which $\theta = 0$. We first consider the case $U = p = -M$. In this case,

$$\boldsymbol{\gamma}(u) = h\Theta\left(\frac{1}{1-e^{-\frac{2u}{\sigma^2}\Theta}} - \frac{1}{\frac{2u}{\sigma^2}\Theta}\right).$$

Because $(1-e^{-y}) - y^{-1}$ is increasing in y, $\boldsymbol{\gamma}(u)$ is constant if $h=0$, increasing if $h>0$, and decreasing if $h<0$.

We next consider the case $M+U>0$. Observe that

$$\boldsymbol{\gamma}(u) = \frac{(M+U)\sigma^2}{2\Theta}\left(\frac{\frac{2u}{\sigma^2}\Theta}{1-e^{-\frac{2u}{\sigma^2}\Theta}} + a\frac{e^{-\frac{2u}{\sigma^2}\Theta} + \frac{2u}{\sigma^2}\Theta - 1}{\left(1-e^{-\frac{2u}{\sigma^2}\Theta}\right)\frac{2u}{\sigma^2}\Theta} - \frac{2u}{\sigma^2}\Theta b\right),$$

Where $a = 2h\Theta^2/(M+U)\sigma^2$, $b = U-p/M+U$, and $0 \leq b \leq 1$ by assumption. Because $\boldsymbol{\gamma}'(u) = (M+U)(q(2u/\sigma^2\Theta,a)-b)$, where

$$q(y,a) = \frac{y^2e^y(e^y-y-1) + aq_1(y)}{(e^y-1)^2y^2}$$

and

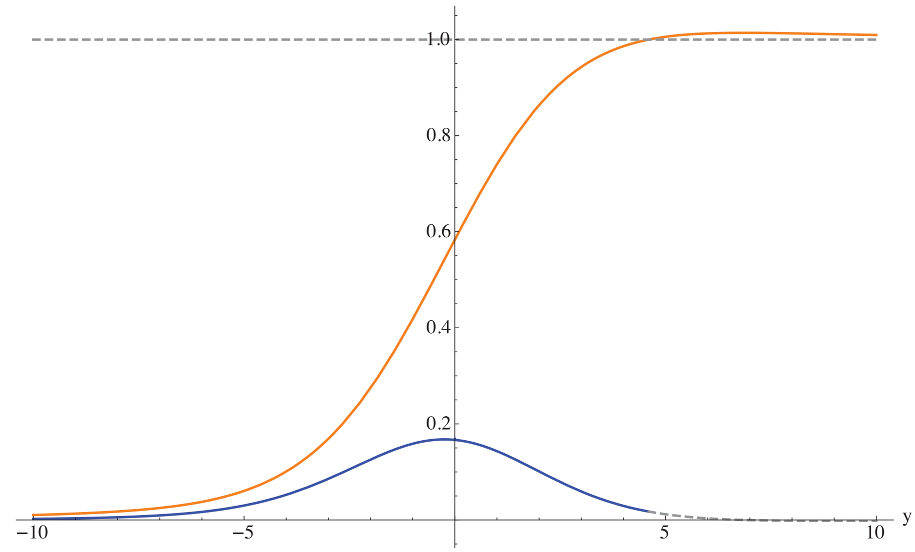
$$q_1(y) = e^{2y} - e^y\left(y^2+2\right) + 1,$$

u is a root of $\boldsymbol{\gamma}'(\cdot)$ if and only if $q(2u\Theta/\sigma^2\Theta,a) = b$. Further, because $\lim_{y\to-\infty}q(y,a) = 0$ and $\lim_{y\to\infty}q(y,a) = 1$, such a root exists if $0<b<1$.

To see that any root of $\boldsymbol{\gamma}'(\cdot)$ is unique, we argue that

$$\boldsymbol{\gamma}''(u) = \frac{2\Theta(M+U)}{\sigma^2}\partial_y q\left(\frac{2u}{\sigma^2}\Theta,a\right) > 0$$

Figure A.1. When $0 \le q(2u\Theta/\sigma^2\Theta, a) \le 1\, \boldsymbol{\gamma}''(u) > 0$ (Scaled by $\sigma^2/2\Theta(M+U) > 0$) and $\boldsymbol{\gamma}'(\cdot)$ Is Increasing, Hence, $\boldsymbol{\gamma}'(\cdot)$ Has a Unique Real Root u^*



for all u such that $0 \le q(2u/\sigma^2\Theta, a) \le 1$ (see Figure A.1), where

$$\partial_y q(y,a) = \frac{e^y y^3 (e^y(y-2) + y + 2) - a q_2(y)}{(e^y-1)^3 y^3} \text{ and}$$

$$q_2(y) = 2e^{3y} - e^{2y}\left(y^3 + 6\right) - e^y\left(y^3 - 6\right) - 2.$$

By Lemma A.3, $q_1(y) > 0$ for all $y \ne 0$, and so $0 \le q(y,a) \le 1$ if and only if

$$\frac{-y^2 e^y (e^y - y - 1)}{q_1(y)} \le a \le \frac{y^2(e^y(y-1)+1)}{q_1(y)}. \tag{A.10}$$

By Lemma A.4, $q_2(y)$ has the same sign as y. We consider separately the two cases: $y > 0$ and $y < 0$.

Case $y > 0$: When $0 \le q(y,a) \le 1$ and $y > 0, q_2(y) > 0$, and so

$$\begin{aligned}\partial_y q(y,a) &\ge \frac{e^y y^3 (e^y(y-2) + y + 2) - \dfrac{y^2(e^y(y-1)+1)}{q_1(y)} q_2(y)}{(e^y-1)^3 y^3} \\ &= \frac{e^{2y}\left(y^2 - 4y + 2\right) + e^y\left(y^3 + y^2 + 4y - 4\right) + 2}{(e^y-1) y q_1(y)}\end{aligned} \tag{A.11}$$

Because $(e^y - 1)y > 0$ and $q_1(y) > 0$ for all $y \ne 0$, the denominator of (A.11) is positive. By Lemma A.5, the numerator is positive as well.

Case $y < 0$: When $0 \le q(y,a) \le 1$ and $y > 0$, $q_2(y) < 0$, and so

$$\begin{aligned}\partial_y q(y,a) &\ge \frac{e^y y^3 (e^y(y-2) + y + 2) - \dfrac{-y^2 e^y(e^y - y - 1)}{q_1(y)} q_2(y)}{(e^y-1)^3 y^3} \\ &= \frac{e^y\left(2e^{2y} - e^y\left(y^3 - y^2 + 4y + 4\right) + y^2 + 4y + 2\right)}{(e^y-1) y q_1(y)}.\end{aligned}$$

Again, the denominator is positive, and by Lemma A.6, the numerator is positive as well.

Lemma A.3. *The function*

$$q_1(y) = e^{2y} - e^y\left(y^2 + 2\right) + 1$$

is strictly positive for all $y \ne 0$.

Proof. Observe that $q_1(0) = 0$ and $q_1'(y) = e^y(2e^y - y^2 - 2y - 2)$, which has the same sign as $q_{1.1}(y) = 2e^y - y^2 - 2y - 2$. Because $q_{1.1}(0) = 0$ and $q_{1.1}'(y) = 2(e^y - y - 1) > 0$ for all $y \neq 0$, we see that $q_{1.1}(y)$, and hence, also $q_1'(y)$ has the same sign as y, which, together with the fact that $q_1(0) = 0$, implies that $q_1(y) > 0$ for all $y \neq 0$.

Lemma A.4. *The function*

$$q_2(y) = 2e^{3y} - e^{2y}\left(y^3 + 6\right) - e^y\left(y^3 - 6\right) - 2$$

has the same sign as y.

Proof. Observe that $q_2(0) = 0$ and $q_2'(y) = e^y(6e^{2y} - e^y(2y^3 + 3y^2 + 12) - y^3 - 3y^2 + 6)$, which has the same sign as $q_{2.1}(y) = 6e^{2y} - e^y(2y^3 + 3y^2 + 12) - y^3 - 3y^2 + 6$. Further, $q_{2.1}(0) = 0$ and $q_{2.1}'(y) = 12e^{2y} - e^y(2y^3 + 9y^2 + 6y + 12) - 3y(y + 2)$. Continuing in this fashion, we see that

$$q_{2.1}'(0) = 0 \text{ and } q_{2.1}''(y) = 24e^{2y} - e^y\left(2y^3 + 15y^2 + 24y + 18\right) - 6(y + 1)$$
$$q_{2.1}''(0) = 0 \text{ and } q_{2.1}'''(y) = 48e^{2y} - e^y\left(2y^3 + 21y^2 + 54y + 42\right) - 6$$
$$q_{2.1}'''(0) = 0 \text{ and } q_{2.1}''''(y) = e^y\left(96e^y - 2y^3 - 27y^2 - 96y - 96\right),$$

which has the same sign as $q_{2.2}(y) = 96e^y - 2y^3 - 27y^2 - 96y - 96$. Further,

$$q_{2.2}(0) = 0 \text{ and } q_{2.2}'(y) = 96e^y - 6\left(y^2 + 9y + 16\right)$$
$$q_{2.2}'(0) = 0 \text{ and } q_{2.2}''(y) = 6(16e^y - 2y - 9).$$

Now, $q_{2.2}''(0) = 42$ and any root y of $q_{2.2}''(y)$ must satisfy

$$-8e^{-\frac{9}{2}} = -e^{-y-\frac{9}{2}}\left(y + \frac{9}{2}\right). \tag{A.12}$$

Because $-e^{-1} < -8e^{-9/2} < 0$, (A.12) has the two real solutions $y_0 = -9/2 - W_0(-8e^{-9/2}) \approx -4.40198$ and $y_{-1} = -9/2 - W_{-1}(-8e^{-9/2}) \approx -0.760486$. Because $\lim_{y\to-\infty} q_{2.2}''(y) = \lim_{y\to\infty} q_{2.2}''(y) = \infty$, we see that $q_{2.2}''(y) < 0$ for $y_0 < y < y_{-1}$ and $q_{2.2}''(y) \geq 0$ otherwise.

Because $\lim_{y\to-\infty} q_{2.2}'(y) = -\infty$, $q_{2.2}'(y_0) = 51/2 - 6W_0(-8e^{-9/2})(2 + W_0(-8e^{-9/2})) \approx 26.6186$, $q_{2.2}'(y_{-1}) = 51/2 - 6W_{-1}(-8e^{-9/2})(2 + W_{-1}(-8e^{-9/2})) \approx -13.5296$, and $\lim_{y\to\infty} q_{2.2}'(y) = \infty$, we see that $q_{2.2}'(y)$ has three real roots: $y_1 \approx -6.567$, $y_2 \approx -1.94154$, and zero.

Repeating this analysis for $q_{2.2}(\cdot)$, we see that, because $\lim_{y\to-\infty} q_{2.2}(y) = \infty$, $q_{2.2}(y_1) \approx -63.4111$, $q_{2.2}(y_2) \approx 17.0212$, $q_{2.2}(0) = 0$, and $\lim_{y\to\infty} q_{2.2}(y) = \infty$, we see that $q_{2.2}(\cdot)$, and so also $q_{2.1}''''(\cdot)$ has three real roots $y_3 \approx -8.85478$, $y_4 \approx -3.40428$, and zero.

Likewise, because $\lim_{y\to-\infty} q_{2.1}'''(y) = -6$, $q_{2.1}'''(y_3) \approx -5.97395$, $q_{2.1}'''(y_4) \approx -6.69918$, $q_{2.1}'''(0) = 0$, and $\lim_{y\to\infty} q_{2.1}'''(y) = \infty$, we see that $q_{2.1}'''(y)$ has the unique real root zero and has the same sign as y. It follows that $q_{2.1}''(y) > 0$ for $y \neq 0$, and so $q_{2.1}'(y)$ has the same sign as y and $q_{2.1}(y) > 0$ for $y \neq 0$, which implies $q_2(y)$ has the same sign as y.

Lemma A.5. *The function*

$$q_3(y) = e^{2y}\left(y^2 - 4y + 2\right) + e^y\left(y^3 + y^2 + 4y - 4\right) + 2$$

is strictly positive for all $y \neq 0$.

Proof. Observe that $q_3(0) = 0$ and $q_3'(y) = e^y y(2e^y(y - 3) + y^2 + 4y + 6)$, which has the same sign as $yq_{3.1}(y)$, where $q_{3.1}(y) = 2e^y(y - 3) + y^2 + 4y + 6$. Note that $q_{3.1}(0) = 0$ and $q_{3.1}'(y) = 2(e^y(y - 2) + y + 2)$. Further, $q_{3.1}'(0) = 0$ and $q_{3.1}''(y) = 2(e^y(y - 1) + 1)$. Finally, observe that $q_{3.1}''(0) = 0$ and $q_{3.1}'''(y) = 2e^y y$, which has the same sign as y. Thus, we see that $q_{3.1}''(y) > 0$ for $y \neq 0$, and so $q_{3.1}'(y)$ has the same sign as y, which implies $q_{3.1}(y) > 0$ for $y \neq 0$, and so $yq_{3.1}(y)$ and $q_3'(y)$ have the same sign as y. Thus, we conclude that $q_3(y) > 0$ for $y \neq 0$.

Lemma A.6. *The function*

$$q_4(y) = 2e^{2y} - e^y\left(y^3 - y^2 + 4y + 4\right) + y^2 + 4y + 2$$

is strictly positive for all $y \neq 0$.

Proof. Observe that

$$q_4(0) = 0 \text{ and } q_4'(y) = 4e^{2y} - e^y\left(y^3 + 2y^2 + 2y + 8\right) + 2(y + 2),$$
$$q_4'(0) = 0 \text{ and } q_4''(y) = 8e^{2y} - e^y\left(y^3 + 5y^2 + 6y + 10\right) + 2,$$
$$q_4''(0) = 0 \text{ and } q_4'''(y) = e^y\left(16e^y - y^3 - 8y^2 - 16y - 16\right),$$

which has the same sign as $q_{4.1}(y) = 16e^y - y^3 - 8y^2 - 16y - 16$. Now,

$$q_{4.1}(0) = 0 \text{ and } q'_{4.1}(y) = 16e^y - 3y^2 - 16y - 16,$$
$$q'_{4.1}(0) = 0 \text{ and } q''_{4.1}(y) = 2(8e^y - 3y - 8).$$

Any root of $q''_{4.1}(\cdot)$ must satisfy $-e^{-y-8/3}(y+8/3) = -8/3e^{-8/3}$. Because $-e^{-1} < -8/3e^{-8/3} < 0$, this system has the two real solutions $y_0 = -8/3 - W_0(-8/3e^{-8/3})$ and $y_{-1} = -8/3 - W_{-1}(-8/3e^{-8/3}) = 0$. Because $\lim_{y\to-\infty} q'_{4.1}(y) = -\infty$, $q'_{4.1}(y_0) = 16/3 - 3W_0(-8/3e^{-8/3})(2 + W_0(-8/3e^{-8/3})) \approx 6.5739$, $q'_{4.1}(0) = 0$, and $\lim_{y\to\infty} q'_{4.1}(y) = \infty$, we see that $q'_{4.1}(\cdot)$ has the two real roots $y_1 \approx -4.03492$ and zero. Similarly, because $\lim_{y\to-\infty} q_{4.1}(y) = \infty$, $q_{4.1}(y_1) \approx -15.7121$, $q_{4.1}(0) = 0$, and $\lim_{y\to\infty} q_{4.1}(y) = \infty$, we see that $q_{4.1}(\cdot)$, and so also $q'''_4(\cdot)$ has the two real roots $y_2 \approx -5.67607$ and zero. Finally, because $\lim_{y\to-\infty} q''_4(y) = 2$, $q''_4(y_2) \approx 2.15718$, $q''_4(0) = 0$, and $\lim_{y\to\infty} q''_4(y) = \infty$, we see that $q''_4(\cdot)$ has the unique real root zero and is strictly positive for $y \neq 0$. It follows that $q'_4(y)$ has the same sign as y and $q_4(y) > 0$ for $y \neq 0$.

