SIAM J. FINANCIAL MATH.
Vol. 11, No. 4, pp. 1007–1062

Executive Stock Option Exercise with Full and Partial Information on a Drift Change Point*

Vicky Henderson†, Kamil Kladívko‡, Michael Monoyios§, and Christoph Reisinger§

Abstract. We analyze the optimal exercise of an American call executive stock option (ESO) written on a stock whose drift parameter falls to a lower value at a *change point*, an exponentially distributed random time independent of the Brownian motion driving the stock. Two agents, who do not trade the stock, have differing information on the change point and seek to optimally exercise the option by maximizing its discounted payoff under the physical measure. The first agent has full information and observes the change point. The second agent has partial information and filters the change point from price observations. This scenario is designed to mimic the positions of two employees of varying seniority, a fully informed executive and a partially informed less senior employee, each of whom receives an ESO. The partial information scenario yields a model under the observation filtration $\widehat{\mathbb{F}}$ in which the stock drift becomes a diffusion driven by the innovations process, an $\widehat{\mathbb{F}}$-Brownian motion also driving the stock under $\widehat{\mathbb{F}}$, and the partial information optimal stopping value function has two spatial dimensions. We rigorously characterize the free boundary PDEs for both agents, establish shape and regularity properties of the associated optimal exercise boundaries, and prove the smooth pasting property in both information scenarios, exploiting some stochastic flow ideas to do so in the partial information case. We develop finite difference algorithms to numerically solve both agents' exercise and valuation problems and illustrate that the additional information of the fully informed agent can result in exercise patterns which exploit the information on the change point, lending credence to empirical studies which suggest that privileged information of bad news is a factor leading to early exercise of ESOs prior to poor stock price performance.

Key words. optimal stopping, free boundary problems, executive stock options, American options, smooth pasting, stochastic flows, Kalman–Bucy filter

AMS subject classifications. 91G80, 93E11, 93E20

DOI. 10.1137/18M1222909

1. Introduction.

In this paper we consider two pure optimal stopping problems involving a constant volatility stock whose drift parameter suffers a *change point*. At an exponentially distributed random time θ (the change point), independent of the Brownian motion W driving the stock, its drift falls from its initial constant value μ_0 to a lower constant value $\mu_1 < \mu_0$. The two problems we study are distinguished by *full information*, in which the change point is observed, or by *partial information*, in which the change point is not observable, and so is filtered from observations of the stock price.

The optimal stopping problems arise from the exercise of an executive stock option (ESO),

*Received by the editors October 25, 2018; accepted for publication (in revised form) July 10, 2020; published electronically October 28, 2020.
https://doi.org/10.1137/18M1222909
†Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK (Vicky.Henderson@warwick.ac.uk).
‡School of Business, Örebro University, 701 82 Örebro, Sweden (kamil.kladivko@oru.se).
§Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Oxford University, Oxford, OX2 6GG, UK (monoyios@maths.ox.ac.uk, christoph.reisinger@maths.ox.ac.uk).

an American call on a stock that is not traded by the option holders. Such a scenario is sometimes referred to as a "pure buyer's position," wherein an agent acquires an option, is not able to hedge the option due to trading restrictions, and seeks only to optimally exercise the claim. The objective we use for this completely unhedgeable payoff is to maximize the discounted payoff under the physical measure $\mathbb{P}$ over stopping times of the agent's filtration. Our two ESO-holding agents thus differ only in the respective filtrations to which each has access, and one of our goals is to understand how this information differential affects their exercise strategies. Our aim is to capture a firm specific disastrous event, which happens at a random time and is immediately known by the firm's top executives, but it is not revealed publicly, at least not immediately, and thus it is unknown to less senior employees. Recent examples of such disastrous events could be the Volkswagen emissions scandal ("Dieselgate"), the Facebook–Cambridge Analytica data scandal, or Boeing 737 MAX groundings.

The first agent has "full information." He observes the change point process $Y \in \{0,1\}$ (the indicator that the change point has occurred) as well as the Brownian motion W, so his filtration, $\mathbb{F}$ (the "large" filtration, or *background filtration*), is the augmentation of the filtration generated by (W, Y). In this case, the (random) drift process of the stock is $\mu(Y)$, given by a linear function of the change point process $Y \in \{0,1\}$, such that at all times the drift is equal to one of the distinct values (μ_0 before the change point, μ_1 afterwards; see (2.5)).

The second agent has "partial information." She does not observe the change point and filters Y (and thus the change point) from stock price observations. The partially informed agent's filtration, $\widehat{\mathbb{F}}$ (the *observation filtration*), is thus the augmentation of the stock price filtration, and $\widehat{\mathbb{F}} \subset \mathbb{F}$. In this partial information scenario, the filtered change point process $\widehat{Y}$ turns out to be a diffusion in $[0,1]$ driven by the innovations process $\widehat{W}$, which is the $\widehat{\mathbb{F}}$-Brownian motion also driving the stock under the observation filtration. In this case, the random drift turns out to be $\mu(\widehat{Y})$, featuring the same linear function as in the full information case, but now of the filtered process $\widehat{Y}$ (see (2.11)). The process $\widehat{Y}$, adapted to the stock price filtration, turns out to be a functional of the path history of the stock price.

For both the full and partial information problems, we carry out a detailed and rigorous free boundary analysis of the associated value function for the option. For each problem this involves a classical program of steps, which we generalize from the (typical) constant drift case to each of our two random drift scenarios, as follows. The two-state drift of the full information problem naturally leads to a pair of value functions (one for each possible initial drift state $i \in \{0,1\}$) characterizing the ESO value. Equally naturally, in the partial information problem, dependent on the diffusion $\widehat{Y} \in [0,1]$, the value function depends on a variable $y \in [0,1]$, representing the initial value of the change point process (in addition to the usual temporal and stock price dependence).

We first derive basic convexity, monotonicity, and time decay properties of the value functions (Lemma 3.1 (full information) and Lemma 4.2 (partial information)), the latter using some stochastic flow ideas applied to $\widehat{Y}(y)$, the filtered change point process viewed as a function of its initial value y. From these results we infer the form of the continuation and stopping regions, the existence and form of optimal exercise thresholds, and (later) their limiting values as we approach the ESO maturity time.

We show that, for the full information problem, there is a pair of ordered, nonincreasing, time-dependent exercise boundaries $x_0^*(\cdot) \geq x_1^*(\cdot)$, such that optimal early exercise can occur in the state where the drift is $\mu_i, i \in \{0,1\}$, when the stock breaches $x_i^*(\cdot)$ from below, or if such a breach is triggered by the change point. On the other hand, in the partial information case the exercise boundary $x^*(\cdot,\cdot)$ is a *surface*, with an additional spatial, nonincreasing dependence on the variable $y \in [0,1]$, arising from the dependence of the drift on the filtered change point process, and such that the partial information exercise surface lies between the full information exercise thresholds. This can lead to an interesting range of possible exercise patterns (such as immediate exercise by the fully informed agent in response to the change point, a strategy unavailable to the agent who does not see the jump in drift), which we describe (and later examine numerically). We also consider how our stopping problems are changed with the inclusion of an option vesting period. In practice, vesting periods during which the option holder is not permitted to exercise are used by the company to maintain the employee's incentives or exposure to the stock price.

We then give a rigorous characterization of the ESO value functions in terms of free boundary PDEs (Proposition 3.5 (full information) and Proposition 4.6 (partial information)) with associated smooth pasting conditions at the exercise thresholds (Theorem 3.6 (full information) and Theorem 4.7 (partial information)). Using these results we are able to derive Doob–Meyer decompositions of the supermartingales which represent the discounted ESO value processes (Theorem 3.7 (full information) and Lemma 4.8 (partial information)). These in turn are used in proving the results on the limiting values of the boundaries as we approach maturity T (Proposition 3.4 (full information) and Lemma 4.5 (partial information)). Although not needed elsewhere, we also show that the boundaries for the full information problem are continuous over $[0,T)$, as stated in Proposition 3.4.

Our mathematical results are obtained by implementing, broadly speaking, the classical program for obtaining properties of American options (see, for example, Karatzas and Shreve [31, Chapter 2] for the American put in the Black–Scholes model) and carefully modifying and extending these arguments to our random drift scenarios, augmenting them in places with new tools, such as the stochastic flow ideas mentioned above. These results are novel compared to existing literature, as we now describe.

The full information case has some similarities with papers on American option valuation with regime switching, such as the infinite horizon put in Guo and Zhang [28] and the finite horizon put in Buffington and Elliott [8] (who assume all required regularity properties of the value function). Le and Wang [34] also treat the American put with regime switching, and do prove the smooth pasting property, by extending a fairly involved iterative procedure originally due to Bayraktar [4]. As well as being lengthy, some steps exploit the boundedness of the put payoff function, so it is not clear if they are directly applicable to our model. Here, therefore, we exploit our explicit one-switch scenario and show how more classical techniques can be extended to the random drift case, both for the free boundary characterization and then for the smooth pasting property. The latter requires an analysis of the optimal stopping time given a particular starting state, and here we use our derived structures for the stopping and continuation regions.