A.3. Proof of Proposition 3

Proposition 3 (Repeated for Convenience).
For each $u \neq 0$ and pair (s,z) with $s < z$, the function $D(v,u,s,z)/(v-u)$, where

$$D(v,u,s,z) = \int_s^z g(x,v,\gamma) - g(x,u,\gamma)\,dx$$

is strictly decreasing in v so long as (i) *$v > u$,* (ii) *$g(s,u,\gamma) = g(s,v,\gamma) < -p$,* (iii) *$g(x,u,\gamma) \geq -p$ for some point $x \geq z$, and* (iv) *$D(v,u,s,z) > 0$. Similarly, for each pair (z,S) with $z < S$, the function $D(v,u,z,S)/(v-u)$ is strictly increasing in u so long as* (i) *$v > u$,* (ii) *$g(S,u,\gamma) = g(S,v,\gamma) > -p$,* (iii) *$g(x,v,\gamma) \leq -p$ for some point $x \leq z$, and* (iv) *$D(v,u,z,S) > 0$.*

Proof. We first consider the function $D(v,u,s,z)/(v-u)$, where $s < z$, $u < v$, $V = g(s,u,\gamma) = g(s,v,\gamma) < -p$, and $g(z,u,\gamma) = \omega$. We argue that, in this case,

$$(v-u)^2 \partial_v\left(\frac{D(v,u,s,z)}{v-u}\right) = \partial_v D(v,u,s,z)(v-u) - D(v,u,s,z) < 0 \tag{A.13}$$

whenever $D(v,u,s,z) > 0$. Let $V = g(s,u,\gamma) = g(s,v,\gamma)$, where $V < -p$, and then

$$C_\mu = e^{\frac{2\mu}{\sigma^2}s}\left(p + V - \frac{\gamma}{\mu} + \frac{hs}{\mu} - \frac{h\sigma^2}{2\mu^2}\right)$$

for $\mu \in \{u,v\}$ and

$$\int g(x,\mu,\gamma)\,dx = -\frac{h}{2\mu}x^2 + \left(-p + \frac{h\sigma^2}{2\mu^2} + \frac{\gamma}{\mu}\right)x - \frac{\sigma^2}{2\mu}e^{-\frac{2\mu}{\sigma^2}x}C_\mu$$

from which we conclude that

$$\int_s^z g(x,\mu,\gamma)\,dx = (z-s)^3\left(\alpha(\mathfrak{f}(\mu))\frac{p+V}{(z-s)^2} + \chi(\mathfrak{f}(\mu))\frac{h}{\sigma^2} + \eta(\mathfrak{f}(\mu))\frac{2(\gamma - hs)}{\sigma^2(z-s)} - \frac{p}{(z-s)^2}\right),$$

where

$$\mathfrak{f}(\mu) = \frac{2\mu}{\sigma^2}(z-s),$$
$$\alpha(y) = (1 - e^{-y})/y$$
$$\eta(y) = (e^{-y} + y - 1)/y^2 \text{ and}$$
$$\chi(y) = \left(2e^{-y} - y^2 + 2y - 2\right)/y^3.$$

Note that we use the scaling $\mathfrak{f}(\cdot)$ as a terse way to maintain the relationship with the drift rates u and v, but we generally use y for $\mathfrak{f}(v)$ and w for $\mathfrak{f}(u)$.

Thus, we see that

$$D(v,u,s,z) = (z-s)^3\big(\eta(\mathfrak{f}(v)) - \eta(\mathfrak{f}(u))\big)\left(P(\mathfrak{f}(v),\mathfrak{f}(u))\frac{p+V}{(z-s)^2} + \Phi(\mathfrak{f}(v),\mathfrak{f}(u))\frac{h}{\sigma^2} + \frac{2(\gamma - hs)}{\sigma^2(z-s)}\right),$$

where

$$P(y,w) = (\alpha(y) - \alpha(w))/(\eta(y) - \eta(w)) \text{ and}$$
$$\Phi(y,w) = (\chi(y) - \chi(w))/(\eta(y) - \eta(w)).$$

Note that, by Lemma A.10, $\eta(\cdot)$ is decreasing, and so $D(v,u,s,z) > 0$ if and only if

$$P(\mathfrak{f}(v),\mathfrak{f}(u))\frac{p+V}{(z-s)^2}+\Phi(\mathfrak{f}(v),\mathfrak{f}(u))\frac{h}{\sigma^2}+\frac{2(\gamma-hs)}{\sigma^2(z-s)}<0. \tag{A.14}$$

By Lemmas A.9–A.11,

$$\alpha'(y) = -a(y)/y^2$$
$$\eta'(y) = -b(y)/y^2 \text{ and}$$
$$\chi'(y) = 2c(y)/y^2,$$

where

$$a(y) = 1-e^{-y}-ye^{-y}$$
$$b(y) = (y-2+e^{-y}(y+2))/y \text{ and}$$
$$c(y) = -\left(2y-3-y^2/2+3e^{-y}+ye^{-y}\right)/y^2.$$

It follows that

$$\begin{aligned}\partial_v D(v,u,s,z) &= (z-s)^3\frac{2(z-s)}{\sigma^2}\left(\alpha'(\mathfrak{f}(v))\frac{p+V}{(z-s)^2}+\chi'(\mathfrak{f}(v))\frac{h}{\sigma^2}+\eta'(\mathfrak{f}(v))\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right)\\ &= (z-s)^3\frac{2}{\sigma^2}(z-s)\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}\left(\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2}+\phi(\mathfrak{f}(v))\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right),\end{aligned}$$

where

$$\rho(y) = a(y)/b(y) \text{ and}$$
$$\phi(y) = c(y)/b(y).$$

We argue that

$$\begin{aligned}\partial_v D(v,u,s,z)(v-u)-D(v,u,s,z) &= (z-s)^3\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}(\mathfrak{f}(v)-\mathfrak{f}(u))\left(\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2}+\phi(\mathfrak{f}(v))\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right)\\ &\quad+(z-s)^3(\eta(\mathfrak{f}(v))-\eta(\mathfrak{f}(u)))\left(P(\mathfrak{f}(v),\mathfrak{f}(u))\frac{-(p+V)}{(z-s)^2}-\Phi(\mathfrak{f}(v),\mathfrak{f}(u))\frac{h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right)<0\end{aligned}$$

whenever (A.14) is satisfied.

We consider the two cases: $h \le 0$ and $h > 0$.

Case $h \le 0$: First note that, if $\partial_v D(v,u,s,z) \le 0$, (A.13) is trivially satisfied when $D(v,u,s,z) > 0$. Thus, we focus on the case in which $\partial_v D(v,u,s,z) > 0$.

Because $b(y)$ is strictly positive for all $y \ne 0$, by Lemma A.7, $\partial_v D(v,u,s,z) > 0$ if and only if

$$\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2}+\phi(\mathfrak{f}(v))\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}>0. \tag{A.15}$$

By Lemma A.12, $P(y,w) > \rho(y)$ for all $y > w$, and by Lemma A.17, $\Phi(y,w) > -2\phi(y)$ for all $y > w$. Because $-p+V/(z-s)^2 > 0$ and $h \le 0$,

$$P(y,w)\frac{-(p+V)}{(z-s)^2}-\Phi(y,w)\frac{h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}>\rho(y)\frac{-(p+V)}{(z-s)^2}+\phi(y)\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}>0$$

for all nonzero $y > w$. Further, because $\eta(\cdot)$ is decreasing by Lemma A.10,

$$\begin{aligned}\partial_v D(v,u,s,z)(v-u)-D(v,u,s,z) &= (z-s)^3\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}(\mathfrak{f}(v)-\mathfrak{f}(u))\left(\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2}+\phi(\mathfrak{f}(v))\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right)\\ &\quad+(z-s)^3(\eta(\mathfrak{f}(v))-\eta(\mathfrak{f}(u)))\left(P(\mathfrak{f}(v),\mathfrak{f}(u))\frac{-(p+V)}{(z-s)^2}-\Phi(\mathfrak{f}(v),\mathfrak{f}(u))\frac{h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right)\\ &< (z-s)^3\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}(\mathfrak{f}(v)-\mathfrak{f}(u))\left(\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2}+\phi(\mathfrak{f}(v))\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right)\\ &\quad+(z-s)^3(\eta(\mathfrak{f}(v))-\eta(\mathfrak{f}(u)))\left(\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2}+\phi(\mathfrak{f}(v))\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right)\\ &= (z-s)^3\left(\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}(\mathfrak{f}(v)-\mathfrak{f}(u))+\eta(\mathfrak{f}(v))-\eta(\mathfrak{f}(u))\right)\left(\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2}+\phi(\mathfrak{f}(v))\frac{2h}{\sigma^2}-\frac{2(\gamma-hs)}{\sigma^2(z-s)}\right).\end{aligned}$$

Finally, because

$$\varphi(y,w) = \frac{b(y)}{y^2}(y-w) + \eta(y) - \eta(w) < 0 \text{ for all } y > w,$$

by Lemma A.15, we see that, when $\partial_v D(v,u,s,z) > 0$,

$$\partial_v D(v,u,s,z)(v-u) - D(v,u,s,z) < (z-s)^3 \varphi(\mathfrak{f}(v),\mathfrak{f}(u))\left(\rho(\mathfrak{f}(v))\frac{-(p+V)}{(z-s)^2} + \phi(\mathfrak{f}(v))\frac{2h}{\sigma^2} - \frac{2(\gamma-hs)}{\sigma^2(z-s)}\right) < 0.$$

Case $h>0$: When $h>0$, we wish to show that, for $v>u$,

$$\begin{aligned}\partial_v D(v,u,s,z)(v-u) - D(v,u,s,z) &= (z-s)^3\frac{h}{\sigma^2}\frac{b(\mathfrak{f}(v))}{(\mathfrak{f}(v))^2}((\mathfrak{f}(v))-\mathfrak{f}(u))\left(\rho((\mathfrak{f}(v)))\frac{-(p+V)\sigma^2}{h(z-s)^2} + 2\phi((\mathfrak{f}(v))) - \frac{2(\gamma-hs)}{h(z-s)}\right)\\ &\quad + (z-s)^3\frac{h}{\sigma^2}(\eta((\mathfrak{f}(v))) - \eta(\mathfrak{f}(u)))\left(P((\mathfrak{f}(v)),\mathfrak{f}(u))\frac{-(p+V)\sigma^2}{h(z-s)^2} - \Phi((\mathfrak{f}(v)),\mathfrak{f}(u)) - \frac{2(\gamma-hs)}{h(z-s)}\right)\\ &= (z-s)^3\frac{h}{\sigma^2}\left(\frac{b((\mathfrak{f}(v)))}{(\mathfrak{f}(v))^2}((\mathfrak{f}(v))-\mathfrak{f}(u))(\zeta((\mathfrak{f}(v)),n)-m) + (\eta((\mathfrak{f}(v)))-\eta(\mathfrak{f}(u)))(Z((\mathfrak{f}(v)),\mathfrak{f}(u),n)-m)\right) < 0\end{aligned}$$

whenever

$$Z(\mathfrak{f}(v),\mathfrak{f}(u),n) - m > 0, \tag{A.16}$$

where $n = -(p+V)\sigma^2/h(z-s)^2 > 0$, $m = 2(\gamma-hs)/h(z-s)$,

$$\zeta(y,n) = \rho(y)n + 2\phi(y) \text{ and}$$

$$Z(y,w,n) = P(y,w)n - \Phi(y,w).$$

In Lemma A.16, we prove that, for each $n>0$ and $w \neq 0$, $y=w$ is a root of $Z(\cdot,w,n)-\zeta(\cdot,n)$ and $Z(\cdot,w,n)-\zeta(\cdot,n)$ changes sign at most once on (w,∞). In particular, we show that there is a possibly infinite value $y^* \geq w$ such that $Z(y,w,n) > \zeta(y,n)$ for $w<y<y^*$ and $Z(y,w,n) < \zeta(y,n)$ for $y>y^*$.

We first consider the case of $\mathfrak{f}(u) < \mathfrak{f}(v) < y^*$. Again, if $\partial_v D(v,u,s,z) \leq 0$, (A.13) is trivially satisfied when $D(v,u,s,z) > 0$. Thus, we focus on the case in which $\partial_v D(v,u,s,z) > 0$, that is, $\zeta(\mathfrak{f}(v),n) > m$.

When $\mathfrak{f}(u) < \mathfrak{f}(v) \leq y^*$, $Z(\mathfrak{f}(v),\mathfrak{f}(u),n) \geq \zeta(\mathfrak{f}(v),n)$, and so

$$\begin{aligned}\partial_v D(v,u,s,z)(v-u) - D(v,u,s,z) &= (z-s)^3\frac{h}{\sigma^2}\left(\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}(\mathfrak{f}(v)-\mathfrak{f}(u))(\zeta(\mathfrak{f}(v),n)-m) + (\eta(\mathfrak{f}(v))-\eta(\mathfrak{f}(u)))(Z(\mathfrak{f}(v),\mathfrak{f}(u),n)-m)\right)\\ &\leq (z-s)^3\frac{h}{\sigma^2}\left(\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}(\mathfrak{f}(v)-\mathfrak{f}(u))(\zeta(\mathfrak{f}(v),n)-m) + (\eta(\mathfrak{f}(v))-\eta(\mathfrak{f}(u)))(\zeta(\mathfrak{f}(v),n)-m)\right)\\ &= (z-s)^3\frac{h}{\sigma^2}\left(\frac{b(\mathfrak{f}(v))}{\mathfrak{f}(v)^2}(\mathfrak{f}(v)-\mathfrak{f}(u)) + \eta(\mathfrak{f}(v)) - \eta(\mathfrak{f}(u))\right)(\zeta(\mathfrak{f}(v),n)-m).\end{aligned}$$

By Lemma A.15, $\varphi(y,w) = b(y)/y^2(y-w) + \eta(y) - \eta(w) < 0$ for $y>w$, and so, when $\zeta(\mathfrak{f}(v),n) > m$, $\partial_v D(v,u,s,z)(v-u) - D(v,u,s,z) < 0$.

We next consider $\mathfrak{f}(v) > y^*$ and argue that $Z(y,w,n) < m$ for $y > y^*$. First, observe that, by Lemma A.19, y^* is finite if and only if $n < T(w)$, where $T(w) \equiv \eta(w) + \chi(w)/\alpha(w) - \eta(w)$. Next, observe that, for $y>y^*$, $Z(y,w,n) < \zeta(y,n)$, and so, by Lemma A.16, $\partial_y Z(y,w,n) > 0$. Thus, to show that $Z(y,w,n) < m$ for $y > y^*$, it is enough to show that $\lim_{y\to\infty} Z(y,w,n) \leq m$ when $n < T(w)$. In Lemma A.20, we prove that, when $h>0$ and $n<T(w)$,

$$\lim_{y\to\infty} Z(y,w,n) \leq \frac{wn+2}{e^w-1} + 2 - \frac{2}{w}.$$

Thus, it remains to show that

$$m = \frac{2(\gamma-hs)}{h(z-s)} \geq \frac{wn+2}{e^w-1} + 2 - \frac{2}{w},$$

that is, that

$$\gamma \geq \frac{h(z-s)}{2}\left(\frac{wn+2}{e^w-1} + 2 + 2\frac{s}{z-s} - \frac{2}{w}\right).$$

Now, if $\omega \geq -p$, then

$$\gamma = \gamma(u,s,z,-V,\omega) \geq \gamma(u,s,z,-V,-p) = \frac{h(z-s)}{2}\left(\frac{\mathfrak{f}(u)n+2}{e^{\mathfrak{f}(u)}-1} + 2 + 2\frac{s}{z-s} - \frac{2}{\mathfrak{f}(u)}\right),$$

and so $Z(\mathfrak{f}(v),\mathfrak{f}(u),n) < m$ for $\mathfrak{f}(v) > y^*$.

If $\omega < -p$ and $g(x,u,\gamma) \geq -p$ for some point $x > z$, we argue that $n > T(\mathfrak{f}(u))$, and so y^* is infinite. First, observe that $n < T(\mathfrak{f}(u))$ is equivalent to

$$n = -\frac{(p+V)\sigma^2}{h(z-s)^2} < T(\mathfrak{f}(u)) = -\left(\frac{\sigma^2}{u(z-s)} + \frac{e^{\frac{2u}{\sigma^2}(z-s)}-1}{1-e^{\frac{2u}{\sigma^2}(z-s)}+\frac{2u}{\sigma^2}(z-s)}\right),$$

and because $h > 0$, $n < T(\mathfrak{f}(u))$ if and only if

$$V > -p + h\frac{z-s}{u} + h\frac{(z-s)^2}{\sigma^2}\frac{e^{\frac{2u}{\sigma^2}(z-s)}-1}{1-e^{\frac{2u}{\sigma^2}(z-s)}+\frac{2u}{\sigma^2}(z-s)} = \underline{V}. \tag{A.17}$$

Next, observe that, because $g(s,u,\gamma) = V < -p$,

$$C_u = e^{\frac{2u}{\sigma^2}s}\left(\frac{hs}{u} + (p+V) - \frac{h\sigma^2}{2u^2} - \frac{\gamma}{u}\right),$$

and because

$$\begin{aligned} g(x,u,\gamma) &= -p - \frac{hx}{u} + \frac{h\sigma^2}{2u^2} + \frac{\gamma}{u} + e^{-\frac{2u}{\sigma^2}x}C_u \\ &= -p - \frac{hx}{u} + \frac{h\sigma^2}{2u^2} + \frac{\gamma}{u} + e^{-\frac{2u}{\sigma^2}(x-s)}\left(\frac{hs}{u} + (p+V) - \frac{h\sigma^2}{2u^2} - \frac{\gamma}{u}\right) > -p, \\ \gamma &> \frac{hx - \frac{h\sigma^2}{2u} - e^{-\frac{2u}{\sigma^2}(x-s)}\left(hs + u(p+V) - \frac{h\sigma^2}{2u}\right)}{1-e^{-\frac{2u}{\sigma^2}(x-s)}} \equiv \tilde{\gamma}(V,x). \end{aligned} \tag{A.18}$$

Further, because

$$\omega = g(z,u,\gamma) = -p - \frac{hz}{u} + \frac{h\sigma^2}{2u^2} + \frac{\gamma}{u} + e^{-\frac{2u}{\sigma^2}(z-s)}\left(\frac{hs}{u} + (p+V) - \frac{h\sigma^2}{2u^2} - \frac{\gamma}{u}\right) < -p,$$

by assumption,

$$\gamma < \frac{hz - \frac{h\sigma^2}{2u} - e^{-\frac{2u}{\sigma^2}(z-s)}\left(hs + u(p+V) - \frac{h\sigma^2}{2u}\right)}{1-e^{-\frac{2u}{\sigma^2}(z-s)}} = \tilde{\gamma}(V,z). \tag{A.19}$$

We argue that, when V satisfies (A.17), the lower bound (A.18) exceeds the upper bound (A.19), proving that, when $n < T(\mathfrak{f}(u))$, $\omega \geq -p$. To see this, we first observe that

$$\partial_x\tilde{\gamma}(\underline{V},x) = \frac{e^{\frac{2u}{\sigma^2}(x-s)}hq_5(u,s,z,x)}{\left(e^{\frac{2u}{\sigma^2}(x-s)}-1\right)^2},$$

where

$$q_5(u,s,z,x) = -1 + e^{\frac{2u}{\sigma^2}(x-s)} - \frac{2u(x-z)}{\sigma^2} + \frac{2u^2}{\sigma^4}(z-s)^2\frac{e^{\frac{2u}{\sigma^2}(z-s)}-1}{1-e^{\frac{2u}{\sigma^2}(z-s)}+\frac{2u}{\sigma^2}(z-s)}.$$

Because $q_5(u,s,z,x) > 0$ for all $u \neq 0$ and $x > z > s$, by Lemma A.22, we see that $\tilde{\gamma}(\underline{V},x) > \tilde{\gamma}(\underline{V},z)$ for all $x > z$. Finally, observe that

$$\partial_V\tilde{\gamma}(V,x) - \partial_V\tilde{\gamma}(V,z) = \frac{\left(e^{\frac{2u}{\sigma^2}(x-z)}-1\right)u}{\left(e^{\frac{2u}{\sigma^2}(x-s)}-1\right)\left(1-e^{-\frac{2u}{\sigma^2}(z-s)}\right)} > 0,$$

for all $u \neq 0$ and $s < z < x$, leading to the contradiction that $\gamma > \tilde{\gamma}(V,x) > \tilde{\gamma}(V,z) > \gamma$, from which we conclude that $V < \underline{V}$ and y^* is infinite.

We next consider the function $D(v,u,z,S)/(v-u)$, where $z < S$, $u < v$, and $V = g(S,u,\gamma) = g(S,v,\gamma) > -p$ and $g(z,v,\gamma) = \omega$. To show that $D(v,u,z,S)/(v-u)$ is increasing in u when $D(v,u,z,S) > 0$, we argue that, in this case,

$$(v-u)^2\partial_u\left(\frac{D(v,u,z,S)}{v-u}\right) = \partial_u D(v,u,z,S)(v-u) + D(v,u,z,S) \geq 0 \tag{A.20}$$

whenever $D(v,u,z,S) > 0$. Because $V = g(S,u,\gamma) = g(S,v,\gamma)$,

$$C_\mu = e^{\frac{2\mu}{\sigma^2}S}\left(p + V - \frac{\gamma}{\mu} + \frac{hS}{\mu} - \frac{h\sigma^2}{2\mu^2}\right)$$

for $\mu \in \{u,v\}$,

$$D(v,u,z,S) = (S-z)^3(\eta(\mathfrak{f}(u)) - \eta(\mathfrak{f}(v)))\left(P(\mathfrak{f}(u),\mathfrak{f}(v))\frac{-(p+V)}{(S-z)^2} - \Phi(\mathfrak{f}(u),\mathfrak{f}(v))\frac{h}{\sigma^2} - \frac{2(hS-\gamma)}{\sigma^2(S-z)}\right),$$

and

$$\partial_u D(v,u,z,S)(v-u) = (S-z)^3\frac{b(\mathfrak{f}(u))}{\mathfrak{f}(u)^2}(\mathfrak{f}(u) - \mathfrak{f}(v))\left(\rho(\mathfrak{f}(u))\frac{-(p+V)}{(S-z)^2} + \phi(\mathfrak{f}(u))\frac{2h}{\sigma^2} - \frac{2(hS-\gamma)}{\sigma^2(S-z)}\right),$$

where $\mathfrak{f}(\mu) = -2\mu/\sigma^2(S-z)$. Note that we use the modified scaling $\mathfrak{f}(\cdot)$ as a terse way to maintain the relationship with the drift rates u and v, but in this case, we generally use y for $\mathfrak{f}(u)$ and w for $\mathfrak{f}(v)$.

When $\partial_u D(v,u,z,S) \geq 0$, $\partial_u D(v,u,z,S)(v-u) + D(v,u,z,S)$ is trivially nonnegative whenever $D(v,u,z,S) \geq 0$. Thus, we focus on the case in which $D(v,u,z,S) \geq 0$ and $\partial_u D(v,u,z,S) < 0$, that is,

$$P(\mathfrak{f}(u),\mathfrak{f}(v))\frac{-(p+V)}{(S-z)^2} - \Phi(\mathfrak{f}(u),\mathfrak{f}(v))\frac{h}{\sigma^2} - \frac{2(hS-\gamma)}{\sigma^2(S-z)} \leq 0 \text{ and}$$
$$\rho(\mathfrak{f}(u))\frac{-(p+V)}{(S-z)^2} + \phi(\mathfrak{f}(u))\frac{2h}{\sigma^2} - \frac{2(hS-\gamma)}{\sigma^2(S-z)} < 0.$$

We consider the two cases, $h \geq 0$ and $h < 0$.

Case $h \geq 0$: By Lemma A.12, $P(y,w) > \rho(y)$ for all $y > w$, and by Lemma A.17, $\Phi(y,w) > -2\phi(y)$ for all $y > w$. Because $-p + V/(S-z)^2 < 0$ and $h \geq 0$,

$$P(y,w)\frac{-(p+V)}{(S-z)^2} - \Phi(y,w)\frac{h}{\sigma^2} < \rho(y)\frac{-(p+V)}{(S-z)^2} + \phi(v)\frac{2h}{\sigma^2}$$

for all $y > w$. Thus, because $\varphi(y,w) < 0$ for $y > w$,

$$\begin{aligned}\partial_u D(v,u,z,S)(v-u) + D(v,u,z,S) &= (S-z)^3\frac{b(\mathfrak{f}(u))}{\mathfrak{f}(u)^2}(\mathfrak{f}(u)-\mathfrak{f}(v))\left(\rho(\mathfrak{f}(u))\frac{-(p+V)}{(S-z)^2} + \phi(\mathfrak{f}(u))\frac{2h}{\sigma^2} - \frac{2(hS-\gamma)}{\sigma^2(S-z)}\right)\\ &\quad + (S-z)^3(\eta(\mathfrak{f}(u)) - \eta(\mathfrak{f}(v)))\left(P(\mathfrak{f}(u),\mathfrak{f}(v))\frac{-(p+V)}{(S-z)^2} - \Phi(\mathfrak{f}(u),\mathfrak{f}(v))\frac{h}{\sigma^2} - \frac{2(hS-\gamma)}{\sigma^2(S-z)}\right)\\ &> (S-z)^3\varphi(\mathfrak{f}(u),\mathfrak{f}(v))\left(\rho(\mathfrak{f}(u))\frac{-(p+V)}{(S-z)^2} + \phi(\mathfrak{f}(u))\frac{2h}{\sigma^2} - \frac{2(hS-\gamma)}{\sigma^2(S-z)}\right) > 0.\end{aligned}$$

Case $h < 0$: When $h < 0$, we wish to show that, for $v > u$,

$$\begin{aligned}\partial_u D(v,u,z,S)(v-u) + D(v,u,z,S) &= (S-z)^3\frac{h}{\sigma^2}\frac{b(\mathfrak{f}(u))}{\mathfrak{f}(u)^2}(\mathfrak{f}(u)-\mathfrak{f}(v))\left(\rho(\mathfrak{f}(u))\frac{-(p+V)\sigma^2}{h(S-z)^2} + 2\phi(\mathfrak{f}(u)) - \frac{2(hS-\gamma)}{h(S-z)}\right)\\ &\quad + (S-z)^3\frac{h}{\sigma^2}(\eta(\mathfrak{f}(u)) - \eta(\mathfrak{f}(v)))\left(P(\mathfrak{f}(u),\mathfrak{f}(v))\frac{-(p+V)\sigma^2}{h(S-z)^2} - \Phi(\mathfrak{f}(u),\mathfrak{f}(v)) - \frac{2(hS-\gamma)}{h(S-z)}\right)\\ &= (S-z)^3\frac{h}{\sigma^2}\left(\frac{b(\mathfrak{f}(u))}{\mathfrak{f}(u)^2}(\mathfrak{f}(u)-\mathfrak{f}(v))(\zeta(\mathfrak{f}(u),n) - m) + (\eta(\mathfrak{f}(u)) - \eta(\mathfrak{f}(v)))(Z(\mathfrak{f}(u),\mathfrak{f}(v),n) - m)\right) > 0\end{aligned}$$

whenever

$$Z(\mathfrak{f}(u),\mathfrak{f}(v),n) - m > 0, \tag{A.21}$$

where $n = -(p+V)\sigma^2/h(S-z)^2 > 0$, $m = 2(hS-\gamma)/h(S-z)$,

$$\zeta(y,n) = \rho(y)n + 2\phi(y) \text{ and}$$
$$Z(y,w,n) = P(y,w)n - \Phi(y,w).$$

In Lemma A.16, we prove that, for each $n > 0$ and $w \neq 0$, $y = w$ is a root of $Z(\cdot,w,n) - \zeta(\cdot,n)$ and $Z(\cdot,w,n) - \zeta(\cdot,n)$ changes sign at most once on (w,∞). More specifically, Lemma A.16 proves that there is a possibly infinite value $y^* > w$ such that $Z(y,w,n) > \zeta(y,n)$ for $w < y < y^*$ and $Z(y,w,n) < \zeta(y,n)$ for $y > y^*$.

We first consider the case of $\mathfrak{f}(v) < \mathfrak{f}(u) < y^*$. Again, if $\partial_u D(v,u,z,S) \geq 0$, (A.20) is trivially satisfied when $D(v,u,z,S) > 0$. Thus, we focus on the case in which $\partial_u D(v,u,z,S) < 0$, that is, $\zeta(\mathfrak{f}(u),n) > m$. When $\mathfrak{f}(v) < \mathfrak{f}(u) \leq y^*$, $Z(\mathfrak{f}(u),\mathfrak{f}(v),n) \geq \zeta(\mathfrak{f}(u),n)$ and so

$$\begin{aligned}\partial_u D(v,u,z,S)(v-u) + D(v,u,z,S) &= (S-z)^3 \frac{h}{\sigma^2}\left(\frac{b(\mathfrak{f}(u))}{\mathfrak{f}(u)^2}(\mathfrak{f}(u)-\mathfrak{f}(v))(\zeta(\mathfrak{f}(u),n)-m) + (\eta(\mathfrak{f}(u))-\eta(\mathfrak{f}(v)))(Z(\mathfrak{f}(u),\mathfrak{f}(v),n)-m)\right)\\ &\geq (S-z)^3 \frac{h}{\sigma^2}\left(\frac{b(\mathfrak{f}(u))}{\mathfrak{f}(u)^2}(\mathfrak{f}(u)-\mathfrak{f}(v))(Z(\mathfrak{f}(u),\mathfrak{f}(v),n)-m) + (\eta(\mathfrak{f}(u))-\eta(\mathfrak{f}(v)))(Z(\mathfrak{f}(u),\mathfrak{f}(v),n)-m)\right)\\ &= (S-z)^3 \frac{h}{\sigma^2}\left(\frac{b(\mathfrak{f}(u))}{\mathfrak{f}(u)^2}(\mathfrak{f}(u)-\mathfrak{f}(v)) + \eta(\mathfrak{f}(u))-\eta(\mathfrak{f}(v))\right)(Z(\mathfrak{f}(u),\mathfrak{f}(v),n)-m).\end{aligned}$$

Recall that by Lemma A.15, $\varphi(y,w) = b(y)/y^2(y-w) + \eta(y) - \eta(w) < 0$ for $y \geq w$, and so, when $Z(\mathfrak{f}(u),\mathfrak{f}(v),n) > m$, $\partial_u D(v,u,z,S)\ (v-u) + D(v,u,z,S) > 0$.

We next consider $\mathfrak{f}(u) > y^*$ and argue that $Z(\mathfrak{f}(u),\mathfrak{f}(v),n) \leq m$ for $\mathfrak{f}(u) > y^*$. First, observe that by Lemma A.19, y^* is finite if and only if $n < T(w)$. Next, observe that, for $y > y^*$, $Z(y,w,n) < \zeta(y,n)$, and so, by Lemma A.16, $\partial_y Z(y,w,n) > 0$. Thus, to show that $Z(y,w,n) \leq m$ for $y > y^*$, it is enough to show that $\lim_{y\to\infty} Z(y,w,n) \leq m$ when $n < T(w)$. In Lemma A.20, we prove that, when $n < T(w)$,

$$\lim_{y\to\infty} Z(y,w,n) \leq \frac{wn+2}{e^w - 1} + 2 - \frac{2}{w}.$$

Thus, it remains to show that

$$m = \frac{2(hS-\gamma)}{h(S-z)} \geq \frac{\mathfrak{f}(v)n+2}{e^{\mathfrak{f}(v)}-1} + 2 - \frac{2}{\mathfrak{f}(v)},$$

that is, that

$$\gamma \geq \frac{h(S-z)}{2}\left(-\frac{\mathfrak{f}(v)n+2}{e^{\mathfrak{f}(v)-1}} + 2\frac{z}{(S-z)} + \frac{2}{\mathfrak{f}(v)}\right).$$

Now, if $\omega \leq -p$, then

$$\gamma = \gamma(v,z,S,-\omega,V) \geq \gamma(v,z,S,p,V) = \frac{h(S-z)}{2}\left(-\frac{\mathfrak{f}(v)n+2}{e^{\mathfrak{f}(v)}-1} + 2\frac{z}{(S-z)} + \frac{2}{\mathfrak{f}(v)}\right),$$

and so $Z(\mathfrak{f}(u),\mathfrak{f}(v),n) < m$ for $\mathfrak{f}(u) > y^*$.