In the partial information case, our results are entirely new. The rigorous characterization of the value function as a solution of a free boundary PDE with an associated smooth pasting condition, has not been demonstrated before to the best of our knowledge. We achieve this, also show that the exercise surface is decreasing in time and in the initial value $y \in [0,1]$ of the filtered change point process, and give its limiting terminal value. An infinite horizon American put with partial information on a switching dividend process was studied by Gapeev [25], but the regularity of the value function and the smooth pasting property were assumed to hold. We resolve these issues in our partial information problem. Note that, with our objective of maximizing the discounted expected payoff under the physical measure, our problems map to conventional American option pricing problems under a martingale measure, but with a random dividend yield. Thus, our results also give the required regularity for the problems studied in [25].

Finally, there is a strand of papers (Décamps, Mariotti, and Villeneuve [14, 15], Klein [32], Ekström and Lu [19], and Ekström and Vannestål [20]) which study optimal stopping problems in a partial information scenario when a drift parameter is assumed to take on one of two values, but the agent is unsure which value pertains in reality. These models correspond to the limit that the parameter of the exponential time in our model approaches zero, so an explicit change point is absent (they are models of an uncertain drift, as opposed to uncertainty in the timing of a change of drift). This renders them simpler than our partial information model, because the dependence of the filtered process on the entire history of the stock disappears. These papers are then able to reduce the dimensionality of the problem under some circumstances, a simplification not available in our model.

We complete the picture by solving both problems numerically, using finite difference schemes, and carry out simulations to illustrate some of the exercise patterns that can occur. The partial information case is substantially more difficult numerically due to the second spatial dimension, but with a single Brownian driver, resulting in a reduced rank diffusion matrix, and the degeneracy of some of the diffusion and drift coefficients at certain boundaries of the domain. This setting requires a novel, tailored approximation scheme for the efficient numerical solution. We propose a first order monotone and a second order nonmonotone penalized backward differentiation formulae (BDF) scheme on nonuniform meshes and prove convergence for the former. Numerical tests demonstrate the stability and achievable accuracy for the scheme.

One of our motivations for studying these issues is a strand of literature in empirical finance which attributes early ESO exercise prior to poor stock performance in part to privileged information, particularly on imminent bad news. Early studies (Huddart and Lang [29] and Carpenter and Remmers [10]) provide some evidence that this is the case. More recent works that partition the exercises according to the particular exercise strategy employed find much stronger evidence of informed exercise (Brooks, Chance, and Cline [7], Cicero [12], and Aboody et al. [1]): exercises accompanied by a sale of stock are followed by negative abnormal returns (while other exercises are not). We were thus motivated to construct a model where complete or incomplete information on an adverse event could be compared in the exercise of an American call. Here, we think of the fully informed agent as a senior executive who observes the change point, while the partially informed agent is thought of as a less senior employee who is not privy to board meetings sharing imminent bad news. Our setup considers a stock price whose drift will jump to, and remain at, a lower value. We do not consider a model where the drift can switch repeatedly between two values, as this would not capture a seismic piece of adverse news, though a rigorous analysis of such a model would be interesting and could potentially be built upon our analysis here.

We use our model to conduct a study of mean postexercise returns for agents with full and partial information, motivated by the empirical work of Brooks, Chance, and Cline [7]. Our simulations (in section 7) support the conjecture that indeed the difference between average postexercise returns for fully and partially informed agents is significantly negative. For our simulations, the difference between mean postexercise returns for fully and partially informed agents varies between about -3.8% and -9.7%, depending on the expected stock return μ_0 and volatility, covering the range of values reported by Brooks, Chance, and Cline [7]. Our model thus provides theoretical support for the tests conducted in the empirical literature to evidence so-called insider exercises.

Our analysis leads to our being able to characterize exercise scenarios and to point out scenarios where the change point can induce exercise for the fully informed agent, but of course not necessarily for the partially informed agent, since the change point is not seen. We illustrate this in section 7, where we provide simulations of various exercise scenarios and show that the agent with full information has considerable advantage in exercise timing. An exercise surface $x^*(t, y)$, $t \in [0, T], y \in [0, 1]$, for the agent with partial information, and thresholds $x_0^*(t), x_1^*(t)$, $t \in [0, T]$, for the full information case are computed and shown to be consistent with the theoretical results in earlier sections.

The informational advantage demonstrated in the exercise strategies is reflected in the respective ESO values the agents place on their options. We document that the additional value the agent with full information places on his ESO is significant in magnitude. The early exercise value as a proportion of the European value can be very many times greater for the agent with full rather than only partial information. In Table 2, we also report comparative statics for the ESO value as we vary stock parameters μ_0, μ_1, σ, and λ. ESO values for both agents decrease as the expected return in the bad state, μ_1, decreases or there is a greater probability of a downward jump. However, the early exercise values increase, indicating that the ability to time the exercise of the option is more valuable when the expected return following the change point is worse, or when the chance of entering the bad state is higher. We also report ESO values when option vesting is included in the model and note that, as expected, the early exercise value drops for both agents, while the informational advantage of the agent with full information is still present.

The rest of the paper is organized as follows. In section 2 we introduce the model and the optimal stopping problems under both information scenarios, and we carry out a filtering procedure to derive the model dynamics with respect to the stock price filtration. In sections 3 and 4 we analyze the full and partial information problems, respectively. Section 5 gives a brief discussion of how a vesting period impacts upon exercise. In section 6 we construct and describe numerical methods for solving the two optimal stopping problems, including convergence results. We apply the finite difference methodology in section 7 to perform simulations to compare the exercise patterns of the agents, undertake an analysis of post-exercise returns, and provide ESO valuation.

2. Stock price with a drift change point.

We model a stock price whose drift will jump to a lower value at a random time (a *change point*). The goal is to investigate differences in the ESO exercise strategy between a fully informed agent who observes the change point and a partially informed agent who has to filter the change point from stock price observations. In particular, we seek to explore whether the fully informed agent can exploit his additional information in the exercise strategy.

The setting is a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} := (\mathcal{F}_t)_{t\in\mathbf{T}}$ satisfying the usual hypotheses of right-continuity and augmentation by all the $\mathbb{P}$-null sets of $\mathcal{F}$. The time set $\mathbf{T}$ will be the finite interval $\mathbf{T} = [0, T]$ for some $T < \infty$. The filtration $\mathbb{F}$ will sometimes be referred to as the *background filtration*. It represents the large filtration available to a perfectly informed agent, and all processes will be assumed to be $\mathbb{F}$-adapted in what follows.

Let W denote a standard $(\mathbb{P}, \mathbb{F})$-Brownian motion. Let $\theta \in \mathbb{R}_+$ be a nonnegative random time, independent of W, with initial distribution $\mathbb{P}[\theta = 0] =: y_0 \in [0, 1)$ and subsequent distribution

$$\mathbb{P}[\theta > t|\theta > 0] = \mathrm{e}^{-\lambda t}, \quad \lambda \geq 0, \quad t \in \mathbf{T}.$$

Thus, conditional on the event $\{\omega \in \Omega : \theta(\omega) > 0\} \equiv \{\theta > 0\}$, θ has exponential distribution with parameter λ. Define the single-jump càdlàg process Y by

$$Y_t := \mathbb{1}_{\{t\geq\theta\}}, \quad t \in \mathbf{T}, \tag{2.1}$$

so that $Y_0 = \mathbb{1}_{\{\theta=0\}}$ with $\mathbb{E}[Y_0] = y_0$. We may (and do) take $\mathbb{F}$ to be the $\mathbb{P}$-augmentation of $\mathbb{F}^{W,Y}$, the filtration generated by the pair (W, Y). By Karatzas and Shreve [30, Proposition 2.7.7] this filtration is indeed right-continuous, because (W, Y) is a strong Markov process.

We associate with Y the $(\mathbb{P}, \mathbb{F})$-martingale $M^{(Y)}$ (the compensated jump process), defined by

$$M_t^{(Y)} := Y_t - Y_0 - \lambda \int_0^t (1 - Y_s)\,\mathrm{d}s, \quad t \in \mathbf{T}. \tag{2.2}$$

A stock price process X with constant volatility $\sigma > 0$ has a drift which depends on the process Y. We are given two real constants $\mu_0 > \mu_1$ such that the drift value falls from μ_0 to the lower value μ_1 at the change point. Define the constant $\eta > 0$ by

$$\eta := \frac{\mu_0 - \mu_1}{\sigma}. \tag{2.3}$$

The stock price dynamics with respect to $(\mathbb{P}, \mathbb{F})$ are given by

$$\mathrm{d}X_t = (\mu_0 - \sigma\eta Y_t)X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t. \tag{2.4}$$

Thus, the drift process $\mu(Y)$ of the stock is given by

$$\mu(Y_t) := \mu_0 - \sigma\eta Y_t = \mu_0(1 - Y_t) + \mu_1 Y_t = \begin{cases} \mu_0 & \text{on} \quad \{t < \theta\} = \{Y_t = 0\}, \\ \mu_1 & \text{on} \quad \{t \geq \theta\} = \{Y_t = 1\}, \end{cases} \quad t \in \mathbf{T}. \tag{2.5}$$

Note in particular that for $y_0 = 0$ the change point θ is almost surely strictly positive, and the stock evolution almost surely begins with the higher drift value μ_0.

We assume that the values of the constants $y_0, \mu_0, \mu_1, \sigma, \lambda$ are given. Finally, there is also a cash account paying a constant interest rate $r \geq 0$. Dividends could also be included, and there are several possibilities as to how these could be modeled, but for simplicity we do not do so. For example, a constant dividend yield could be included with minor adjustments by reinterpreting the drifts as being the net of dividends.