If $\omega > -p$ and $g(x,u,\gamma) \leq -p$ for some point $x < z$, we argue that $n > T(\mathfrak{f}(v))$, and so y^* is infinite. First, observe that $n < T(\mathfrak{f}(v))$ is equivalent to

$$n = -\frac{(p+V)\sigma^2}{h(S-z)^2} < T(\mathfrak{f}(v)) = \frac{\sigma^2}{(S-z)v} + \frac{1-e^{\frac{2v}{\sigma^2}(S-z)}}{1-e^{\frac{2v}{\sigma^2}(S-z)} + e^{\frac{2v}{\sigma^2}(S-z)}\frac{2v}{\sigma^2}(S-z)},$$

and because $h < 0$, $n < T(\mathfrak{f}(v))$ if and only if

$$V < -p - h\frac{(S-z)^2}{(S-z)v} - h\frac{(S-z)^2}{\sigma^2}\frac{1-e^{\frac{2v}{\sigma^2}(S-z)}}{1-e^{\frac{2v}{\sigma^2}(S-z)} + e^{\frac{2v}{\sigma^2}(S-z)}\frac{2v}{\sigma^2}(S-z)} = \overline{V}. \tag{A.22}$$

Next, observe that, because $g(S,v,\gamma) = V$,

$$C_v = e^{\frac{2v}{\sigma^2}S}\left(\frac{hS}{v} + p + V - \frac{h\sigma^2}{2v^2} - \frac{\gamma}{v}\right)$$

and because

$$\begin{aligned}g(x,v,\gamma) &= -p - \frac{hx}{v} + \frac{h\sigma^2}{2v^2} + \frac{\gamma}{v} + e^{-\frac{2v}{\sigma^2}x}C_v\\ &= -p - \frac{hx}{v} + \frac{h\sigma^2}{2v^2} + \frac{\gamma}{v} + e^{\frac{2v}{\sigma^2}(S-x)}\left(\frac{hS}{v} + p + V - \frac{h\sigma^2}{2v^2} - \frac{\gamma}{v}\right) < -p,\\ \gamma &> \frac{hx - \frac{h\sigma^2}{2v} - e^{\frac{2v}{\sigma^2}(S-x)}\left(hS + (p+V)v - \frac{h\sigma^2}{2v}\right)}{1-e^{\frac{2v}{\sigma^2}(S-x)}} = \tilde{\gamma}(V,x).\end{aligned} \tag{A.24}$$

Further, because

$$g(z,v,\gamma)=-p-\frac{hz}{v}+\frac{h\sigma^2}{2v^2}+\frac{\gamma}{v}+e^{\frac{2v}{\sigma^2}(S-z)}\left(\frac{hS}{v}+p+V-\frac{h\sigma^2}{2v^2}-\frac{\gamma}{v}\right)=\omega>-p,$$

$$\gamma<\frac{hz-\frac{h\sigma^2}{2v}-e^{\frac{2v}{\sigma^2}(S-z)}\left(hS+(p+V)v-\frac{h\sigma^2}{2v}\right)}{1-e^{\frac{2v}{\sigma^2}(S-z)}}=\tilde{\gamma}(V,z). \tag{A.24}$$

We argue that, when V satisfies (A.22), the lower bound (A.23) exceeds the upper bound (A.24), proving that, when $n<T(\mathfrak{f}(v))$, $\omega<-p$. To see this, we first observe that

$$\partial_x\tilde{\gamma}(\overline{V},x)=\frac{h}{\left(e^{\frac{2v}{\sigma^2}(S-x)}-1\right)^2}q_6(v,x,z,S),$$

where

$$q_6(v,x,z,S)=1+e^{\frac{2v}{\sigma^2}(S-x)}\frac{1-\frac{2v}{\sigma^2}(z-x)+\frac{2v^2}{\sigma^4}(S-z)^2-e^{\frac{2v}{\sigma^2}(S-z)}\left(1-\frac{2v}{\sigma^2}(S-x)+\frac{2v^2}{\sigma^4}\left((S-x)^2-(z-x)^2\right)\right)}{-1+e^{\frac{2v}{\sigma^2}(S-z)}-e^{\frac{2v}{\sigma^2}(S-z)}\frac{2v}{\sigma^2}(S-z)}.$$

Lemma A.21 proves that $q_6(v,x,z,S)>0$ for all $v\neq 0$ and $0\le x\le z<S$, and because $h<0$, $\partial_x\tilde{\gamma}(\overline{V},x)<0$, proving that $\tilde{\gamma}(\overline{V},x)>\tilde{\gamma}(\overline{V},z)$ for $0\le x<z<S$. Finally, observe that

$$\partial_V\tilde{\gamma}(V,x)-\partial_V\tilde{\gamma}(V,z)=e^{\frac{2v}{\sigma^2}(S-z)}\frac{\left(1-e^{\frac{2v}{\sigma^2}(z-x)}\right)v}{\left(e^{\frac{2v}{\sigma^2}(S-x)}-1\right)\left(e^{\frac{2v}{\sigma^2}(S-z)}-1\right)}<0$$

for $0\le x<z<S$, proving that $\tilde{\gamma}(V)>\tilde{\gamma}(V,z)$ for $V<\overline{V}$ and $0\le x<z<S$.

Lemma A.7. *The function*

$$b(y)=(e^{-y}(y+2)+y-2)/y$$

is strictly positive for all $y\neq 0$ and satisfies $\lim_{y\to-\infty}b(y)=\infty$ and $\lim_{y\to\infty}b(y)=1$. Furthermore, $b'(y)$ has the same sign as y and $a(y)>b(y)$ for all $y\neq 0$, where

$$a(y)=1-e^{-y}-ye^{-y}.$$

Proof. First observe that $a(0)=0$ and $a'(y)=ye^{-y}$ has the same sign as y, proving that $a(y)>0$ for $y\neq 0$. Next observe that $A(y)=yb(y)$ satisfies $A(0)=0$ and

$$A'(y)=b(y)+yb'(y)=a(y), \tag{A.25}$$

which is positive for $y\neq 0$. Thus, $A(y)$ has the same sign as y for $y\neq 0$, proving that $b(\cdot)$ is positive for $y\neq 0$.

Further,

$$\lim_{y\to-\infty}b(y)=1+\lim_{y\to-\infty}\frac{e^{-y}(y+2)}{y}=1+\lim_{y\to-\infty}-e^{-y}(y+1)=\infty,$$

and it is easy to see that $\lim_{y\to\infty}b(y)=1$.

Finally, note that, by (A.25),

$$b'(y)=\frac{a(y)-b(y)}{y} \tag{A.26}$$

for $y\neq 0$. Now recall that

$$a(y)-b(y)=1-e^{-y}-ye^{-y}-\frac{1}{y}(y-2+e^{-y}(y+2))=\frac{2}{y}-e^{-y}\left(2+y+\frac{2}{y}\right),$$

and so

$$e^y b'(y)/y=e^y(a(y)-b(y))/y^2=\left(2e^y-y^2-2y-2\right)/y^3=-\chi(-y). \tag{A.27}$$

By Lemma A.11, $\chi(y)<0$ for all $y\neq 0$, and so $b'(y)$ has the same sign as y, which, together with (A.26), proves $a(y)>b(y)$ for $y\neq 0$.

Lemma A.8. *The function*

$$c(y)=-\frac{e^{-y}(y+3)-\frac{y^2}{2}+2y-3}{y^2}$$

is positive for $y\neq 0$. Further, $\lim_{y\to-\infty}c(y)=\infty$ and $\lim_{y\to\infty}c(y)=1/2$.

Proof. First, observe that $y^2c(y)$ has the same sign as $c(y)$, is zero when $y = 0$, and the derivative of $y^2c(y)$ is $2yc(y) + y^2c'(y) = yb(y)$, which, by Lemma A.7, has the same sign as y. Thus, $c(y) > 0$ for $y \neq 0$.

Because

$$\lim_{y\to-\infty} y^2c(y) = \lim_{y\to-\infty} -\left(e^{-y}(y+3) - \frac{y^2}{2} + 2y - 3\right) = 3 + \lim_{y\to-\infty} -e^{-y}(y+3) + \frac{y^2}{2} - 2y = \infty$$

and

$$\lim_{y\to\infty} y^2c(y) = 3 + \lim_{y\to-\infty} -(3+y)e^{-y} + \frac{y(y-4)}{2} = \infty,$$

$$\lim_{y\to-\infty} c(y) = \lim_{y\to-\infty} \frac{y^2c(y)}{y^2} = \lim_{y\to-\infty} \frac{yb(y)}{2y} = \lim_{y\to-\infty} \frac{b(y)}{2} = \infty$$

and

$$\lim_{y\to\infty} c(y) = \lim_{y\to\infty} \frac{b(y)}{2} = \frac{1}{2}.$$

Lemma A.9. *The function*

$$\alpha(y) = (1 - e^{-y})/y$$

is strictly positive and decreasing for all $y \neq 0$. Furthermore, $\lim_{y\to-\infty}\alpha(y) = \infty$, $\lim_{y\to\infty}\alpha(y) = 0$, and $\alpha'(y) = -a(y)/y^2$.

Proof. Clearly, $\alpha(y) > 0$ for $y \neq 0$.

$$\alpha'(y) = \frac{e^{-y}}{y} - \frac{1-e^{-y}}{y^2} = -\frac{1 - e^{-y} - ye^{-y}}{y^2} = -\frac{a(y)}{y^2}.$$

Further,

$$\lim_{y\to-\infty} \alpha(y) = \lim_{y\to-\infty} \frac{1-e^{-y}}{y} = \lim_{y\to-\infty} e^{-y} = \infty.$$

And

$$\lim_{y\to\infty} \alpha(y) = \lim_{y\to\infty} \frac{1-e^{-y}}{y} = 0.$$

Lemma A.10. *The function*

$$\eta(y) = (e^{-y} + y - 1)/y^2$$

is strictly positive and decreasing for all $y \neq 0$. Furthermore, $\lim_{y\to-\infty}\eta(y) = \infty$, $\lim_{y\to\infty}\eta(y) = 0$, $\alpha(y) > \eta(y)$, and $\eta'(y) = -b(y)/y^2$.

Proof. Observe that

$$\eta'(y) = \frac{1-e^{-y}}{y^2} - \frac{2(e^{-y}+y-1)}{y^3} = -\frac{y(e^{-y}-1) + 2(e^{-y}-1) + 2y}{y^3} = -\frac{y - 2 + e^{-y}(y+2)}{y^3} = -\frac{b(y)}{y^2},$$

and so $\eta(\cdot)$ is decreasing by Lemma A.7. Note that

$$\alpha(y) - \eta(y) = \frac{a(y)}{y^2} > 0.$$

Because

$$\lim_{y\to\infty} \eta(y) = \lim_{y\to\infty} \frac{e^{-y}+y-1}{y^2} = \lim_{y\to\infty} \frac{1-e^{-y}}{2y} = 0,$$

$\eta(\cdot)$ is positive. Finally,

$$\lim_{y\to-\infty} \eta(y) = \lim_{y\to-\infty} \frac{y-1}{y^2} + \lim_{y\to-\infty} \frac{e^{-y}}{y^2} = \lim_{y\to-\infty} \frac{e^{-y}}{y^2} = \lim_{y\to-\infty} \frac{-e^{-y}}{2y} = \lim_{y\to-\infty} \frac{e^{-y}}{2} = \infty.$$

Lemma A.11. *The function*

$$\chi(y) = \left(2e^{-y} - y^2 + 2y - 2\right)/y^3$$

is strictly negative and increasing for all $y \neq 0$. Furthermore, $\lim_{y\to-\infty}\chi(y) = -\infty$, $\lim_{y\to\infty}\chi(y) = 0$, and $\chi'(y) = 2c(y)/y^2$.

Proof. Observe that

$$\chi'(y) = \frac{1}{y^3}(-2e^{-y} - 2y + 2) - \frac{3}{y^4}\left(2e^{-y} - y^2 + 2y - 2\right) = -\frac{2}{y^4}\left(e^{-y}(y+3) - \frac{y^2}{2} + 2y - 3\right) = \frac{2c(y)}{y^2} > 0,$$

and so $\chi(\cdot)$ is increasing. Further,

$$\lim_{y\to\infty}\chi(y) = \lim_{y\to\infty}\frac{2e^{-y} - y^2 + 2y - 2}{y^3} = \lim_{y\to\infty}\frac{2e^{-y} - (y-2)y - 2}{y^3} = \lim_{y\to\infty}\frac{2}{3y^2}(-e^{-y} - y + 1)$$
$$= \lim_{y\to\infty}\frac{2}{6y}(e^{-y} - 1) = 0,$$

and so $\chi(\cdot)$ is negative. Finally,

$$\lim_{y\to-\infty}\chi(y) = \lim_{y\to-\infty}\frac{2e^{-y} - (y-2)y - 2}{y^3} = \lim_{y\to-\infty}\frac{2}{3y^2}(-e^{-y} - y + 1) = \lim_{y\to-\infty}\frac{2}{6y}(e^{-y} - 1) = -\infty.$$

Lemma A.12. *For $y > w$, the function*

$$P(y,w) = (\alpha(y) - \alpha(w))/(\eta(y) - \eta(w))$$

is nonnegative and decreasing in y. Furthermore, $\lim_{y\downarrow w}P(y,w) = \rho(w)$, $\lim_{y\to\infty}P(y,w) = \alpha(w)/\eta(w)$, $P(y,w) > \rho(y)$, and $\lim_{y\downarrow w}\partial_y P(y,w) = 1/2\rho'(w)$, where

$$\begin{aligned}
&\rho(y) = a(y)/b(y),\\
&\alpha(y) = (1 - e^{-y})/y,\\
&\eta(y) = (e^{-y} + y - 1)/y^2,\\
&a(y) = 1 - e^{-y} - ye^{-y} \text{ and}\\
&b(y) = (y - 2 + e^{-y}(y+2))/y.
\end{aligned}$$

Proof. Note that, because $\eta'(y) = -b(y)/y^2 < 0$ and $\alpha'(y) = -a(y)/y^2$ for $y \neq 0$,

$$\begin{aligned}
\partial_y P(y,w) &= \frac{1}{(\eta(y) - \eta(w))^2}\left(\alpha'(y)(\eta(y) - \eta(w)) - (\alpha(y) - \alpha(w))\eta'(y)\right)\\
&= \frac{\eta'(y)}{(\eta(y) - \eta(w))^2}\left(\frac{\alpha'(y)}{\eta'(y)}(\eta(y) - \eta(w)) - (\alpha(y) - \alpha(w))\right)\\
&= \frac{\eta'(y)}{(\eta(y) - \eta(w))^2}\left(\frac{a(y)}{b(y)}(\eta(y) - \eta(w)) - (\alpha(y) - \alpha(w))\right)\\
&= -\frac{b(y)}{y^2(\eta(y) - \eta(w))^2}(\rho(y)(\eta(y) - \eta(w)) - (\alpha(y) - \alpha(w))).
\end{aligned}$$

Because $b(y) > 0$ for $y \neq 0$, $\partial_y P(y,w)$ has the same sign as

$$f_3(y,w) = -(\rho(y)(\eta(y) - \eta(w)) - (\alpha(y) - \alpha(w))).$$

Clearly, $\lim_{y\downarrow w} f_3(y,w) = 0$ and

$$\begin{aligned}
\partial_y f_3(y,w) &= -\left(\rho'(y)(\eta(y) - \eta(w)) - \alpha'(y) + \rho(y)\eta'(y)\right)\\
&= -\left(\rho'(y)(\eta(y) - \eta(w)) + \frac{a(y)}{y^2} - \frac{a(y)}{b(y)}\frac{b(y)}{y^2}\right) = -\rho'(y)(\eta(y) - \eta(w)).
\end{aligned}$$

Now, by Lemma A.13, $\rho'(y) < 0$ for $y \neq 0$, and by Lemma A.10, $\eta(y) - \eta(w) < 0$ for $y > w$. Thus, we see that $\partial_y P(y,w) < 0$, and so $P(y,w)$ is decreasing in y for $y > w$.