We may write the stock price evolution as

$$\mathrm{d}X_t = \sigma X_t \,\mathrm{d}\xi_t, \tag{2.6}$$

where ξ is the volatility-scaled return process given by

$$\xi_t := \frac{1}{\sigma}\int_0^t \frac{\mathrm{d}X_s}{X_s} = \left(\frac{\mu_0}{\sigma}\right) t - \eta \int_0^t Y_s \,\mathrm{d}s + W_t =: \int_0^t h_s \,\mathrm{d}s + W_t, \quad t \in \mathbf{T}, \tag{2.7}$$

with the process h defined by

$$h_t := \frac{\mu_0}{\sigma} - \eta Y_t, \quad t \in \mathbf{T}, \tag{2.8}$$

so h and W are independent. The process ξ will be used as an observation process in a filtering algorithm in section 2.2.

Define the *observation filtration* $\widehat{\mathbb{F}} = (\widehat{\mathcal{F}}_t)_{t\in\mathbf{T}}$ as the $\mathbb{P}$-augmentation of the filtration generated by the stock price (equivalently by the process ξ in (2.7)):

$$\widehat{\mathcal{F}}_t := \sigma(\mathcal{F}_t^X \cup \mathcal{N}), \quad t \in \mathbf{T},$$

where $\mathcal{F}_t^X := \sigma(X_s : 0 \leq s \leq t)$, and $\mathcal{N}$ denotes the $\mathbb{P}$-null sets of $\mathcal{F}$. We have $\widehat{\mathbb{F}} \subset \mathbb{F}$ and, moreover, it turns out that the filtration $\widehat{\mathbb{F}}$ is right-continuous,[1] as we shall justify in Remark 2.3.

An ESO on X is an American call option with strike $K \geq 0$ and maturity T, and so has payoff $(X_t - K)^+$ if exercised at $t \in \mathbf{T}$. We assume the ESO holder receives the cash payoff on exercise. We consider two agents in this scenario, each of whom is awarded at time zero an ESO on X, and who have access to different filtrations but are identical in other respects. In practice, employees holding such ESOs are prohibited from trading the company stock X (see Carpenter [9] and section 16c of the Securities and Exchange Act), and this motivates our assumption that neither agent trades the stock.

The first agent has *full information*. He knows the values of all the model parameters and has full access to the background filtration $\mathbb{F}$, so in particular can observe the Brownian motion W and the one-jump process Y. The second agent has *partial information*. She also knows the values of the constant model parameters and observes the stock price X, but not the one-jump process Y. The partially informed agent's filtration is therefore the observation filtration $\widehat{\mathbb{F}}$. The only difference between the agents is that the partially informed agent does not know the value of the process Y, which she will filter from stock price observations.

[1]This is a consequence of the strong Markov property of the pair $(X, \widehat{Y})$, where $\widehat{Y}$ is the filtered estimate of Y given $\widehat{\mathbb{F}}$.

We have assumed that the stock volatility is constant and in particular does not depend on the single-jump process Y. If we allowed the volatility process to depend on Y, then with continuous stock price observations the partially informed agent could infer the value of Y from the rate of increase of the quadratic variation of the stock. This would remove the distinction between the agents and thus nullify our intention of building a model where the agents have distinctly different information on the performance of the stock. In principle, the constant volatility assumption could be relaxed to allow the volatility to depend on Y, but only at the expense of requiring a necessarily more complicated model of differential information between the agents. For instance, the partially informed agent could be rendered ignorant of the values μ_0, μ_1, so these could be modeled (for example) as random variables whose values would be filtered from price observations. This would have significant ramifications for the tractability of the ESO optimal stopping problems, and our constant volatility model is the simplest model one can envisage with differential information on a change point.

2.1. The ESO optimal stopping problems. We assume that each agent will maximize, over stopping times of their respective filtration, the discounted expectation of the ESO payoff under the physical measure $\mathbb{P}$. Given the absence of trading opportunities, the ESO payoff constitutes a completely unhedgeable claim, so the agents each face a pure exercise decision. In this case, for simplicity, we take the most straightforward objective possible. This objective was used in Monoyios and Ng [40], where ESO valuation with inside information was considered. It also appears in works which consider American options in the absence of classical hedging opportunities, sometimes called a pure buyer's position: an agent holds a long position in an American option but, for reasons of (say) liquidity or transaction costs, does not hedge this position (see Ekström and Vannestål [20], for example). If we were to allow the agents to trade other securities, one could envisage adding risk aversion by considering utility-based valuation and hedging, yielding combined optimal stopping and control problems. Such ESO problems have been considered for constant drift models by Leung and Sircar [35, 36] and Grasselli and Henderson [27] using classical utility, and by Leung, Sircar, and Zariphopoulou [37] using forward utility. These works take the required regularity of value functions as given. Utility-based valuation of European claims on nontraded assets in a random parameter framework has been considered by Monoyios [39], where both traded and nontraded assets are geometric Brownian motions with unobserved constant drifts modeled as Gaussian random variables. Filtering then leads to a random parameter basis risk model that is significantly less tractable than its constant parameter counterpart. As both our information models have random parameters, their rigorous treatment via a risk-averse utility-based methodology, including verification of regularity where needed, is an open problem left for future research. Our contribution here is thus to use our risk-neutral objective, in a random parameter framework, to give a fully rigorous free boundary PDE treatment of both the full and partial information ESO problems. The absence of risk aversion in our model gives us the tractability we need for our analysis and arguably focuses on the informational, as opposed to risk aversion, aspects of the agents' exercise and valuation decisions.

For $t \in [0, T]$, let $\mathcal{T}_{t,T}$ denote the set of $\mathbb{F}$-stopping times with values in $[t, T]$, and let $\widehat{\mathcal{T}}_{t,T}$ denote the corresponding set of $\widehat{\mathbb{F}}$-stopping times. For any such starting time $t \in [0, T]$, the fully informed agent's ESO value process is V, an $\mathbb{F}$-adapted process defined by

$$V_t := \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(X_\tau - K)^+ \middle| \mathcal{F}_t \right], \quad t \in [0,T]. \tag{2.9}$$

We shall call (2.9) the full information problem.

Similarly, the partially informed agent's ESO value process is U, an $\widehat{\mathbb{F}}$-adapted process defined by

$$U_t := \operatorname*{ess\,sup}_{\tau \in \widehat{\mathcal{T}}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(X_\tau - K)^+ \middle| \widehat{\mathcal{F}}_t \right], \quad t \in [0,T]. \tag{2.10}$$

We shall call (2.10) the partial information problem.

Naturally, the salient distinction between (2.9) and (2.10) is the filtration with respect to which the stopping time and essential supremum are defined. For the full information problem (2.9) the stock dynamics will be (2.4). For the partial information problem (2.10) we must derive the model dynamics under the observation filtration. This is done in section 2.2 below.

Recipients of company ESOs are often contractually restricted from exercising their options during a vesting period, $[0, t_v)$, so that stopping times may lie in the interval $[t_v, T]$; see, for example, Carpenter, Stanton, and Wallace [11]. Later, in section 5, we outline how the problems may be modified to incorporate vesting, and in section 7.3 we demonstrate the impact of vesting on ESO values.

Remark 2.1 (formal equivalence to random-dividend no-arbitrage valuation). The optimal stopping problems (2.9) and (2.10), formulated under the physical measure $\mathbb{P}$ with some random stock drift $\mu(\cdot)$, of course map formally to problems written under a martingale measure $\mathbb{Q}$ where the stock drift will be $r - \delta(\cdot)$, for some random dividend yield $\delta(\cdot)$, related to $\mu(\cdot)$ by $\mu(\cdot) = r - \delta(\cdot)$. The results we obtain are thus applicable to classical no-arbitrage valuation with a random dividend yield.

The scenario we have set up, with a drift value for a log-Brownian motion which switches at a random time to a new value, has obvious similarities with the so-called "quickest detection of a Wiener process" problem, which has a long history and is discussed in Chapter VI of Peskir and Shiryaev [41] (see Gapeev and Shiryaev [26] for a recent example involving diffusion processes). The difference between these problems and ours is that our objective functional will be the expected discounted payoff of an ESO, so errors in detecting the change point are transmitted through the prism of the ESO exercise decision. In contrast, the classical change point detection problem has some explicit objective functional which directly penalizes a detection delay or a false alarm (where the change point is incorrectly deduced to have occurred).

2.2. Dynamics under the observation filtration. Let the signal process be Y in (2.1), and take the observation process to be ξ in (2.7), with the augmented filtration generated by ξ equivalent to the augmented stock price filtration $\widehat{\mathbb{F}}$.

Introduce the notation $\widehat{\phi}_t := \mathbb{E}[\phi_t | \widehat{\mathcal{F}}_t]$, $t \in \mathbf{T}$, for any process ϕ. In particular, we are interested in the filtered estimate of Y, defined by

$$\widehat{Y}_t := \mathbb{E}[Y_t | \widehat{\mathcal{F}}_t], \quad t \in \mathbf{T}.$$

A standard filtering procedure gives the stock price dynamics with respect to the observation filtration $\widehat{\mathbb{F}}$, along with the dynamics of $\widehat{Y}$, resulting in the following lemma. We give a short proof for completeness.

Lemma 2.2 (observation filtration dynamics). *With respect to the observation filtration $\widehat{\mathbb{F}}$ the stock price follows*

$$\mathrm{d}X_t = (\mu_0 - \sigma\eta\widehat{Y}_t)X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}\widehat{W}_t, \tag{2.11}$$

where $\widehat{W}$ is the innovations process, given by

$$\widehat{W}_t := \xi_t - \int_0^t \widehat{h}_s\,\mathrm{d}s = \xi_t - \frac{\mu_0}{\sigma}t + \eta\int_0^t \widehat{Y}_s\,\mathrm{d}s, \quad t \in \mathbf{T}, \tag{2.12}$$

where, analogously to (2.8), $\widehat{h}_t := \frac{\mu_0}{\sigma} - \eta\widehat{Y}_t$, $t \in \mathbf{T}$, and $\widehat{W}$ is a $(\mathbb{P}, \widehat{\mathbb{F}})$-Brownian motion.

The filtered process $\widehat{Y}$ has dynamics given by

$$\mathrm{d}\widehat{Y}_t = \lambda(1 - \widehat{Y}_t)\,\mathrm{d}t - \eta\widehat{Y}_t(1 - \widehat{Y}_t)\,\mathrm{d}\widehat{W}_t, \quad \widehat{Y}_0 = \mathbb{E}[Y_0] = y_0 \in [0, 1). \tag{2.13}$$

Proof. We use the innovations approach to filtering, as discussed in Rogers and Williams [45, Chapter VI.8] or Bain and Crisan [2, Chapter 3], for instance.

By Theorem VI.8.4 in [45], the innovations process $\widehat{W}$, defined by (2.12), is a $(\mathbb{P}, \widehat{\mathbb{F}})$-Brownian motion. Using (2.12) in the stock price SDE (2.6) then yields (2.11).

It remains to prove (2.13). For any bounded, measurable test function f, write $f_t \equiv f(Y_t)$, $t \in \mathbf{T}$, for brevity. Define a process $(\mathcal{G}f_t)_{t\in\mathbf{T}}$, satisfying $\mathbb{E}\big[\int_0^t |\mathcal{G}f_s|^2\,\mathrm{d}s\big] < \infty$ for all $t \in \mathbf{T}$, such that

$$M_t^{(f)} := f_t - f_0 - \int_0^t \mathcal{G}f_s\,\mathrm{d}s, \quad t \in \mathbf{T},$$

is a $(\mathbb{P}, \mathbb{F})$-martingale. With h, W independent, we have the (Kushner–Stratonovich) fundamental filtering equation (see Theorem 3.30 in [2], for example)

$$\widehat{f}_t = \widehat{f}_0 + \int_0^t \widehat{\mathcal{G}f}_s\,\mathrm{d}s + \int_0^t \left(\widehat{f_s h_s} - \widehat{f}_s\widehat{h}_s\right)\mathrm{d}\widehat{W}_s, \quad t \in \mathbf{T}. \tag{2.14}$$

Take $f(y) = y$. Then the martingale $M^{(f)} = M^{(Y)}$, as defined in (2.2), so that $\mathcal{G}f = \lambda(1 - Y)$ and the filtering equation (2.14) reads as

$$\widehat{Y}_t = y_0 + \lambda\int_0^t (1 - \widehat{Y}_s)\,\mathrm{d}s + \int_0^t (\widehat{Y_s h_s} - \widehat{Y}_s\widehat{h}_s)\,\mathrm{d}\widehat{W}_s, \quad t \in \mathbf{T}, \tag{2.15}$$

where we have used $\widehat{Y}_0 = \mathbb{E}[Y_0] = y_0$.

Now,

$$\widehat{Y_t h_t} = \mathbb{E}\left[Y_t\left(\frac{\mu_0}{\sigma} - \eta Y_t\right)\middle|\,\widehat{\mathcal{F}}_t\right] = \left(\frac{\mu_0}{\sigma}\right)\widehat{Y}_t - \eta\mathbb{E}[Y_t^2|\widehat{\mathcal{F}}_t] = \left(\frac{\mu_0}{\sigma} - \eta\right)\widehat{Y}_t, \quad t \in \mathbf{T}, \tag{2.16}$$

the last equality a consequence of $Y^2 = Y$.

On the other hand,

$$\widehat{Y}_t\widehat{h}_t = \widehat{Y}_t\mathbb{E}\left[\frac{\mu_0}{\sigma} - \eta Y_t \middle| \widehat{\mathcal{F}}_t\right] = \left(\frac{\mu_0}{\sigma}\right)\widehat{Y}_t - \eta\left(\widehat{Y}_t\right)^2, \quad t \in \mathbf{T}. \tag{2.17}$$

Using (2.16) and (2.17) in (2.15) then yields the integral form of (2.13). ■

Remark 2.3 (right-continuity of observation filtration). Note that $\widehat{Y}$ in (2.13) is an $\widehat{\mathbb{F}}$-adapted diffusion in $[0, 1]$ with an absorbing state at $\widehat{Y} = 1$. Note also that, since observations of the stock price are sufficient to specify $\widehat{Y}$, the observation filtration is also the $\mathbb{P}$-augmentation of the filtration generated by the two-dimensional diffusion $(X, \widehat{Y})$. Then Karatzas and Shreve [30, Proposition 2.7.7] guarantee that $\widehat{\mathbb{F}}$ is right-continuous, as it is the augmented filtration generated by the strong Markov process $(X, \widehat{Y})$.

3. The full information ESO problem.

In this section we focus on the full information problem defined in (2.9). Define the (continuous) *reward process* R as the discounted payoff process:

$$R_t := \mathrm{e}^{-rt}(X_t - K)^+, \quad t \in \mathbf{T}. \tag{3.1}$$

The reward process is assumed to satisfy

$$\mathbb{E}\left[\sup_{t\in[0,T]} R_t\right] < \infty. \tag{3.2}$$

The discounted full information ESO value process is $\widetilde{V}$, given by

$$\widetilde{V}_t := \mathrm{e}^{-rt}V_t = \operatorname*{ess\,sup}_{t\in\mathcal{T}_{t,T}} \mathbb{E}[R_\tau|\mathcal{F}_t], \quad t \in \mathbf{T}. \tag{3.3}$$

Classical optimal stopping theory for continuous time processes, as described in Karatzas and Shreve [31, Appendix D], characterizes the solution to problem (3.3) as follows. First, by [31, Proposition D.2], $\widetilde{V}$ is a $(\mathbb{P}, \mathbb{F})$-supermartingale. Further, by [31, Proposition D.3 and Corollary D.4], there exists a càdlàg modification $\widetilde{V}^0$ of $\widetilde{V}$, called the *Snell envelope* of R, that by [31, Theorem D.7] satisfies $\widetilde{V}^0_t = \widetilde{V}_t$ almost surely, for all $t \in [0, T]$, and is the smallest càdlàg $(\mathbb{P}, \mathbb{F})$-supermartingale that dominates (in the sense of [31, Definition D.5], so $\mathbb{P}[\widetilde{V}^0_t \geq R_t \,\forall\, 0 \leq t \leq T] = 1$) the reward R. Then, by [31, Theorem D.9], a stopping time $\tau^* \in \mathcal{T}$ is optimal for problem (3.3) starting at time zero if and only if $\widetilde{V}^0_{\tau^*} = R_{\tau^*}$ almost surely, and if and only if the stopped supermartingale $(\widetilde{V}^0_{\tau^*\wedge t})_{t\in[0,T]}$ is a $(\mathbb{P}, \mathbb{F})$-martingale. Finally, under (3.2) and with a continuous reward process, [31, Theorem D.12] gives that the smallest optimal stopping time in $\mathcal{T}_{t,T}$ for the problem (3.3) is $\tau^*(t)$, the first time that the Snell envelope coincides with the reward, and so is given by

$$\tau^*(t) := \inf\{\tau \in [t, T) : \widetilde{V}^0_\tau = R_\tau\} \wedge T, \quad t \in [0, T]. \tag{3.4}$$

Given this characterization of the full information ESO problem via the Snell envelope, from now on we identify the discounted ESO value process with the Snell envelope and adopt the standard notational convention of not distinguishing between them, so $\widetilde{V} \equiv \widetilde{V}^0$. The ESO value process is then given by $V_t = \mathrm{e}^{rt}\widetilde{V}_t, t \in [0,T]$, with the understanding that $\widetilde{V}$ is the Snell envelope of the reward. With this standard convention, the optimal stopping time in (3.4) is given by the first time the ESO value process hits the payoff:

$$\tau^*(t) = \inf\{\tau \in [t,T) : V_\tau = (X_\tau - K)^+\} \wedge T, \quad t \in [0,T].$$

3.1. Full information value function. Introduce the value function $v : [0,T] \times \mathbb{R}_+ \times \{0,1\} \to \mathbb{R}_+$ for the full information optimal stopping problem (2.9) as

$$v(t,x,i) := \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(X_\tau - K)^+ \middle| X_t = x, Y_t = i\right], \quad i = 0,1, \quad t \in [0,T], \tag{3.5}$$

and write $v_i(\cdot,\cdot) \equiv v(\cdot,\cdot,i)$, $i = 0,1$. Thus, the value function in the full information scenario is a pair of functions of time and current stock price, such that $v_0(t,x)$ (respectively, $v_1(t,x)$) represents the value of the ESO to the insider at time $t \in [0,T]$ given $X_t = x$ and $Y_t = 0$ (respectively, $Y_t = 1$). In other words, the value process V in (2.9) has the representation

$$V_t = v(t,X_t,Y_t) = (1-Y_t)v_0(t,X_t) + Y_t v_1(t,X_t), \quad t \in [0,T]. \tag{3.6}$$

Very general results on optimal stopping in a continuous-time Markov setting (see, for instance, El Karoui, Lepeltier, and Millet [22]) imply that each $v_i(\cdot,\cdot)$, $i = 0,1$, is a continuous function of time and current stock price, and the process $(\mathrm{e}^{-rt}v(t,X_t,Y_t))_{t\in[0,T]}$ is the Snell envelope of the reward process R.