Because $\rho(y)(\eta(y) - \eta(w)) - (\alpha(y) - \alpha(w)) > 0$ and, by Lemma A.10, $\eta(y) - \eta(w) < 0$ for $y > w$, it follows that

$$0 < \rho(y) < \frac{\alpha(y) - \alpha(w)}{\eta(y) - \eta(w)} = P(y,w) \text{ for } y > w.$$

Lemma A.13. *The function*

$$\rho(y) = a(y)/b(y)$$

is positive and decreasing for $y \neq 0$. *Furthermore,* $\lim_{y\to-\infty}\rho(y) = \infty$ *and* $\lim_{y\to\infty}\rho(y) = 1$, *where*

$$a(y) = 1 - e^{-y} - ye^{-y} \text{ and}$$
$$b(y) = \left(y - 2 + e^{-y}(y+2)\right)/y.$$

Proof. From Lemma A.7, it is clear that $\rho(y) > 0$ for $y \neq 0$ and $\lim_{y\to\infty}\rho(y) = 1$. To see that $\lim_{y\to-\infty}\rho(y) = \infty$, recall that, by (A.26), $b'(y) = a(y) - b(y)/y$. Further,

$$\lim_{y\to-\infty}\rho(y) = \lim_{y\to-\infty}\frac{a(y)}{b(y)} = \lim_{y\to-\infty}\frac{ya(y)}{A(y)},$$

where $A(y) = yb(y)$. Because $\lim_{y\to-\infty} ya(y) = \lim_{y\to-\infty} y(1 - e^{-y} - ye^{-y}) = -\infty$, $\lim_{y\to-\infty}A(y) = -\infty$, and $A'(y) = a(y)$ by (A.25),

$$\lim_{y\to-\infty}\rho(y) = \lim_{y\to-\infty}\frac{ya(y)}{A(y)} = \lim_{y\to\infty}\frac{ya'(y) + a(y)}{a(y)} = 1 + \lim_{y\to-\infty}\frac{ya'(y)}{a(y)}.$$

Similarly, because $\lim_{y\to-\infty} ya'(y) = \lim_{y\to-\infty} y^2e^{-y} = \infty$ and $\lim_{y\to-\infty}a(y) = \infty$,

$$\lim_{y\to-\infty}\rho(y) = 1 + \lim_{y\to-\infty}\frac{ya'(y)}{a(y)} = 1 + \lim_{y\to-\infty}\frac{ya''(y) + a'(y)}{a'(y)} = 2 + \lim_{y\to-\infty}\frac{y(e^{-y} - ye^{-y})}{ye^{-y}} = 2 + \lim_{y\to-\infty} 1 - y = \infty.$$

To see that ρ is decreasing, observe that

$$\rho'(y) = \frac{a'(y)}{b(y)} - \frac{a(y)b'(y)}{b(y)^2} = \frac{a'(y)b(y) - a(y)b'(y)}{b(y)^2} = \frac{ya'(y)A(y) - y^2a(y)b'(y)}{A(y)^2}.$$

We argue that

$$f_4(y) = ya'(y)A(y) - ya(y)(a(y) - b(y)) = \left(ya'(y) + a(y)\right)A(y) - ya(y)^2 < 0 \text{ for } y \neq 0.$$

Observe that $f_4(0) = 0$ and

$$f_4'(y) = ye^{-y}f_{4.1}(y),$$

where

$$f_{4.1}(y) = (3 - y)A(y) - ya(y).$$

Observe that $f_{4.1}(0) = 0$ and $f_{4.1}'(y) = -2A(y) = -2yb(y)$, and so $f_{4.1}'(y) > 0$ for $y < 0$ and $f_{4.1}'(y) < 0$ for $y > 0$. Thus, $f_{4.1}(y) < f_{4.1}(0) = 0$. It follows that $f_4'(y) > 0$ for $y < 0$ and $f_4'(y) < 0$ for $y > 0$, and so $f_4(y) < f_4(0) = 0$ for $y \neq 0$. Thus, $\rho'(y) < 0$ for $y \neq 0$, proving that $\rho(\cdot)$ is decreasing.

Lemma A.14. *The function*

$$\phi(y) = \frac{c(y)}{b(y)}$$

is nonnegative and strictly increasing. Further, $\lim_{y\to\infty}\phi(y) = 1/2$.

Proof. By Lemmas A.7 and A.8, $\phi(\cdot)$ is nonnegative. We argue that

$$\phi'(y) = \frac{1}{b(y)^2}\left(c'(y)b(y) - c(y)b'(y)\right) > 0$$

for $y \neq 0$. Note that $\phi'(y)$ has the same sign as

$$f_5(y) = 2y^4e^y\left(c'(y)b(y) - c(y)b'(y)\right).$$

To see that $f_5(y) > 0$ for $y \neq 0$, observe that $f_5(0) = 0$ and

$$\begin{aligned} f_5'(y) &= \left(8y^3e^y + 2y^4e^y\right)\left(c'(y)b(y) - c(y)b'(y)\right) + 2y^4e^y\left(c''(y)b(y) + c'(y)b'(y) - c'(y)b'(y) - c(y)b''(y)\right) \\ &= \left(8y^3e^y + 2y^4e^y\right)\left(c'(y)b(y) - c(y)b'(y)\right) + 2y^4e^y\left(c''(y)b(y) - c(y)b''(y)\right) \\ &= 2e^{-y}yf_{5.1}(y), \end{aligned} \tag{A.28}$$

where

$$f_{5.1}(y) = e^{2y}(y - 4) - (y + 4) + 2e^y\left(y^2 + 4\right).$$

So $f_5'(y)$ has the same sign as $yf_{5.1}(y)$. We argue that $yf_{5.1}(y) > 0$ for all $y \neq 0$.

Note that $f_{5.1}(0) = 0$,

$$f'_{5.1}(y) = 2e^{2y}(y-4) + e^{2y} - 1 + 2e^{y}\left(y^2+4\right) + 4ye^{y} = -1 + e^{2y}(2y-7) + 2e^{y}\left(y^2+2y+4\right)$$

and

$$f''_{5.1}(y) = 2e^{2y}(2y-7) + 2e^{2y} + 2e^{y}\left(y^2+2y+4\right) + 2e^{y}(2y+2) = 2e^{y}\left(2e^{y}(y-3) + y^2 + 4y + 6\right).$$

We argue that $f''_{5.1}(y) \geq 0$ for all y and so $f_{5.1}(\cdot)$ is convex. Clearly, $f''_{5.1}(y)$ has the same sign as

$$f_{5.2}(y) = 2e^{y}(y-3) + y^2 + 4y + 6$$

and $f_{5.2}(0) = 0$ and

$$f'_{5.2}(y) = 2e^{y}(y-3) + 2e^{y} + 2y + 4 = 2(e^{y}(y-2) + (y+2)) = 2ye^{y}b(y)$$

And so has the same sign as y, proving that $f_{5.2}(y) > 0$ for $y \neq 0$, and hence, $f_{5.1}(\cdot)$ is strictly convex. Thus, the fact that $f'_{5.1}(0) = 0$ implies that $0 = f_{5.1}(0) < f_{5.1}(y)$ for all $y \neq 0$. Thus, we see that $f'_5(y)$ has the same sign as y for all $y \neq 0$. Because $f_5(0) = 0$, this implies that $f_5(y) > 0$, and so $\phi'(y) > 0$ for all $y \neq 0$.

Finally, by Lemmas A.7 and A.8, $\lim_{y\to\infty}\phi(y) = \lim_{y\to\infty} c(y)/b(y) = 1/2$.

Lemma A.15. *The function*

$$\varphi(y,w) = \frac{b(y)}{y^2}(y-w) + \eta(y) - \eta(w) < 0 \text{ for all } y > w,$$

where

$$\eta(y) = (e^{-y} + y - 1)/y^2 \text{ and}$$
$$b(y) = (e^{-y}(y+2) + y - 2)/y.$$

Proof. Observe that $\lim_{y\downarrow w}\varphi(y,w) = 0$, and by Lemma A.10,

$$\partial_y\varphi(y,w) = \partial_y\left(\frac{b(y)}{y^2}\right)(y-w) + \frac{b(y)}{y^2} + \eta'(y) = \partial_y\left(\frac{b(y)}{y^2}\right)(y-w).$$

We argue that $\partial_y(b(y)/y^2) < 0$, and hence, $\partial_y\varphi(y,w) < 0$ for all $y > w$.

Recall that $b'(y) = (a(y) - b(y))/y$, and so

$$\partial_y\left(\frac{b(y)}{y^2}\right) = \frac{b'(y)}{y^2} - 2\frac{b(y)}{y^3} = \frac{1}{y^3}\left(yb'(y) - 2b(y)\right) = \frac{1}{y^3}(a(y) - 3b(y)).$$

We argue that

$$f_6(y) = a(y) - 3b(y)$$

Is decreasing and has the same sign as $-y$, and so $\partial_y(b(y)/y^2)$ has the same sign as $-y^{-2} < 0$ for $y \neq 0$.

Observe that, because $\lim_{y\to 0} a(y) = 0$ and $\lim_{y\to 0} b(y) = 0$, $\lim_{y\to 0} f_6(y) = 0$. We argue that

$$f'_6(y) = a'(y) - 3b'(y) = \frac{e^{-y}}{y^2}\left(6 - 6e^{y} + 6y + 3y^2 + y^3\right),$$

which has the same sign as

$$f_{6.1}(y) = 6 - 6e^{y} + 6y + 3y^2 + y^3,$$

and this is strictly negative for $y \neq 0$.

Observe that

$$f'_{6.1}(y) = -6e^{y} + 6 + 6y + 3y^2,$$
$$f''_{6.1}(y) = -6e^{y} + 6 + 6y,$$
$$f'''_{6.1}(y) = 6(1 - e^{y}),$$

$f_{6.1}(0) = f'_{6.1}(0) = f'''_{6.1}(y) = 0$, and $f'''_{6.1}(y)$ has the same sign as $-y$. Thus, we see that $f''_{6.1}(y) < 0$ for all $y \neq 0$, $f'_{6.1}(y)$ has the same sign as $-y$ for all y, and so $f_{6.1}(y) < 0$ and $f'_6(y) < 0$ for $y \neq 0$. Thus, $f_6(y)$ has the same sign as $-y$ and $\partial_y(b(y)/y^2)$ has the same sign as $-y^{-2} < 0$ for $y \neq 0$, proving that $\partial_y\varphi(y,w) < 0$, and hence, $\varphi(y,w) < 0$ for all $y > w$.

Lemma A.16. *For $y > w$,* (i)

$$\partial_y Z(y,w,n) = \frac{b(y)}{y^2(\eta(y)-\eta(w))}(Z(y,w,n)-\zeta(y,n)).$$

Further, (ii) *$\partial_y\zeta(\cdot,n)$ admits exactly one real root in $[w,\infty)$ when $t(w) = -2\phi'(w)/\rho'(w) \le n < 1$ and no real roots in $[w,\infty)$ otherwise.* (iii) *$Z(w,w,n)-\zeta(w,n) = 0$ for all $w \neq 0$ and* (iv) *if $t(w) < n < 1$, $Z(\cdot,w,n)-\zeta(\cdot,n)$ changes sign at most once on (w,∞). Otherwise, $Z(\cdot,w,n)-\zeta(\cdot,n)$ does not change sign on (w,∞). Moreover,* (v) *if $Z(\cdot,w,n)-\zeta(\cdot,n)$ changes sign on (w,∞), there is $y^* > w$ such that $Z(y,w,n)-\zeta(y,n) > 0$ for $w < y < y^*$ and $Z(y,w,n)-\zeta(y,n) < 0$ for $y > y^*$. Finally,* (vi) *when it exists, y^* is increasing in n.*

Proof. To prove (i), observe that, by Lemmas A.9–A.11,

$$\begin{aligned}
\partial_y Z(y,w,n) &= \partial_y P(y,w)n - \partial_y \Phi(y,w) = \partial_y\left(\frac{\alpha(y)-\alpha(w)}{\eta(y)-\eta(w)}\right)n - \partial_y\left(\frac{\chi(y)-\chi(w)}{\eta(y)-\eta(w)}\right)\\
&= \left(\frac{\alpha'(y)}{\eta(y)-\eta(w)} - \frac{\alpha(y)-\alpha(w)}{(\eta(y)-\eta(w))^2}\eta'(y)\right)n - \left(\frac{\chi'(y)}{\eta(y)-\eta(w)} - \frac{\chi(y)-\chi(w)}{(\eta(y)-\eta(w))^2}\eta'(y)\right)\\
&= \frac{\eta'(y)}{\eta(y)-\eta(w)}\left(\left(\frac{\alpha'(y)}{\eta'(y)} - \frac{\alpha(y)-\alpha(w)}{\eta(y)-\eta(w)}\right)n - \left(\frac{\chi'(y)}{\eta'(y)} - \frac{\chi(y)-\chi(w)}{\eta(y)-\eta(w)}\right)\right)\\
&= \frac{\eta'(y)}{\eta(y)-\eta(w)}\left(\frac{\alpha'(y)}{\eta'(y)}n - \frac{\chi'(y)}{\eta'(y)} - \left(\frac{\alpha(y)-\alpha(w)}{\eta(y)-\eta(w)}n - \frac{\chi(y)-\chi(w)}{\eta(y)-\eta(w)}\right)\right)\\
&= \frac{\eta'(y)}{\eta(y)-\eta(w)}\left(\frac{a(y)}{b(y)}n + \frac{2c(y)}{b(y)} - \left(\frac{\alpha(y)-\alpha(w)}{\eta(y)-\eta(w)}n - \frac{\chi(y)-\chi(w)}{\eta(y)-\eta(w)}\right)\right)\\
&= \frac{\eta'(y)}{\eta(y)-\eta(w)}(\rho(y)n + 2\phi(y) - (P(y,w)n - \Phi(y,w)))\\
&= \frac{b(y)}{y^2(\eta(y)-\eta(w))}(Z(y,w,n)-\zeta(y,n)).
\end{aligned} \tag{A.29}$$

To prove (ii), recall that $\zeta(y,n) = \rho(y)n + 2\phi(y)$ and, by Lemma A.13, $\rho'(y) < 0$, and so y^* is a root of $\partial_y\zeta(\cdot,n)$ if and only if $\rho'(y^*)n + 2\phi'(y^*) = 0$, that is, if and only if

$$n = -2\frac{\phi'(y)}{\rho'(y)}.$$

By Lemma A.18, $t(y) = -2\phi'(y)/\rho'(y)$ is increasing and $\lim_{y\to\infty} t(y) = 1$. Thus, $\partial_y\zeta(\cdot,n)$ admits exactly one real root in $[w,\infty)$ if $t(w) \le n < 1$ and no real roots otherwise.

To prove (iii), observe that, by Lemmas A.12 and A.17,

$$\lim_{y\downarrow w} Z(y,w,n) = \rho(w)n + 2\phi(w) = \zeta(w,n),$$

and so $Z(w,w,n)-\zeta(w,n) = 0$. Further, by (A.29),

$$\partial_y Z(y,w,n) - \partial_y\zeta(y,n) = \frac{b(y)}{y^2(\eta(y)-\eta(w))}(Z(y,w,n)-\zeta(y,n)) - \partial_y\zeta(y,n),$$

and so $\partial_y Z(\cdot,w,n) - \partial_y\zeta(\cdot,n) = -\partial_y\zeta(\cdot,n)$ at the roots of $Z(\cdot,w,n)-\zeta(\cdot,n)$.

Next, we argue that, if $y^* \ge w$ is a root of $Z(\cdot,w,n)-\zeta(\cdot,n)$ and a root of $\partial_y Z(\cdot,w,n) - \partial_y\zeta(\cdot,n)$, then $y^* = w$ and $n = t(w)$. Observe that, by arguments analogous to those used in (A.29),

$$\partial_w Z(y,w,n) = \frac{b(w)}{w^2(\eta(y)-\eta(w))}((\rho(w)-P(y,w))n + \Phi(y,w) + 2\phi(w)),$$

where, for $y > w$, $\rho(w) > P(y,w)$ by Lemma A.12, $\Phi(y,w) + 2\phi(w) > 0$ by Lemma A.17, $b(w) > 0$ by Lemma A.7, and $\eta(y) - \eta(w) < 0$ by Lemma A.10. Thus, $\partial_w Z(y,w,n) < 0$ for $y > w$, and so $Z(y,w,t(y)) - \zeta(y,t(y))$ is decreasing in w for $w < y$. But $\lim_{w\uparrow y} Z(y,w,t(y)) = \zeta(y,t(y))$, and so $Z(y,w,t(y)) - \zeta(y,t(y)) > 0$ for $y > w$. Thus, if $y^* \ge w$ is a root of $Z(\cdot,w,n)-\zeta(\cdot,n)$, then $\partial_y Z(y^*,w,n) - \partial_y\zeta(y^*,n) = -\partial_y\zeta(y^*,n)$. If y^* is also a root of $\partial_y Z(\cdot,w,n) - \partial_y\zeta(\cdot,n)$, then $\partial_y\zeta(y^*,n) = 0$, and so $n = t(y^*)$. But we show that $Z(y,w,t(y)) - \zeta(y,t(y)) > 0$ for $y > w$, and so it must be the case that $y^* = w$ and $n = t(w)$.