In what follows, we first establish, in Lemma 3.1, some elementary properties of the full information value function, so as to then characterize the nature of the continuation and stopping regions in Corollary 3.3. As we shall see, the two-drift model leads to two ordered exercise thresholds $x_i^* : [0.T] \to [K,\infty)$, $i = 0,1$, and we shall establish that these thresholds are right-continuous on $[0,T)$. Later, using the free boundary system (Proposition 3.5) and smooth pasting property (Theorem 3.6) satisfied by the value function, as well as the Doob–Meyer decomposition of the supermartingale characterizing the discounted ESO value process (Theorem 3.7), we shall obtain the limiting values $x_i^*(T-)$ of the exercise boundaries, given in Proposition 3.4, where we also show that the exercise boundaries are continuous on $[0,T)$.

With respect to $\mathbb{F}$, the dynamics of the stock are given in (2.4). For $0 \le s \le t \le T$, define the accumulation factor

$$H_{s,t} := \exp\left\{\left(\mu_0 - \frac{1}{2}\sigma^2\right)(t-s) - \sigma\eta\int_s^t Y_u \,\mathrm{d}u + \sigma(W_t - W_s)\right\}, \quad 0 \le s \le t \le T. \tag{3.7}$$

Then, given $X_s = x \in \mathbb{R}_+$, the stock price at $t \in [s,T]$ is $X_t \equiv X_t^{s,x}$, given by

$$X_t \equiv X_t^{s,x} = xH_{s,t}, \quad 0 \le s \le t \le T.$$

When $s = 0$, write $H_t \equiv H_{0,t}$ and $X_t^x \equiv X_t^{0,x}$, so that

$$X_t^x = xH_t, \quad t \in [0,T].$$

For use further below, also define the accumulation factor when the stock is exclusively in state $i \in \{0,1\}$ by

$$H^{(i)}_{s,t} := \exp\left\{\left(\mu_i - \frac{1}{2}\sigma^2\right)(t-s) + \sigma(W_t - W_s)\right\}, \quad 0 \le s \le t \le T, \quad i = 0,1, \tag{3.8}$$

and, as before, for $s = 0$ write $H^{(i)}_t \equiv H^{(i)}_{0,t}$, $i = 0,1$ for $t \in [0,T]$.

Note, in particular, that if the stock starts at time zero at $X_0 = x$, and the change point occurs in $[0,T]$, then the stock price at $t \in [\theta, T]$ (thus at or beyond the change point) is $X_t \equiv X^x_t$ given by

$$X_t = x\exp(\sigma\eta\theta)H^{(1)}_t, \quad 0 \le \theta \le t \le T. \tag{3.9}$$

With these definitions in place, the value function in (3.5) is expressed in the form

$$v_i(t,x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(xH_{t,\tau} - K)^+ \middle| Y_t = i\right], \quad (t,x) \in [0,T] \times \mathbb{R}_+, \quad i = 0,1, \tag{3.10}$$

where $H_{t,\tau}$ is the process in (3.7) over the interval $[t,\tau]$:

$$H_{t,\tau} := \exp\left\{\left(\mu_0 - \frac{1}{2}\sigma^2\right)(\tau - t) - \sigma\eta\int_t^\tau Y_u\,\mathrm{d}u + \sigma(W_\tau - W_t)\right\}, \quad \tau \in [t,T]. \tag{3.11}$$

Now, the Brownian increment $W_\tau - W_t$ in the interval $[t,\tau]$ is identical in Law to $W_{\tau-t} - W_0 = W_{\tau-t}$. Further, the integral over Y in (3.11) may be rewritten according to $\int_t^\tau Y_u\,\mathrm{d}u = \int_0^{\tau-t} Y_{t+s}\,\mathrm{d}s$, and the absence of memory property of the exponential distribution ($\mathbb{P}[\theta > t+s|\theta > t] = \mathbb{P}[\theta > s]$ for any $s,t \ge 0$) means that $\mathrm{Law}(Y_{t+s}|Y_t = i) = \mathrm{Law}(Y_s|Y_0 = i)$. Therefore, in (3.10), the integral of Y over $[t,\tau]$ with conditioning on the value of Y_t may be replaced by one over $[0,\tau-t]$ with conditioning on the value of Y_0. In other words, stationarity of Brownian increments and the memoryless property of the exponential distribution imply that optimizing over $\mathcal{T}_{t,T}$ is equivalent to optimizing over $\mathcal{T}_{0,T-t}$, so the value function in (3.10) may be recast into the form

$$v_i(t,x) = \sup_{\tau \in \mathcal{T}_{0,T-t}} \mathbb{E}\left[\mathrm{e}^{-r\tau}(xH_\tau - K)^+ \middle| Y_0 = i\right], \quad (t,x) \in [0,T] \times \mathbb{R}_+, \quad i = 0,1. \tag{3.12}$$

Thus, the ESO value with maturity T and starting time $t \in [0,T]$ is the same as the ESO value with maturity $T-t$ and initial time zero. This recasting of the ESO value will be helpful below in demonstrating some properties of the value function and is frequently utilized in American option valuation problems (see, for example, the proof of Proposition 31 in Detemple [16, Chapter 4] for the same recasting in the (simpler) case of a stock with constant drift).

The following lemma gives the elementary properties of the full information value function.

Lemma 3.1 (convexity, monotonicity, time decay: full information). *The functions $v(\cdot,\cdot,i) \equiv v_i : [0,T] \times \mathbb{R}_+, i = 0,1,$ in (3.12) or (3.5) characterizing the full information ESO value function (and the ESO value process via (3.6)) have the following properties:*

1. *For $i = 0,1$ and $t \in [0,T]$, the map $x \to v_i(t,x)$ is convex and nondecreasing.*

2. *For any fixed* $(t, x) \in [0, T] \times \mathbb{R}_+$, $v_0(t, x) \geq v_1(t, x)$.
3. *For* $i = 0, 1$ *and* $x \in \mathbb{R}_+$, *the map* $t \to v_i(t, x)$ *is nonincreasing.*

Proof.

1. Convexity and monotonicity of the map $x \to v_i(t, x)$ follow from the representation (3.12), along with convexity and monotonicity properties of the payoff function $x \to (x - K)^+$ and the linearity of the map $x \to X^x_\tau = xH_\tau$. For example, to show convexity, consider $0 \leq x_1 < x_2 < \infty$ and some $\gamma \in [0, 1]$. For each $i \in \{0, 1\}$ we then have, on using (3.12), that

$$\begin{aligned} &\gamma v_i(t, x_1) + (1 - \gamma) v_i(t, x_2) \\ &= \sup_{\tau \in \mathcal{T}_{0,T-t}} \mathbb{E}\left[\mathrm{e}^{-r\tau}\left(\gamma(x_1 H_\tau - K)^+ + (1-\gamma)(x_2 H_\tau - K)^+\right) \middle| Y_0 = i\right] \\ &\geq \sup_{\tau \in \mathcal{T}_{0,T-t}} \mathbb{E}\left[\mathrm{e}^{-r\tau}\left((\gamma x_1 + (1-\gamma)x_2)H_\tau - K\right)^+ \middle| Y_0 = i\right] \\ &= v_i(t, \gamma x_1 + (1-\gamma)x_2), \end{aligned}$$

where the inequality follows from convexity of the payoff function. This establishes convexity of $x \to v_i(t, x)$. Monotonicity is established in the same manner.

2. At maturity we have $v_0(T, x) = v_1(T, x) = (x - K)^+$ for all $x \in \mathbb{R}_+$. For $t \in [0, T)$, using the representation (3.10) and the definition (3.8) for $i = 0$, we have

$$\begin{aligned} v_0(t, x) &= \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau - t)}(xH_{t,\tau} - K)^+ \middle| Y_t = 0\right] \\ &= \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}\left(xH^{(0)}_{t,\tau}\exp\left(-\sigma\eta\int_t^\tau Y_u\,\mathrm{d}u\right) - K\right)^+ \middle| Y_t = 0\right]. \end{aligned} \tag{3.13}$$

Now, if $Y_t = 0$ (so $\theta > t$), then for any $\mathbb{F}$-stopping time $\tau \in [t, T)$ we have $\int_t^\tau Y_u\,\mathrm{d}u = (\tau - \theta)\mathbb{1}_{\{\tau \geq \theta\}} \leq \tau - t$, which implies that

$$H_{t,\tau} \equiv H^{(0)}_{t,\tau}\exp\left(-\sigma\eta\int_t^\tau Y_u\,\mathrm{d}u\right) \geq H^{(0)}_{t,\tau}\mathrm{e}^{-\sigma\eta(\tau-t)} = H^{(1)}_{t,\tau}.$$

Using this in the representation (3.13), we have

$$v_0(t, x) \geq \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(xH^{(1)}_{t,\tau} - K)^+ \middle| Y_t = 0\right].$$

But $xH^{(1)}_{t,\tau}$ is also the value of the stock at time τ given $X_t = x$ and $Y_t = 1$ (since the drift appearing in $H^{(1)}$ is μ_1), so we have

$$v_0(t, x) \geq \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(xH_{t,\tau} - K)^+ \middle| Y_t = 1\right] = v_1(t, x), \quad t \in [0, T).$$

3. This is the classical time decay property of American claims, which follows from the representation (3.12) and the fact that $\mathcal{T}_{0,T-t'} \subseteq \mathcal{T}_{0,T-t}$ for $t' \geq t$. That is, given

the time-homogeneity of the stock price model (that is, the absence of explicit time dependence in the model parameters), the possible stopping strategies starting at the later time t' are a subset of the available strategies starting at an earlier time, leading immediately to $v_i(t', x) \le v_i(t, x)$ for any fixed x and $t' \ge t$. This time decay property is well known to hold in time-homogeneous models, as discussed by Ekström [17] and Monoyios and Ng [40]. ∎

3.2. Full information continuation and stopping regions. Define the continuation regions $\mathcal{C}_i$ and stopping regions $\mathcal{S}_i$ when the one-jump process Y is in state $i \in \{0, 1\}$ by

$$\mathcal{C}_i := \{(t, x) \in [0, T) \times \mathbb{R}_+ : v_i(t, x) > (x - K)^+\}, \quad i = 0, 1,$$
$$\mathcal{S}_i := \{(t, x) \in [0, T) \times \mathbb{R}_+ : v_i(t, x) = (x - K)^+\}, \quad i = 0, 1.$$

Since the functions $v_i(\cdot, \cdot)$ are continuous, the continuation regions $\mathcal{C}_i$, $i = 0, 1$, are open sets and their respective complements $\mathcal{S}_i$, $i = 0, 1$, are closed sets. At maturity, by definition one cannot continue, so exercise takes place if the terminal stock price exceeds the strike.

Remark 3.2 (minimal conditions for early exercise: full information). If the drift process $\mu(Y)$ of the stock in (2.5) satisfies $\mu(Y) \ge r$ almost surely, then the reward process is a $(\mathbb{P}, \mathbb{F})$-submartingale, so no early exercise is optimal, and the American ESO value coincides with that of its European counterpart. In particular, if $\mu_0 \ge r$, then we expect no early exercise when $Y = 0$ (thus before the change point).

The properties in Lemma 3.1 imply that for each $i = 0, 1$, the boundary between $\mathcal{C}_i, \mathcal{S}_i$ will take the form of a nonincreasing critical stock price function (or exercise boundary) $x_i^* : [0, T) \to [K, \infty)$, with $x_0^*(t) \ge x_1^*(t) \ge K$ for all $t \in [0, T)$. The optimal exercise policy when Y is in state $i \in \{0, 1\}$ is to exercise the ESO the first time the stock price crosses $x_i^*(\cdot)$ from below, unless the change point occurs at a juncture when the exercise boundaries are strictly ordered and the stock price satisfies $x_1^*(\theta) \le X_\theta < x_0^*(\theta)$, in which case the change point causes the system to immediately switch from being in $\mathcal{C}_0$ to $\mathcal{S}_1$, and the ESO is exercised immediately after the change point. At the maturity time itself, exercise takes place if the terminal stock price exceeds the strike, so the exercise boundaries may be extended to maturity by defining $x_i^*(T) := K$, $i = 0, 1$ (though as we shall see shortly in Proposition 3.4 there exists the possibility of a discontinuity in the boundaries at maturity, with $x_i^*(T-)$ possibly not equal to K). We formalize these properties in the corollary below.

Corollary 3.3. *For $i = 0, 1$, if $\mu_i < r$, then there exist two nonincreasing right-continuous functions $x_i^* : [0, T) \to [K, \infty)$, $i = 0, 1$, satisfying*

$$x_1^*(t) \le x_0^*(t), \quad t \in [0, T), \tag{3.14}$$

such that the continuation and stopping regions in state $i \in \{0, 1\}$ are given by

$$\mathcal{C}_i = \{(t, x) \in [0, T) \times \mathbb{R}_+ : x < x_i^*(t)\}, \quad i = 0, 1, \tag{3.15}$$
$$\mathcal{S}_i = \{(t, x) \in [0, T) \times \mathbb{R}_+ : x \ge x_i^*(t)\}, \quad i = 0, 1. \tag{3.16}$$

The smallest optimal stopping time for the full information problem (2.9) *starting at time zero is* $\tau^*(0) \equiv \tau^*$*, given by*

$$\tau^* = \inf\left\{t \in [0,T) : \mathbb{1}_{\{Y_t=0\}} X_t \geq x_0^*(t) + \mathbb{1}_{\{Y_t=1\}} X_t \geq x_1^*(t)\right\} \wedge T.$$

For $i = 0, 1$*, if* $\mu_i \geq r$*, then the exercise thresholds satisfy* $x_i^*(t) = +\infty$ *for* $t \in [0,T)$*, in accordance with Remark* 3.2.

At maturity, regardless of the values of μ_i*,* $i = 0, 1$*, we have* $x_i^*(T) = K$*,* $i = 0, 1$*.*

Before giving the proof of this corollary, we state in Proposition 3.4 below some further properties of the exercise boundaries, which it is natural to give here, and which we shall prove later, after establishing free boundary PDEs and smooth pasting properties for the value functions in sections 3.3 and 3.4, along with the Doob–Meyer decomposition of the Snell envelope of the reward process (that is, the discounted full information ESO process) in section 3.5.

When $\mu_i < r$, $i = 0, 1$, so that bounded exercise thresholds exist prior to maturity, it turns out that the exercise boundaries are continuous over $[0,T)$, with a possible discontinuity at T, as we show below in Proposition 3.4. This mirrors the classical situation in the Black–Scholes model for an American call, in which the critical stock price satisfies $x^*_{\mathrm{BS}}(T-) = \max(K, (r/\delta)K)$ and $x^*_{\mathrm{BS}}(T) = K$, where δ is the dividend yield (see, for example, Detemple [16, Chapter 4, Proposition 33]). The proposition below shows that these formulae extend to the random dividend yield case, where the dividend yield can switch from its initial value to another, and where we invoke Remark 2.1 to map our problem to a classical no-arbitrage valuation of an American call. A similar remark will pertain to the partial information problem as well, where the random dividend yield will depend on a diffusion with values in $[0,1]$.

Proposition 3.4. *Suppose, for* $i = 0, 1$*, that* $\mu_i < r$*. The optimal exercise boundaries* $x_i^*(\cdot)$*,* $i = 0, 1$*, for the full information ESO problem are continuous over* $[0,T)$*, with limiting values as we approach maturity given by*

$$\lim_{t\uparrow T} x_i^*(t) \equiv x_i^*(T-) = \max\left(K, \frac{r}{r-\mu_i}K\right), \quad i = 0, 1. \tag{3.17}$$

At maturity itself, we have $x_i^*(T) = K$ *for* $i = 0, 1$*.*

The proof of this proposition will be given later in section 3.5, after we establish the free boundary PDE for the full information value function in Proposition 3.5, the smooth pasting condition in Theorem 3.6, as well as the Doob–Meyer decomposition of the Snell envelope process in Theorem 3.7, these results being utilized in the proof of Proposition 3.4.

We now turn to proving Corollary 3.3.

Proof of Corollary 3.3. For $i = 0, 1$, take $\mu_i < r$, as the case $\mu_i \geq r$ is covered by Remark 3.2. First, if early exercise has not occurred prior to maturity, then it will occur at maturity, provided the stock price is not below the strike, so we have terminal critical stock prices $x_i^*(T) = K$, $i = 0, 1$.

Next, let us show that the continuation and stopping regions have the threshold forms shown in in (3.15) and (3.16), respectively. Fix $i \in \{0, 1\}$ and $t \in [0,T)$, and suppose that $(t,x) \in [0,T) \times \mathbb{R}_+$ is such that $(t,x) \in \mathcal{S}_i$, so we have $v_i(t,x) = x - K$. Now take $\bar{x} > x$. We want to show that $(t,\bar{x}) \in \mathcal{S}_i$. Suppose, to the contrary, that $(t,\bar{x}) \notin \mathcal{S}_i$, so that $v_i(t,\bar{x}) > \bar{x} - K$. But we also have, with $\bar{\tau}$ denoting the time interval to the optimal exercise time for starting state $(t,\bar{x},i)$ in the representation (3.12), that

$$\begin{aligned}
v_i(t,\bar{x}) &= \mathbb{E}\left[\mathrm{e}^{-r\bar{\tau}}(\bar{x}H_{\bar{\tau}} - K)^+ \middle| Y_0 = i\right] \\
&= \mathbb{E}\left[\mathrm{e}^{-r\bar{\tau}}(xH_{\bar{\tau}} + (\bar{x}-x)H_{\bar{\tau}} - K)^+ \middle| Y_0 = i\right] \\
&\le \mathbb{E}\left[\mathrm{e}^{-r\bar{\tau}}(xH_{\bar{\tau}} - K)^+ \middle| Y_0 = i\right] + \mathbb{E}[\mathrm{e}^{-r\bar{\tau}}(\bar{x}-x)H_{\bar{\tau}} | Y_0 = i] \\
&\le v_i(t,x) + (\bar{x}-x)\mathbb{E}[\mathrm{e}^{-r\bar{\tau}}H_{\bar{\tau}} | Y_0 = i] \\
&< v_i(t,x) + \bar{x} - x \\
&= \bar{x} - K.
\end{aligned}$$

Above, the first inequality follows from the inequality $(a+b)^+ \le a^+ + b^+$, the second inequality follows from the suboptimality of $\bar{\tau}$ for starting state (t,x,i), and the third inequality is due to the strict supermartingale property of $(\mathrm{e}^{-rt}H_t)_{t\in[0,T]}$ when $\mu_i < r$, which we now show.