To prove (iv), suppose, without loss of generality, that $Z(\cdot,w,n)-\zeta(\cdot,n)$ is positive on (w,y_1), negative on (y_1,y_2), and positive on (y_2,y_3) for some $w < y_1 < y_2 < y_3$. Then, because w, y_1, and y_2 are roots of $Z(\cdot,w,n)-\zeta(\cdot,n)$,

- $\partial_y Z(w,w,n) - \partial_y\zeta(w,n) \ge 0$ and so $\partial_y\zeta(w,n) \le 0$.
- $\partial_y Z(y_1,w,n) - \partial_y\zeta(y_1,n) < 0$ and so $\partial_y\zeta(y_1,n) \ge 0$.
- $\partial_y Z(y_2,w,n) - \partial_y\zeta(y_2,n) > 0$ and so $\partial_y\zeta(y_2,n) \le 0$.

But that is only possible if $\partial_y \zeta(\cdot, n)$ has at least two real roots in $[w, \infty)$. Thus, $Z(\cdot, w, n) - \zeta(\cdot, n)$ changes sign at most once on (w, ∞) and can only change sign if $\partial_y \zeta(\cdot, n)$ admits a root in (w, ∞), that is, if $t(w) \le n < 1$.

To prove (v), we argue that, if $Z(\cdot, w, n) - \zeta(\cdot, n)$ changes sign on (w, ∞), then there is $y^* > w$ such that $Z(y, w, n) - \zeta(y, n) > 0$ for $w < y < y^*$ and $Z(y, w, n) - \zeta(y, n) < 0$ for $y > y^*$. In particular, we show that, if $\partial_y Z(w, w, n) - \partial_y \zeta(w, n) \le 0$, then $\partial_y \zeta(\cdot, n)$ has no real root in (w, ∞).

By Lemmas A.12 and A.17,

$$\lim_{y \downarrow w} \partial_y Z(y, w, n) = \lim_{y \downarrow w} \partial_y P(y, w) n - \partial_y \Phi(y, w) = 1/2 \rho'(w) n + \phi'(w)$$

and

$$\lim_{y \downarrow w} \partial_y \zeta(y, n) = \lim_{y \downarrow w} \rho'(y) n + 2\phi'(y) = \rho'(w) n + 2\phi'(w),$$

and by Lemma A.13, $\rho'(\cdot) < 0$, so $\partial_y Z(w, w, n) - \partial_y \zeta(w, n) \le 0$ if and only if $\partial_y \zeta(w, n) = \rho'(w) n + 2\phi'(w) \ge 0$, that is, $n \le -2\phi'(w)/\rho'(w) = t(w)$, which we proved implies that $\partial_y \zeta(\cdot, n)$ has no real root in (w, ∞), and so $Z(\cdot, w, n) - \zeta(\cdot, n)$ cannot change sign on (w, ∞).

Finally, to prove (vi), observe that, expressed as a function of n, when $y^*(n)$ exists,

$$y^{*\prime}(n) = -\frac{P(y^*(n), w) - \rho(y^*(n))}{\partial_y Z(y^*(n), w, n) - \partial_y \zeta(y^*(n), n)},$$

where the numerator is positive by Lemma A.12 and the denominator is negative by (v). Thus, we see that y^* is increasing in n.

Lemma A.17. *The function*

$$\Phi(y, w) = \frac{\chi(y) - \chi(w)}{\eta(y) - \eta(w)}$$

is negative and decreasing in y for $y > w$. Further, $\Phi(y, w) > -2\phi(y)$ for $y > w$, $\lim_{y \downarrow w} \Phi(y, w) = -2\phi(w)$, $\lim_{y \to \infty} \Phi(y, w) = \chi(w)/\eta(w)$, and $\lim_{y \downarrow w} \partial_y \Phi(y, w) = -\phi'(w)$, where

$$\begin{aligned}
\chi(y) &= \left(2e^{-y} + 2y - y^2 - 2\right)/y^3, \\
\phi(y) &= c(y)/b(y), \\
\rho(y) &= a(y)/b(y), \\
a(y) &= 1 - e^{-y} - ye^{-y}, \\
b(y) &= (e^{-y}(y + 2) + y - 2)/y \text{ and} \\
c(y) &= -\left(2y - 3 - y^2/2 + 3e^{-y} + ye^{-y}\right)/y^2.
\end{aligned}$$

Proof. Observe that, by Lemmas A.10 and A.11,

$$\begin{aligned}
\partial_y \Phi(y, w) &= \frac{\chi'(y)}{\eta(y) - \eta(w)} - \frac{\chi(y) - \chi(w)}{(\eta(y) - \eta(w))^2} \eta'(y) \\
&= \frac{1}{(\eta(y) - \eta(w))^2} \left(\chi'(y)(\eta(y) - \eta(w)) - \eta'(y)(\chi(y) - \chi(w))\right) \\
&= \frac{b(y)}{y^2 (\eta(y) - \eta(w))^2} (2\phi(y)(\eta(y) - \eta(w)) + \chi(y) - \chi(w)),
\end{aligned}$$

which, by Lemma A.7, has the same sign as

$$f_7(y, w) = 2\phi(y)(\eta(y) - \eta(w)) + \chi(y) - \chi(w).$$

Now, $\lim_{y \downarrow w} f_7(y, w) = 0$ and

$$\begin{aligned}
\partial_y f_7(y, w) &= 2\phi'(y)(\eta(y) - \eta(w)) + 2\phi(y)\eta'(y) + \chi'(y) \\
&= 2\phi'(y)(\eta(y) - \eta(w)) - 2\frac{c(y)}{b(y)}\frac{b(y)}{y^2} + 2\frac{c(y)}{y^2} = 2\phi'(y)(\eta(y) - \eta(w)).
\end{aligned}$$

By Lemma A.10, $\eta(y) - \eta(w) < 0$, and $\partial_y f_7(y, w)$ has the same sign as $-\phi'(y)$, which is negative for $y \ne 0$. Therefore, $f_7(y, w) < 0$, and hence, $\partial_y \Phi(y, w) < 0$ for $y > w$.

The fact that

$$f_7(y, w) = 2\phi(y)(\eta(y) - \eta(w)) + \chi(y) - \chi(w) < 0$$

for all $y > w$ implies that $-2\phi(y) < \chi(y) - \chi(w)/\eta(y) - \eta(w) = \Phi(y, w)$ for all $y > w$.

Further,

$$\lim_{y \downarrow w} \Phi(y, w) = \lim_{y \downarrow w} \frac{\chi(y) - \chi(w)}{\eta(y) - \eta(w)} = \lim_{y \downarrow w} \frac{\chi'(y)}{\eta'(y)} = -2 \lim_{y \downarrow w} \frac{c(y)}{b(y)} = -2\frac{c(w)}{b(w)} = -2\phi(w).$$

By Lemmas A.10 and A.11,

$$\lim_{y \to \infty} \Phi(y, w) = \lim_{y \to \infty} \frac{\chi(y) - \chi(w)}{\eta(y) - \eta(w)} = \frac{\chi(w)}{\eta(w)}.$$

Finally,

$$\lim_{y \downarrow w} \partial_y \Phi(y, w) = -\phi'(w).$$

Lemma A.18. *The function*

$$t(y) = -2\frac{\phi'(y)}{\rho'(y)}$$

is increasing, and $\lim_{y \to \infty} t(y) = 1$, *where*

$$\begin{aligned}
&\phi(y) = c(y)/b(y),\\
&\rho(y) = a(y)/b(y),\\
&a(y) = 1 - e^{-y} - ye^{-y},\\
&b(y) = (e^{-y}(y+2) + y - 2)/y \text{ and}\\
&c(y) = -\left(2y - 3 - y^2/2 + 3e^{-y} + ye^{-y}\right)/y^2.
\end{aligned}$$

Proof. Observe that

$$t(y) = -2\frac{\phi'(y)}{\rho'(y)} = -2\frac{c'(y)b(y) - b'(y)c(y)}{a'(y)b(y) - a(y)b'(y)} = -2\frac{b(y) - 2c(y) - \phi(y)(a(y) - b(y))}{y^2e^{-y} - \rho(y)(a(y) - b(y))}.$$

Now, $\lim_{y\to\infty} a(y) = 1$, $\lim_{y\to\infty} b(y) = 1$ by Lemma A.7, $\lim_{y\to\infty} c(y) = 1/2$ by Lemma A.9, $\lim_{y\to\infty} \rho(y) = 1$ by Lemma A.13 and $\lim_{y\to\infty} \phi(y) = 1/2$ by Lemma A.14. So $\lim_{y\to\infty} b(y) - 2c(y) - \phi(y)(a(y) - b(y)) = \lim_{y\to\infty} y^2e^{-y} - \rho(y)(a(y) - b(y)) = 0$ and

$$\begin{aligned}
\lim_{y\to\infty} t(y) &= -2\lim_{y\to\infty} \frac{b'(y) - 2c'(y) - \phi'(y)(a(y) - b(y)) - \phi(y)(a'(y) - b'(y))}{2ye^{-y} - y^2e^{-y} - \rho'(y)(a(y) - b(y)) - \rho(y)(a'(y) - b'(y))}\\
&= -2\lim_{y\to\infty} \frac{\dfrac{a(y) - b(y)}{y} - 2\dfrac{b(y) - 2c(y)}{y} - \phi'(y)(a(y) - b(y)) - \phi(y)\left(ye^{-y} - \dfrac{a(y) - b(y)}{y}\right)}{2ye^{-y} - y^2e^{-y} - \rho'(y)(a(y) - b(y)) - \rho(y)\left(ye^{-y} - \dfrac{a(y) - b(y)}{y}\right)}.\\
&= -2\lim_{y\to\infty} \frac{a(y) - b(y) - 2(b(y) - 2c(y)) - \phi'(y)y(a(y) - b(y)) - \phi(y)\left(y^2e^{-y} - (a(y) - b(y))\right)}{2y^2e^{-y} - y^3e^{-y} - \rho'(y)y(a(y) - b(y)) - \rho(y)\left(y^2e^{-y} - (a(y) - b(y))\right)}
\end{aligned}$$

By Lemma A.7, $a(y) - b(y) > 0$ for $y \neq 0$, and so

$$\lim_{y\to\infty} t(y) = -2\lim_{y\to\infty} \frac{1 - 2\frac{b(y)-2c(y)}{a(y)-b(y)} - \phi'(y)y - \phi(y)\left(\frac{y^2e^{-y}}{a(y)-b(y)} - 1\right)}{\frac{2y^2e^{-y}-y^3e^{-y}}{a(y)-b(y)} - \rho'(y)y - \rho(y)\left(\frac{y^2e^{-y}}{a(y)-b(y)} - 1\right)}.$$

Now,

$$\begin{aligned}
\lim_{y\to\infty} \rho'(y)y &= \lim_{y\to\infty} \left(\frac{a'(y)}{b(y)} - \frac{a(y)b'(y)}{b(y)^2}\right)y = \lim_{y\to\infty} \left(\frac{y^2e^{-y}}{b(y)} - \frac{a(y)(a(y) - b(y))}{b(y)^2}\right)\\
&= -\lim_{y\to\infty} \rho(y)(\rho(y) - 1) = 0.
\end{aligned}$$

And

$$\begin{aligned}
\lim_{y\to\infty} \phi'(y)y &= \lim_{y\to\infty} \frac{y}{b(y)^2}\left(c'(y)b(y) - c(y)b'(y)\right) = \lim_{y\to\infty} \frac{1}{b(y)^2}((b(y) - 2c(y))b(y) - c(y)(a(y) - b(y)))\\
&= \lim_{y\to\infty} (1 - 2\phi(y)) - \lim_{y\to\infty} \phi(y)(\rho(y) - 1) = 0.
\end{aligned}$$

Further, by (A.27)

$$\lim_{y\to\infty}\frac{y^2e^{-y}}{a(y)-b(y)}=\lim_{y\to\infty}\frac{y^3}{2e^y-2y-y^2-2}=\lim_{y\to\infty}\frac{3y^2}{2e^y-2-2y}=\lim_{y\to\infty}\frac{6y}{2e^y-2}=0.$$

Similar arguments show that $\lim_{y\to\infty}y^3e^{-y}/a(y)-b(y)=0$.

Finally, we argue that

$$\lim_{y\to\infty}\frac{b(y)-2c(y)}{a(y)-b(y)}=1. \tag{A.30}$$

Observe that

$$\begin{aligned}b(y)-2c(y)&=\frac{2y-6+e^{-y}\left(y^2+4y+6\right)}{y^2}\quad\text{and}\\ a(y)-b(y)&=\frac{2(1-e^{-y})-(2+y)ye^{-y}}{y}.\end{aligned} \tag{A.31}$$

Thus,

$$\begin{aligned}\lim_{y\to\infty}\frac{b(y)-2c(y)}{a(y)-b(y)}&=\lim_{y\to\infty}\frac{2y-6+e^{-y}\left(y^2+4y+6\right)}{y\left(2(1-e^{-y})-(2+y)ye^{-y}\right)}\\ &=\lim_{y\to\infty}\frac{2-e^{-y}\left(y^2+4y+6\right)+e^{-y}(2y+4)}{y(2e^{-y}-(2+2y)e^{-y}+(2+y)ye^{-y})+\left(2(1-e^{-y})-(2+y)ye^{-y}\right)}\\ &=\lim_{y\to\infty}\frac{2-e^{-y}\left(2+2y+y^2\right)}{2-e^{-y}(2+2y+y^2-y^3)}=1.\end{aligned}$$

It follows that

$$\lim_{y\to\infty}t(y)=-2\lim_{y\to\infty}\frac{1-2\frac{b(y)-2c(y)}{a(y)-b(y)}-\phi'(y)y-\phi(y)\left(\frac{y^2e^{-y}}{a(y)-b(y)}-1\right)}{\frac{2y^2e^{-y}-y^3e^{-y}}{a(y)-b(y)}-\rho'(y)y-\rho(y)\left(\frac{y^2e^{-y}}{a(y)-b(y)}-1\right)}=1.$$

To see that $t(\cdot)$ is increasing, observe that

$$\begin{aligned}\rho'(y)&=\frac{a'(y)b(y)-a(y)b'(y)}{b(y)^2}\\ \rho''(y)&=\frac{a''(y)b(y)-a(y)b''(y)}{b(y)^2}-2\frac{a'(y)b(y)-a(y)b'(y)}{b(y)^3}b'(y)=\frac{1}{y^6b(y)^3}f_8(y)\ \text{and}\\ \phi''(y)&=\frac{1}{y^6b(y)^3}f_9(y),\end{aligned}$$

where

$$\begin{aligned}f_8(y)&=y^6\left(\left(a''(y)b(y)-a(y)b''(y)\right)b(y)-2\left(a'(y)b(y)-a(y)b'(y)\right)b'(y)\right)\ \text{and}\\ f_9(y)&=y^6\left(\left(c''(y)b(y)-c(y)b''(y)\right)b(y)-2\left(c'(y)b(y)-c(y)b'(y)\right)b'(y)\right).\end{aligned}$$

Thus,

$$t'(y)=-2\frac{\phi''(y)\rho'(y)-\phi'(y)\rho''(y)}{\rho'(y)^2}=-2b(y)^4\frac{\phi''(y)\rho'(y)-\phi'(y)\rho''(y)}{\left(a'(y)b(y)-a(y)b'(y)\right)^2}=\frac{b(y)}{y^6}\frac{f_{10}(y)}{\left(a'(y)b(y)-a(y)b'(y)\right)^2},$$

where

$$\begin{aligned}f_{10}(y)&=-2\left(f_9(y)\rho'(y)-\phi'(y)f_8(y)\right)\\ &=24-2e^{-y}\left(36+36y-18y^2+26y^3-7y^4+y^5\right)+e^{-2y}\left(72+144y+32y^3+22y^4+4y^5+y^6\right)\\ &\quad-4e^{-3y}\left(6+18y+9y^2+y^3\right).\end{aligned}$$

We argue that $f_{10}(y)>0$ for $y\neq 0$.