If $Y_0 = 0$, then for $t \in [0,T]$ we have, with $\mathcal{E}(\cdot)$ denoting the stochastic exponential,

$$\mathrm{e}^{-rt}H_t = \mathrm{e}^{-(r-\mu_0)t}\mathcal{E}(\sigma W)_t \exp\left(-\sigma\eta\int_0^t Y_s\,\mathrm{d}s\right) \le \mathrm{e}^{-(r-\mu_0)t}\mathcal{E}(\sigma W)_t, \quad t \in [0,T],$$

which for $\mu_0 < r$ yields a strict supermartingale. If $Y_0 = 1$, the argument is yet simpler, as in that case we obtain

$$\mathrm{e}^{-rt}H_t = \mathrm{e}^{-(r-\mu_1)t}\mathcal{E}(\sigma W)_t, \quad t \in [0,T],$$

again yielding a strict supermartingale. We thus obtain $v_i(t,\bar{x}) < \bar{x} - K$, which contradicts $v_i(t,\bar{x}) > \bar{x} - K$. Hence, $(t,\bar{x}) \in \mathcal{S}_i$, which establishes (3.15) and (3.16).

Next, let us show that the exercise boundaries are nonincreasing. Fix $i \in \{0,1\}$ and $(t,x) \in (0,T) \times \mathbb{R}_+$ such that $(t,x) \in \mathcal{C}_i$, so that $v_i(t,x) > (x-K)^+$ and $x < x_i^*(t)$. Consider a time t_0 satisfying $0 \le t_0 < t < T$. By the time decay property in Lemma 3.1 we have $v_i(t_0,x) \ge v_i(t,x)$, and therefore

$$v_i(t_0,x) - (x-K)^+ \ge v_i(t,x) - (x-K)^+ > 0,$$

so that we also have $(t_0,x) \in \mathcal{C}_i$. In other words, $x < x_i^*(t) \implies x < x_i^*(t_0)$, which can only be true if $x_i^*(\cdot)$ is nonincreasing.

Let us now show the ordering of the boundaries as expressed in (3.14). Suppose $[0,T) \times \mathbb{R}_+ \ni (t,x) \in \mathcal{C}_1$, so that $x < x_1^*(t)$ and $v_1(t,x) > (x-K)^+$. We then have, using the ordering of the value functions established in Lemma 3.1, that $v_0(t,x) \ge v_1(t,x) > (x-K)^+$, so that we also have $(t,x) \in \mathcal{C}_0$ and hence $x < x_0^*(t)$, which implies that $x_0^*(t) \ge x_1^*(t)$ over $[0,T)$.

Finally, let us show that the exercise boundaries are right-continuous over $[0,T)$. Fix $i \in \{0,1\}$ and $t \in [0,T)$, and consider a sequence $(t_n)_{n\in\mathbb{N}}$ of times converging from above to t, that is, $t_n \downarrow t$ as $n \to \infty$. Since $x_i^*(\cdot)$ is nonincreasing, we know that the right-hand limit $x_i^*(t+)$ exists. Now, for each $n \in \mathbb{N}$, $(t_n, x_i^*(t_n)) \in \mathcal{S}_i$, and because the stopping region $\mathcal{S}_i$

is a closed set, we get that $(t, x_i^*(t+)) \in \mathcal{S}_i$. Then, recalling that $\mathcal{S}_i$ has the up-connected representation (3.16), we see that we have $x_i^*(t+) \geq x_i^*(t)$. But we also have the reverse inequality $x_i^*(t+) \leq x_i^*(t)$ from the fact that $x_i^*(\cdot)$ is nonincreasing, so we obtain $x_i^*(t+) = x_i^*(t)$, showing that $x_i^*(\cdot)$ is right-continuous. ■

3.3. Full information free boundary system. Let us now proceed to the free boundary characterization of the full information value function. Define differential operators $\mathcal{L}_i$, $i = 0, 1$, acting on functions $f \in C^{1,2}([0,T] \times \mathbb{R}_+)$, by

$$\mathcal{L}_i f(t,x) := \left(\frac{\partial}{\partial t} + \mu_i x \frac{\partial}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2} - r \right) f(t,x), \quad i = 0, 1.$$

The free boundary problem for the full information value function then involves a pair of coupled PDEs as given in Proposition 3.5 below. The proof illustrates that a classical approach, akin to the proof of Theorem 2.7.7 of Karatzas and Shreve [31] in the Black–Scholes model, can be extended in our random drift scenario. This is in marked contrast to the much more involved proof of the free boundary system satisfied by finite maturity American put options in regime switching models given by Le and Wang [34, Proposition 1]. To the best of our knowledge, our result below constitutes the first time the classical method of proof is extended to a finite horizon American option model with regime switching (for example, no such regularity is established in Buffington and Elliott [8]).

Proposition 3.5 (free boundary problem: full information). *The full information value function $v(t,x,i) \equiv v_i(t,x)$, $i = 0,1$, defined in (3.5) is the unique solution in $[0,T] \times \mathbb{R}_+ \times \{0,1\}$ of the free boundary problem*

$$\mathcal{L}_0 v_0(t,x) = -\lambda\left(v_1(t,x) - v_0(t,x)\right), \quad 0 \leq x < x_0^*(t), \quad t \in [0,T), \tag{3.18}$$

$$\mathcal{L}_1 v_1(t,x) = 0, \quad 0 \leq x < x_1^*(t), \quad t \in [0,T), \tag{3.19}$$

$$v_i(t,x) = x - K, \quad x \geq x_i^*(t), \quad t \in [0,T), \quad i = 0,1, \tag{3.20}$$

$$v_i(T,x) = (x-K)^+, \quad x \in \mathbb{R}_+, \quad i = 0,1, \tag{3.21}$$

$$\lim_{x \downarrow 0} v_i(t,x) = 0, \quad t \in [0,T), \quad i = 0,1. \tag{3.22}$$

Proof. It is clear that $v_i(\cdot,\cdot)$, $i = 0,1$, satisfy the boundary conditions (3.20), (3.21), and (3.22). It remains to verify the PDEs (3.18) and (3.19). To this end, take a pair of points $(t_i, x_i) \in \mathcal{C}_i$, $i = 0,1$, and a pair of rectangles $\mathcal{R}_i := (t_i^{\min}, t_i^{\max}) \times (x_i^{\min}, x_i^{\max})$, $i = 0,1$, with $(t_i, x_i) \in \mathcal{R}_i \subset \mathcal{C}_i$, $i = 0,1$. Let $\partial\mathcal{R}_i$, $i = 0,1$, denote the boundaries of these rectangles, and denote by $\partial_0\mathcal{R}_i := \partial\mathcal{R}_i \setminus [\{t_i^{\min}\} \times (x_i^{\min}, x_i^{\max})]$ the so-called parabolic boundaries of these rectangles. With this setup, consider the terminal-boundary value problem

$$\mathcal{L}_0 f_0 = -\lambda(f_1 - f_0) \quad \text{in } \mathcal{R}_0; \quad f_0 = v_0 \quad \text{on } \partial_0\mathcal{R}_0, \tag{3.23}$$

$$\mathcal{L}_1 f_1 = 0 \quad \text{in } \mathcal{R}_1; \quad f_1 = v_1 \quad \text{on } \partial_0\mathcal{R}_1. \tag{3.24}$$

Classical theory for parabolic PDEs (for example, Friedman [24, Chapter 3]) guarantees the existence of a unique solution to (3.23)–(3.24) with all derivatives appearing in $\mathcal{L}_i$, $i = 0,1$, being continuous. We wish to show that f_i and v_i agree on $\mathcal{R}_i$, $i = 0,1$, respectively.

With $(t_i, x_i) \in \mathcal{R}_i$, $i = 0, 1$, given, define stopping times τ_i, $i = 0, 1$, by

$$\tau_i := \inf\{\rho \in [0, t_i^{\max} - t) : (t_i + \rho, x_i H_\rho) \in \partial_0 \mathcal{R}_i\} \wedge (t_i^{\max} - t), \quad i = 0, 1,$$

and processes N^i, $i = 0, 1$, by

$$N^i_\rho := \mathrm{e}^{-r\rho} f_i(t_i + \rho, x_i H_\rho), \quad 0 \le \rho \le t_i^{\max} - t, \quad i = 0, 1,$$

where $H_\rho \equiv H_{0,\rho}$ is the accumulation factor in (3.7) for the interval $[0, \rho]$. The stopped processes $(N^i_{\rho \wedge \tau_i})_{0 \le \rho \le t_i^{\max} - t_i}$, $i = 0, 1$, are $(\mathbb{P}, \mathbb{F})$-martingales by virtue of the Itô formula and the system (3.23)–(3.24) satisfied by f_i, $i = 0, 1$, and therefore

$$f_i(t_i, x_i) = N^i_{t_i} = \mathbb{E}[N^i_{\tau_i}] = \mathbb{E}[\mathrm{e}^{-r\tau_i} v_i(t_i + \tau_i, x_i H_{\tau_i})], \quad i = 0, 1, \tag{3.25}$$

where we have used the boundary conditions in (3.23)–(3.24) to obtain the last equality for each $i = 0, 1$.