Observe that $f_{10}(0)=0$ and $f'_{10}(y)=2e^{-3y}yf_{10.1}(y)$, where

$$f_{10.1}(y)=6\left(12+8y+y^2\right)+e^{2y}\left(72-96y+54y^2-12y^3+y^4\right)-e^y\left(144-48y-12y^2+12y^3+y^4+y^5\right).$$

To show that $f_{10}(y)>0$ for $y\neq 0$, we show that $f'_{10}(y)$ has the same sign as y by showing that $f_{10.1}(y)>0$ for $y\neq 0$.

Observe that

$$\begin{aligned}
f'_{10.1}(y) &= 12(4+y)+2e^{2y}\left(24-42y+36y^2-10y^3+y^4\right)-e^y\left(96-72y+24y^2+16y^3+6y^4+y^5\right)\\
f''_{10.1}(y) &= 12+4e^{2y}\left(3-6y+21y^2-8y^3+y^4\right)-e^y\left(24-24y+72y^2+40y^3+11y^4+y^5\right)\\
f'''_{10.1}(y) &= e^y yf_{10.2}(y),
\end{aligned}$$

where

$$f_{10.2}(y)=-120-192y-84y^2-16y^3-y^4+8e^y\left(15+9y-6y^2+y^3\right)$$

and $f'_{10.1}(0)=f''_{10.1}(0)=f'''_{10.1}(0)=f_{10.2}(0)=0$.

Observe that

$$\begin{aligned}
f'_{10.2}(y) &= 4\left(-48-42y-12y^2-y^3+2e^y\left(24-3y-3y^2+y^3\right)\right)\\
f''_{10.2}(y) &= 4\left(-3\left(14+8y+y^2\right)+2e^y\left(21-9y+y^3\right)\right)\\
f'''_{10.2}(y) &= 8\left(-3(4+y)+e^y\left(12-9y+3y^2+y^3\right)\right)\\
f''''_{10.2}(y) &= 8\left(-3+e^y\left(3-3y+6y^2+y^3\right)\right)\\
f'''''_{10.2}(y) &= 8e^y y\left(9+9y+y^2\right)
\end{aligned}$$

and $f'_{10.2}(0)=f''_{10.2}(0)=f'''_{10.2}(0)=f''''_{10.2}(0)=f'''''_{10.2}(0)=0$.

Note that $f'''''_{10.2}(\cdot)$ has the three roots zero, $y_1=-3/2(3+\sqrt{5})\approx-7.8541$, and $y_2=-3/2(3-\sqrt{5})\approx-1.1459$, and so $f'''''_{10.2}(y)<0$ for $y<y_1$, $f'''''_{10.2}(y)>0$ for $y_1<y<y_2$, $f'''''_{10.2}(y)<0$ for $y_2<y<0$, and $f'''''_{10.2}(y)>0$ for $y>0$.

Because $\lim_{y\to-\infty}f''''_{10.2}(y)=-24$, $f''''_{10.2}(y_1)=-12e^{-3/2(3+\sqrt{5})}(25+15\sqrt{5}+2e^{3/2(3+\sqrt{5})})\approx-24.2727$, $f''''_{10.2}(y_2)=12(-2+5(-5+3\sqrt{5})e^{-3/2(3-\sqrt{5})})\approx 8.5862$, and $f''''_{10.2}(0)=0$, we see that $f''''_{10.2}(\cdot)$ has the two roots, $y_3\approx-2.26945$ and zero, $f''''_{10.2}(y)<0$ for $y<y_3$, $f''''_{10.2}(y)>0$ for $y_3<y<0$, and $f''''_{10.2}(y)>0$ for $y>0$.

Because $\lim_{y\to-\infty}f'''_{10.2}(y)=\infty$, $f'''_{10.2}(y_3)\approx-11.6077$, and $f'''_{10.2}(0)=0$, we see that $f'''_{10.2}(\cdot)$ has two roots, $y_4\approx-3.7090$ and zero, and $f'''_{10.2}(y)>0$ for $y<y_4$, $f'''_{10.2}(y)<0$ for $y_4<y<0$, and $f'''_{10.2}(y)>0$ for $y>0$.

Because $\lim_{y\to-\infty}f''_{10.2}(y)=-\infty$, $f''_{10.2}(y_4)\approx 23.6419$, and $f''_{10.2}(0)=0$, we see that $f''_{10.2}(\cdot)$ has two roots, $y_5\approx-5.31662$ and zero, and $f''_{10.2}(y)<0$ for $y<y_5$, $f''_{10.2}(y)>0$ for $y_5<y<0$, and $f''_{10.2}(y)>0$ for $y>0$.

Because $\lim_{y\to-\infty}f'_{10.2}(y)=\infty$, $f'_{10.2}(y_5)\approx-62.1332$, and $f'_{10.2}(0)=0$, we see that $f'_{10.2}(\cdot)$ has two roots, $y_6\approx-6.99119$ and zero, and $f'_{10.2}(y)>0$ for $y<y_6$, $f'_{10.2}(y)<0$ for $y_6<y<0$, and $f'_{10.2}(y)>0$ for $y>0$.

Because $\lim_{y\to-\infty}f_{10.2}(y)=-\infty$, $f_{10.2}(y_6)\approx 190.006$, and $f_{10.2}(0)=0$, we see that $f_{10.2}(\cdot)$ has two roots, $y_7\approx-8.69234$ and zero, and $f_{10.2}(y)<0$ for $y<y_7$, $f_{10.2}(y)>0$ for $y_7<y<0$, and $f_{10.2}(y)>0$ for $y>0$.

Thus, $f'''_{10.1}(y)=e^y yf_{10.2}(y)$ has the roots $y_7\approx-8.6923$ and zero, $f'''_{10.1}(y)>0$ for $y<y_7$, $f'''_{10.1}(y)<0$ for $y_7<y<0$, and $f'''_{10.1}(y)>0$ for $y>0$.

Because $\lim_{y\to-\infty}f''_{10.1}(y)=12$, $f''_{10.1}(y_7)\approx 13.2476$, and $f''_{10.1}(0)=0$, we see that $f''_{10.1}(\cdot)$ has only one root, $y=0$, and $f''_{10.1}(y)>0$ for $y\neq 0$. Because $f'_{10.1}(0)=0$, it follows that $f'_{10.1}(y)$ has the same sign as y, and because $f_{10.1}(0)=0$, it follows that $f_{10.1}(y)>0$ for $y\neq 0$. It follows that $f_{10}(y)$ has the same sign as y, and so we see that $f_{10}(y)>0$ and also $t'(y)>0$ for $y\neq 0$.

Lemma A.19. *For given $w\neq 0$, $\zeta(y,n)\le Z(y,w,n)$ for all $y>w$ if and only if*

$$n\ge T(w)=\frac{\eta(w)+\chi(w)}{\alpha(w)-\eta(w)},$$

where

$$\begin{aligned}
\zeta(y,n) &= \rho(y)n+2\phi(y),\\
Z(y,w,n) &= P(y,w)n-\Phi(y,w),\\
P(y,w) &= (\alpha(y)-\alpha(w))/(\eta(y)-\eta(w)),\\
\Phi(y,w) &= (\chi(y)-\chi(w))/(\eta(y)-\eta(w)),\\
\chi(y) &= \left(2e^{-y}+2y-y^2-2\right)/y^3,\\
\phi(y) &= c(y)/b(y),
\end{aligned}$$

$$\begin{aligned}\rho(y) &= a(y)/b(y),\\ \alpha(y) &= (1-e^{-y})/y,\\ \eta(y) &= (e^{-y}+y-1)/y^2,\\ a(y) &= 1-e^{-y}-ye^{-y},\\ b(y) &= (y-2+e^{-y}(y+2))/y \text{ and}\\ c(y) &= -\left(2y-3-y^2/2+3e^{-y}+ye^{-y}\right)/y^2.\end{aligned}$$

Proof. By Lemma A.16, $\zeta(y,n) \le Z(y,w,n)$ for all $y > w$ if and only if $\lim_{y\to\infty}\zeta(y,n) \le \lim_{y\to\infty} Z(y,w,n)$. But $\lim_{y\to\infty}\zeta(y,n) = n+1$ by Lemma A.13 and $\lim_{y\to\infty} Z(y,w,n) = \alpha(w)/\eta(w)n - \chi(w)/\eta(w)$ by Lemmas A.12 and A.17. And so $\zeta(y,n) \le Z(y,w,n)$ for all $y > w$ if and only if $n+1 \le \alpha(w)/\eta(w)n - \chi(w)/\eta(w)$. Because $\alpha(w) > \eta(w)$ by Lemma A.10, this is equivalent to $n \ge T(w)$.

Lemma A.20. *When $n < T(w)$,*

$$\lim_{y\to\infty} Z(y,w,n) = \frac{\alpha(w)}{\eta(w)}n - \frac{\chi(w)}{\eta(w)} \le \frac{wn+2}{e^w-1} + 2 - \frac{2}{w},$$

where

$$\begin{aligned}Z(y,w,n) &= P(y,w)n - \Phi(y,w),\\ P(y,w) &= (\alpha(y)-\alpha(w))/(\eta(y)-\eta(w)),\\ \Phi(y,w) &= (\chi(y)-\chi(w))/(\eta(y)-\eta(w)),\\ \chi(y) &= \left(2e^{-y}+2y-y^2-2\right)/y^3,\\ \alpha(y) &= (1-e^{-y})/y \text{ and}\\ \eta(y) &= (e^{-y}+y-1)/y^2.\end{aligned}$$

Proof. Observe that $\lim_{y\to\infty} Z(y,w,n) = \alpha(w)/\eta(w)n - \chi(w)/\eta(w)$ by Lemmas A.12 and A.17. To see that

$$\frac{\alpha(w)}{\eta(w)}n - \frac{\chi(w)}{\eta(w)} - \left(\frac{wn+2}{e^w-1} + 2 - \frac{2}{w}\right) \le 0 \tag{A.32}$$

when $n < T(w)$, observe that, because

$$\frac{\alpha(w)}{\eta(w)} - \frac{w}{e^w-1} = e^w\frac{\eta(-w)}{a(-w)}\frac{w^3}{e^w-1} > 0$$

by Lemmas A.7 and A.10, the left-hand side of (A.32) is increasing in n, and when $n = T(w)$ the left-hand side of (A.32) is zero.

Lemma A.21. *When $v \ne 0$ and $0 \le x < z < S$,*

$$q_6(v,x,z,S) = 1 + e^{\frac{2v}{\sigma^2}(S-x)}\frac{1-\frac{2v}{\sigma^2}(z-x)+\frac{2v^2}{\sigma^4}(S-z)^2 - e^{\frac{2v}{\sigma^2}(S-z)}\left(1-\frac{2v}{\sigma^2}(S-x)+\frac{2v^2}{\sigma^4}\left((S-x)^2-(z-x)^2\right)\right)}{-1+e^{\frac{2v}{\sigma^2}(S-z)} - e^{\frac{2v}{\sigma^2}(S-z)}\frac{2v}{\sigma^2}(S-z)} > 0.$$

Proof. Observe that $q_6(v,x,z,S) = -q_7(2v/\sigma^2, z-x, S-z)/a(-2v/\sigma^2(S-z))$, where

$$a(y) = 1-e^{-y}-ye^{-y} \text{ and}$$

$$q_7(u,y,b) = -1+e^{ub}(1-ub)+e^{u(y+b)}\left(1-uy+\frac{u^2}{2}b^2-e^{ub}\left(1-u(y+b)+\frac{u^2}{2}\left((y+b)^2-y^2\right)\right)\right).$$

Because $a(y) > 0$ for all $y \ne 0$ by Lemma A.7, we wish to show that $q_7(u,a,b) < 0$ for all $u \ne 0$ and y, $b > 0$. Now,

$$q_7(u,0,b) = -1+e^{ub}(1-ub)+e^{ub}\left(1+\frac{u^2}{2}b^2-e^{ub}\left(1-ub+\frac{u^2}{2}b^2\right)\right) < 0$$

for all $u \ne 0$ and $b > 0$. Further, observe that

$$\partial_y q_7(u,y,b) = \frac{1}{2}e^{(y+b)u}u^2 q_{7.1}(u,y,b),$$

where

$$q_{7.1}(u,y,b) = -\Big(b^2(e^{bu}-1)u + 2y\big(1+e^{bu}(bu-1)\big)\Big).$$

Note that $q_{7.1}(u,0,b) = -b^2(e^{bu}-1)u < 0$ for $u \neq 0$ and $b > 0$ and $\partial_y q_{7.1}(u,y,b) = -2a(-ub) < 0$ for $u \neq 0$ and $b > 0$. Thus, we see that $q_{7.1}(u,y,b) < 0$ for $u \neq 0$ and y, $b > 0$ and so $q_7(u,a,b) < 0$ for $u \neq 0$ and y, $b > 0$, proving that $q_6(v,x,z,S) > 0$ for $v \neq 0$ and $0 \leq x < z < S$.

Lemma A.22. *When $u \neq 0$ and $s < z < x$,*

$$q_5(u,s,z,x) = -1 + e^{\frac{2u}{\sigma^2}(x-s)} - \frac{2u(x-z)}{\sigma^2} + \frac{2u^2}{\sigma^4}(z-s)^2 \frac{e^{\frac{2u}{\sigma^2}(z-s)}-1}{1-e^{\frac{2u}{\sigma^2}(z-s)} + \frac{2u}{\sigma^2}(z-s)} > 0.$$

Proof. Observe that $q_5(u,s,z,z) > 0$ for all $u \neq 0$ and $s < z$ and $\partial_x q_5(u,s,z,x) = 2u/\sigma^2\big(e^{2u/\sigma^2(x-s)}-1\big) > 0$ for all $u \neq 0$ and $s < z$.

A.4. Lemma A.23

Lemma A.23. *Suppose $0 \neq u < v \neq 0$ and C_u and C_v are defined by $g(s,u,\gamma) = g(s,v,\gamma) = -V$ for some point s, value $V > p$, and scalar γ. If $g(S,u,\gamma) = g(S,v,\gamma)$ and $\partial_x g(S,u,\gamma) = \partial_x g(S,v,\gamma)$ for some point $S > s$, then $h > 0$ and $\partial_x g(S,u,\gamma) < 0$. Similarly, if C_u and C_v are defined by $g(S,u,\gamma) = g(S,v,\gamma) = V > -p$, $g(s,u,\gamma) = g(s,v,\gamma)$ and $\partial_x g(s,u,\gamma) = \partial_x g(s,v,\gamma)$, then $h < 0$ and $\partial_x g(s,u,\gamma) < 0$.*

Proof. We first consider the case $g(s,u,\gamma) = g(s,v,\gamma) = -V < -p$, $g(S,u,\gamma) = g(S,v,\gamma)$, and $\partial_x g(S,u,\gamma) = \partial_x g(S,v,\gamma)$. Because $g(S,u,\gamma) = g(S,v,\gamma)$ and $\partial_x g(S,u,\gamma) = \partial_x g(S,v,\gamma)$, we have, by Lemma 1, that $g(S,u,\gamma) = g(S,v,\gamma) = -p$. Because $g(s,u,\gamma) - g(s,v,\gamma) = 0$ and $g(S,u,\gamma) - g(S,v,\gamma) = 0$, the mean value theorem ensures there is a point $x_1 \in (s,S)$ such that $\partial_x g(x_1,u,\gamma) - \partial_x g(x_1,v,\gamma) = 0$, and invoking the mean value theorem a second time, we see that there is a point $x_2 \in (x_1,S)$ such that $\partial_{x,x} g(x_2,u,\gamma) - \partial_{x,x} g(x_2,v,\gamma) = 0$. Clearly, x_2 must be the unique real root

$$x_2 = \frac{\sigma^2}{2(u-v)} \log\left(\frac{u^2 C_u}{v^2 C_v}\right) < S$$

of $\partial_{x,x} g(\cdot,u,\gamma) - \partial_{x,x} g(\cdot,v,\gamma)$, and C_u and C_v must be nonzero and have the same signs. Further, because $g(S,u,\gamma) = g(S,v,\gamma) = -p$,

$$C_\mu = e^{\frac{2\mu}{\sigma^2}S}\left(\frac{hS}{\mu} - \frac{h\sigma^2}{2\mu^2} - \frac{\gamma}{\mu}\right) \text{ for } \mu \in \{u,v\}. \tag{A.33}$$

Thus, we see that

$$x_2 = \frac{\sigma^2}{2(u-v)} \log\left(\frac{u^2 C_u}{v^2 C_v}\right) = \frac{\sigma^2}{2(u-v)} \log\left(e^{\frac{2(u-v)}{\sigma^2}S} \frac{(hS-\gamma)u - \frac{h\sigma^2}{2}}{(hS-\gamma)v - \frac{h\sigma^2}{2}}\right) < S, \tag{A.34}$$

from which it follows that

$$\frac{(hS-\gamma)u - \frac{h\sigma^2}{2}}{(hS-\gamma)v - \frac{h\sigma^2}{2}} > 1. \tag{A.35}$$

Note further that, because $g(\cdot,u,\gamma)$ is the general solution to (16),

$$\frac{\sigma^2}{2} g(x_1,u,\gamma) + ug(x_1,u,\gamma) + hx_1 + pu = \gamma = \frac{\sigma^2}{2} g(x_1,v,\gamma) + vg(x_1,v,\gamma) + hx_1 + pv,$$

from which we conclude that

$$u(g(x_1,u,\gamma) + p) = v(g(x_1,v,\gamma) + p). \tag{A.36}$$

We consider the two cases: $\gamma < hS$ and $\gamma \geq hS$.

Case $\gamma < hS$: In this case, by Lemma 2, the point $x_p(u) \in (s,S)$ and $\partial_x g(x_p(u),u,\gamma) > 0$. Because it must be either convex or concave, we see that $g(\cdot,u,\gamma)$ is concave and $\partial_x g(S,u,\gamma) < 0$. To see that $h > 0$, observe that, by (A.33),

$$\partial_x g(x,u,\gamma) = \frac{2}{\sigma^2}\left(\frac{e^{\frac{2u}{\sigma^2}(S-x)} - 1}{u}\frac{h\sigma^2}{2} - e^{\frac{2u}{\sigma^2}(S-x)}(hS-\gamma)\right),$$

from which we see that, if $h \leq 0$, then $\partial_x g(x,u,\gamma) \leq 0$ for all $x < S$, contradicting the fact that $x_p(u) \in (s,S)$ and $\partial_x g(x_p(u),u,\gamma) > 0$.

Case $\gamma \geq hS$: In this case, the fact that $g(S,u,\gamma) = -p$ implies that

$$\partial_x g(S,u,\gamma) = \frac{2}{\sigma^2}(\gamma - hS) \geq 0 \text{ and}$$
$$\partial_{x,x} g(S,u,\gamma) = \frac{4}{\sigma^4}\left((hS-\gamma)u - \frac{h\sigma^2}{2}\right)$$

from which we see that $S = x_p(u)$ and $g(\cdot,u,\gamma)$ is convex if $(hS-\gamma)u - h\sigma^2/2 > 0$ and concave if $(hS-\gamma)u - h\sigma^2/2 < 0$. Note that, by (A.34), $(hS-\gamma)\mu - h\sigma^2/2 \neq 0$ for $\mu \in \{u,v\}$.

And, because $x_1 \in (s,S)$, $g(s,u,\gamma) = g(s,v,\gamma) < -p$, $g(S,u,\gamma) = g(S,v,\gamma) = -p$, and $\partial_x g(S,u,\gamma) = \partial_x g(S,v,\gamma) > 0$, $g(x_1,u,\gamma) < -p$ and $g(x_1,v,\gamma) < -p$.

We consider the two subcases: $(hS-\gamma)v - h\sigma^2/2 < 0$ and $(hS-\gamma)v - h\sigma^2/2 > 0$.

Subcase $\gamma \geq hS$ and $(hS-\gamma)v - h\sigma^2/2 < 0$:
In this case, by (A.35), $(hS-\gamma)u < (hS-\gamma)v$, contradicting the fact that $v > u$.

Subcase $\gamma \geq hS$ and $(hS-\gamma)v - h\sigma^2/2 > 0$:
In this case,

$$(hS-\gamma)u > (hS-\gamma)v > \frac{h\sigma^2}{2}. \tag{A.37}$$

We argue that $0 > g(x_1,v,\gamma) + p > g(x_1,u,\gamma) + p$, and so by (A.36), $0 < u < v$ and $h < 0$. Because $\partial_x g(s,u,\gamma) - \partial_x g(s,v,\gamma) = \frac{2}{\sigma^2}(v-u)(p-V) < 0$, $g(x,v,\gamma) > g(x,u,\gamma)$ for each $x \in (s,x_1]$ unless there is a point $r_1 \in (s,x_1)$ where the two functions cross, that is, $g(r_1,u,\gamma) = g(r_1,v,\gamma)$. To see that the functions cannot cross in $(s,x_1]$, suppose $r_1 \in (s,x_1]$ is the first point at which they cross; then, we see by Lemma 1 that because $\partial_x g(r_1,u,\gamma) \geq \partial_x g(r_1,v,\gamma)$, $g(r_1,u,\gamma) \geq -p$. But then $g(s,u,\gamma) < -p$, $g(r_1,u,\gamma) \geq -p$, $g(x_1,u,\gamma) < -p$, and $g(S,u,\gamma) = -p$, contradicting the fact that $g(\cdot,u,\gamma)$ must be either convex or concave. Thus, in particular, $0 > g(x_1,v,\gamma) + p > g(x_1,u,\gamma) + p$, and by (A.36), $0 < u < v$, and by (A.37), $h < 0$.

Finally, we argue that these facts imply $g(s,u,\gamma) > -p$, contradicting the assumption that $g(s,u,\gamma) = -V < -p$. To see this, observe that, by (A.33),

$$g(s,u,\gamma) = -p + \frac{1}{u^2}\left(e^{\frac{2u}{\sigma^2}(S-s)} - 1\right)\left((hS-\gamma)u - \frac{h\sigma^2}{2}\right) + \frac{h(S-s)}{u}. \tag{A.38}$$

Further, from the fact that $g(s,u,\gamma) = g(s,v,\gamma)$, $g(S,u,\gamma) = -p$, and $g(S,v,\gamma) = -p$, we see that

$$\gamma - hS = \frac{h\sigma^2}{2}\left(\frac{e^{\frac{2u}{\sigma^2}(S-s)} - e^{\frac{2v}{\sigma^2}(S-s)} - (u-v)\frac{2}{\sigma^2}(S-s)}{e^{\frac{2u}{\sigma^2}(S-s)}v - e^{\frac{2v}{\sigma^2}(S-s)}u + (u-v)} - \frac{1}{v} - \frac{1}{u}\right). \tag{A.39}$$

Combining (A.38) and (A.39), we see that

$$g(s,u,\gamma) = -p + \frac{h(S-s)}{u}\left(\left(e^{\frac{2u}{\sigma^2}(S-s)} - 1\right)\left(1 - \frac{u}{v}\right)\frac{e^{\frac{2v}{\sigma^2}(S-s)} - \frac{2v}{\sigma^2}(S-s) - 1}{e^{\frac{2u}{\sigma^2}(S-s)}\frac{2v}{\sigma^2}(S-s) - e^{\frac{2v}{\sigma^2}(S-s)}\frac{2u}{\sigma^2}(S-s) + \frac{2(u-v)}{\sigma^2}(S-s)} + 1\right). \tag{A.40}$$

The facts that $e^x - 1/x$ is increasing and greater than one for $x > 0$ and that $1/e^x - 1 - (1/x)$ is increasing for $x > 0$, together with a bit of algebra, prove that

$$(e^x - 1)\left(1 - \frac{x}{y}\right)\frac{e^y - y - 1}{e^x y - e^y x + x - y} < -1 \tag{A.41}$$

for all $0 < x < y$.

Now, combining (A.41) with the facts that $h < 0$, $0 < u < v$, and $s < S$ in (A.40), we see that $g(s,u,\gamma) > -p$, contradicting the assumption that $g(s,u,\gamma) = -V < -p$.

The arguments for the case $g(S,u,\gamma) = g(S,v,\gamma) = V > -p$, $g(s,u,\gamma) = g(s,v,\gamma)$, and $\partial_x g(s,u,\gamma) = \partial_x g(s,v,\gamma)$ are similar.

A.5. List of Symbols

- $AC(\Psi)$, the average cost of a given policy Ψ.
- $\beta(\cdot)$, the function identifying the drift rate to change to at the lower limit of a control band.
- $\mathbb{C}_k$, the set of all real-valued functions f on $[\theta,\Theta]$ such that the derivatives $f', f'', \ldots, f^{(k)}$ exist and are continuous except at a finite set of points.
- C_μ, the scalar defining the function $g(\cdot,\mu,\gamma) = C_\mu e^{-2\mu/\sigma^2 x} - h/\mu x + \gamma/\mu + h\sigma^2/2\mu^2 - p$.
- $D(v,u,s,z)$, the integral $\int_s^z g(x,v,\gamma) - g(x,u,\gamma)\,dx$.
- Δ_i, the area between successive members of a nondecreasing sequence $\delta = \{\delta(\cdot,i) : i \in \Lambda\}$ of functions. In particular, $\Delta_i = \int_\theta^\Theta \delta(x,i) - \delta(x,i-1)\,dx$.
- $G(\Psi)$, a directed graph defined on the control band policy Ψ.
- γ, the average cost of a given policy. We also use

— $\gamma(\Lambda,\theta,\Theta,U,M,K)$ for the optimal average cost for the average cost Brownian control problem ACBC$(\Lambda,\theta,\Theta,U,M,K)$.
— $\gamma(\Lambda,\theta,U)$, for example, when the other parameters of the problem are unchanged.
— $\gamma(u)$ as a shorthand for the more cumbersome $\gamma(\{u\},\theta,\Theta,U,M,K)$ when Λ consists of a single drift rate u.

- h, the holding cost charged on orders awaiting processing.
- K, the changeover cost funtion: $K(u, v)$ is the cost to change from drift rate u to drift rate v.
- K_i, the cost to change from drift rate $u(i)$ to drift rate $u(i-1)$ and back.
- $\kappa = \kappa_+ + \kappa_-$, where the cost to change from a drift rate u to a drift rate $v > u$ is $\kappa_+(v-u)$ and the cost to change from the faster rate v back to u is $\kappa_-(v-u)$.
- $L(\cdot)$, the instantaneous controls required as a function of time to keep the controlled diffusion process from falling below the lower limit θ.
- $\Lambda = \{u(i) : i = 1, 2, \ldots, n\}$ the finite set of nonzero drift rates with $u(i) < u(i+1)$ for $i = 1, 2, \ldots, n-1$.
- M, the cost per unit to reject work.
- $\mu(t)$, the drift rate of the diffusion process at time t.
- $\underline{\mu}(\Psi)$, the slowest recurrent drift rate of the unichain control band policy Ψ.
- $\overline{\mu}(\Psi)$, the fastest recurrent drift rate of the unichain control band policy Ψ.
- o, generally used to represent the adjusted lower limit of the process for a subproblem in the inductive argument.
- p, the processing cost charged per unit of capacity per unit time.
- $\mathcal{P}$, the space of all nonanticipating policies.
- Ψ, a nonanticipating policy. We occasionally use Ψ as a general nonanticipating policy defined by a sequence of stopping times, etc. We also use $\Psi = \{\psi(u) : u \in \Lambda\}$ to represent a specific control band policy with $\psi(u) = (u, s_u, S_u, \beta(u), \tau(u))$ defining the band for drift rate $u \in \Lambda$.
- $R(\cdot)$, the instantaneous controls required as a function of time to keep the controlled diffusion process from exceeding the upper limit Θ.
- $\mathcal{R}(\Psi)$, the set of recurrent drift rates of the ordered control band policy Ψ.
- s_μ, the lower limit of the control band for drift rate μ. When $\mu = u(i)$, the i^{th} drift rate, we often use s_i in place of $s_{u(i)}$.
- S_μ, the upper limit of the control band for drift rate μ. When $\mu = u(i)$, the i^{th} drift rate, we often use S_i in place of $S_{u(i)}$.
- σ^2, the variance parameter of the diffusion process.
- $\tau(\cdot)$, the function identifying the drift rate to change to at the upper limit of a control band.
- θ, the lower limit on the diffusion process.
- Θ, the finite upper limit on the diffusion process.
- U, the cost per unit to idle capacity.
- V, generally used to represent the adjusted cost of instantaneous controls at the lower limit of a subproblem in the inductive argument.
- $x_p(u)$, the point x satisfying $g(x, u, \gamma) = -p$ and $\partial_x g(x, u, \gamma) \geq 0$ when C_u and γ are understood from context. When C_u is not clear, we use the more general expression $x_p(u, Z, z, \gamma)$ to define C_u so that $g(z, u, \gamma) = -Z$.

References

Ahuja RK, Magnanti TL, Orlin JB (1993) *Network Flows: Theory, Algorithms, and Applications* (Prentice-Hall, Inc., Upper Saddle River, NJ).

Ata B, Harrison JM, Shepp LA (2005) Drift rate control of a Brownian processing system. *Ann. Appl. Probab.* 15(2):1145–1160.

Avram F, Karaesmen F (1996) A method for computing double band policies for switching between two diffusions. *Probab. Engrg. Inform. Sci.* 10(4):569–590.

Bather J (1968) A diffusion model for the control of a dam. *J. Appl. Probab.* 5(1):55–71.

Chernoff H, Petkau AJ (1978) Optimal control of a Brownian motion. *SIAM J. Appl. Math.* 34(4):717–731.

Chung KL, Williams RJ (1990) *An Introduction to Stochastic Integration. Probability and its Applications. Operator Theory, Advances and Applications*, 2nd ed. (Springer Science+Business Media, New York).

Corless RGH, Gonnet DEG, Hare DJ, Knuth D (1996) On the Lambert *W* function. *Adv. Comput. Math.* 5:329–359.

Euler L (1921) De serie lambertina plurimisque eius insignibus proprietatibus, leonhardi euleri opera omnia, ser. 1. *Opera Omnia* 6:350–369.

Ghosh AP, Weerasinghe AP (2007) Optimal buffer size for a stochastic processing network in heavy traffic. *Queueing Systems* 55:147–159.

Harrison JM (1985) *Brownian motion and stochastic flow systems* (Wiley, New York).

Lambert J (1758) Observationes variae in mathesin puram. *Acta Helvetica, physico-mathematico-anatomico-botanico-medica*, 3:128–168.

Lambert J (1772) Observations analytiques. Nouveaux mémoires de l'Académie royale des sciences et belles-lettres, 1 (Berlin).

Ormeci Matoglu M, Vande Vate J (2011) Drift control with changeover costs. *Oper. Res.* 59(2):427–439.

Ormeci Matoglu M, Vande Vate J (2020), The economic average cost brownian control problem with proportional changeover costs. https://www.isye.gatech.edu/~jvandeva/DriftControlNote.pdf.

Ormeci Matoglu M, Vande Vate J, Wang H (2015) Solving the drift control problem. *Stochastic Systems* 5(2):324–371.

Ormeci Matoglu M, Vande Vate J, Yu H (2019) The economic average cost brownian control problem. *Adv. Appl. Probab.* 51(1):300–337.

Perry D, Bar-Lev SK (1989) A control of a Brownian storage system with two switchover drifts. *Stochastic Anal. Appl.* 7:103–115.

Puterman ML (2005) *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (Wiley, New York).

Rath JH (1977) The optimal policy for a controlled Brownian motion process. *SIAM J. Appl. Math.* 32(1):115–125.

Wu J, Chao X (2014) Optimal control of a Brownian production/inventory system with average cost criterion. *Math. Oper. Res.* 39(1):163–189.