But $\mathcal{R}_i \subset \mathcal{C}_i$, $i = 0, 1$, implies that $(t_i + \tau_i, x_i H_{\tau_i}) \in \mathcal{C}_i$, $i = 0, 1$, which implies that τ_i, $i = 0, 1$, must be less than or equal to the smallest optimal stopping time for starting state (t_i, x_i), $i = 0, 1$, that is,

$$\tau_i \le \tau_i^*(t_i, x_i) := \inf\{\rho \in [0, T - t_i) : v_i(t_i + \rho, x_i H_\rho) = (x_i H_\rho - K)^+\} \wedge (T - t_i), \quad i = 0, 1.$$

Now, the stopped processes

$$\mathrm{e}^{-r(\rho \wedge \tau_i^*(t_i, x_i))} v_i\left(t_i + (\rho \wedge \tau_i^*(t_i, x_i)), x_i H_{\rho \wedge \tau_i^*(t_i, x_i)}\right), \quad 0 \le \rho \le T - t_i, \quad i = 0, 1,$$

are martingales, so this and the optional sampling theorem yield that

$$\mathbb{E}\left[\mathrm{e}^{-r\tau_i} v_i(t_i + \tau_i, x_i H_{\tau_i})\right] = v_i(t_i, x_i). \tag{3.26}$$

Then (3.25) and (3.26) show that, for each $i = 0, 1$, f_i and v_i agree on $\mathcal{R}_i$ (and hence also on $\mathcal{C}_i$ since $\mathcal{R}_i \subset \mathcal{C}_i$ and $(t_i, x_i) \in \mathcal{R}_i$ were arbitrary). Thus, v_i, $i = 0, 1$, satisfy the PDEs (3.18) and (3.19).

Finally, to show uniqueness, let g_i, $i = 0, 1$, defined on the closure of $\mathcal{C}_i$, $i = 0, 1$, respectively, be solutions to the system (3.18)–(3.22). For starting states $(0, x_i, i)$, $i = 0, 1$, such that $x_i < x_i^*(0)$, $i = 0, 1$, define

$$L^i_t := \mathrm{e}^{-rt} g_i(t, x_i H_t), \quad t \in [0, T], \quad i = 0, 1,$$

as well as the smallest optimal stopping times for $v_i(0, x_i)$, $i = 0, 1$, given by

$$\begin{aligned}
\tau_0^*(x_0) &:= \inf\{t \in [0, T) : x_0 H_t^{(0)} \ge x_0^*(t)\} \wedge \inf\{t \in [0, T) : x_0 H_t^{(1)} \mathrm{e}^{\sigma\eta\theta} \ge x_1^*(t)\} \wedge T, \\
\tau_1^*(x_1) &:= \inf\{t \in [0, T) : x_1 H_t^{(1)} \ge x_1^*(t)\} \wedge T.
\end{aligned}$$

In the first equation above, the early exercise times on the right-hand side correspond to exercise before the change point (for $x_0 H_t^{(0)} \ge x_0^*(t)$) and after the change point (for $x_0 H_t^{(1)} \mathrm{e}^{\sigma\eta\theta} \ge x_1^*(t)$), where we have used the form (3.9) of the stock price after the change point.

The Itô formula yields that each $(L^i_{t\wedge\tau^*_i(x_i)})_{t\in[0,T]}$ is a martingale. Then optional sampling, along with the fact that $\tau^*_i(x_i)$, $i=0,1$, attain the respective suprema in (3.12) starting at time zero, yields that

$$\begin{aligned} g_i(0,x_i) = L^i_0 &= \mathbb{E}[L^i_{\tau^*_i(x_i)}] \\ &= \mathbb{E}\left[\mathrm{e}^{-r\tau^*_i(x_i)} g_i(\tau^*_i(x_i), x_i H_{\tau^*_i(x_i)})\right] \\ &= \mathbb{E}\left[\mathrm{e}^{-r\tau^*_i(x_i)} (x_i H_{\tau^*_i(x_i)} - K)^+\right] \\ &= v_i(0,x_i), \quad i=0,1, \end{aligned}$$

so that the solution is unique. ∎

3.4. Full information smooth fit condition. Proposition 3.5 shows that for $i=0,1$, each $v_i(\cdot,\cdot)$ is $C^{1,2}([0,T)\times\mathbb{R}_+)$ in the corresponding continuation region $\mathcal{C}_i$. In the stopping region we know that $v_i(t,x)=x-K$, which is also smooth. At issue then is the smoothness of $v_i(\cdot,\cdot)$ across the exercise boundaries $x^*_i(\cdot)$. This is settled by the smooth pasting property in Theorem 3.6 below. This property has been established for an American put in a model with multiple regime-switching by Le and Wang [34, Lemma 8], though the method of proof is complicated, relying on extending an iterative procedure first developed by Bayraktar [4], and depends upon the boundedness of the put payoff as well. Our proof is more direct, exploiting our specific one-switch model, and showing how classical techniques developed for the Black–Scholes model (see, for example, the proof of Lemma 2.7.8 in Karatzas and Shreve [31]), which proceed by analyzing properties of the smallest optimal stopping time from a given starting state, can be extended to the random drift scenario.

Theorem 3.6 (smooth pasting: full information value function). *The functions $v_i(\cdot,\cdot)$, $i=0,1$, satisfy the smooth pasting property at the optimal exercise thresholds $x^*_i(\cdot)$:*

$$\frac{\partial v_i}{\partial x}(t,x^*_i(t))=1, \quad t\in[0,T), \quad i=0,1.$$

Proof. It entails no loss of generality in this proof if we use the starting time $t=0$, so for simplicity of presentation we do so and write $v_i(x)\equiv v_i(0,x)$, $i=0,1$, $x\in\mathbb{R}_+$, and $x^*_i\equiv x^*_i(0)$ for brevity.

For $x\in\mathbb{R}_+$ and for each $i\in\{0,1\}$, the map $x\to v_i(x)$ is convex and nondecreasing, so we have $0\le v'_i(x)\le 1$ in the continuation region at time zero, $\mathcal{C}^0_i:=\{x\in\mathbb{R}_+ : x<x^*_i\}$, and thus $v'_i(x^*_i-)\le 1$. We also have $v'_i(x)=1$ in the corresponding stopping region $\mathcal{S}^0_i:=\{x\in\mathbb{R}_+ : x\ge x^*_i\}$ and thus $v'_i(x^*_i+)=1$. Hence, the proof will be complete if we can show that $v'_i(x^*_i-)\ge 1$.

First consider the case $i=1$, that is, the stock price evolution begins in the low-drift regime, so the change point happens at the initial time. The stock drift is thus equal to μ_1 throughout $[0,T]$ and the relevant value function is $v_1(\cdot)$. Denote by $\tau_1(x)$ the smallest optimal stopping time given an initial stock price $x\in\mathbb{R}_+$, given by the first time the stock breaches the boundary $x^*_1(\cdot)$,

$$\tau_1(x):=\inf\{t\in[0,T): xH^{(1)}_t\ge x^*_1(t)\}\wedge T,$$

where $H^{(1)}$ is the process in (3.8) for $s = 0$ and $i = 1$, giving the multiplicative random factor by which the stock price appreciates, so that, given $Y_0 = 1$ and $X_0 = x$, the stock price at $t \in [0, T]$ is $X_t \equiv X_t^x$, given by

$$X_t = xH_t^{(1)} = x \exp\left[\left(\mu_1 - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right], \quad t \in [0, T].$$

Set $x = x_1^* \geq K$ (the last inequality due to the fact that exercise below the strike is never optimal), fixed for the remainder of the proof for the case $i = 1$, and define

$$\tau_1(x - \epsilon) := \inf\{t \in [0, T) : (x - \epsilon)H_t^{(1)} \geq x_1^*(t)\} \wedge T$$

for $\epsilon \geq 0$, so that $\tau_1(x) \equiv 0$ and $\tau_1(x - \epsilon)$ is nondecreasing in ϵ. Because $x_1^*(\cdot)$ is nonincreasing, we have

$$\tau_1(x - \epsilon) \leq \inf\{t \in [0, T) : (x - \epsilon)H_t^{(1)} \geq x\} \wedge T. \tag{3.27}$$

The Law of the Iterated Logarithm for the Brownian motion W (Karatzas and Shreve [30, Theorem 2.9.23]) implies that $\mathbb{P}[\sup_{0 \leq t \leq a} H_t^{(1)} > 1] = 1$ for every $a > 0$, so there will exist a sufficiently small $\epsilon > 0$ such that $\sup_{0 \leq t \leq a}(x - \epsilon)H_t^{(1)} \geq x$ almost surely for every $a > 0$. Thus, the right-hand side of (3.27) tends to zero as $\epsilon \downarrow 0$, and therefore

$$\tau_1(x - \epsilon) \downarrow 0 \quad \text{as} \quad \epsilon \downarrow 0 \quad \text{almost surely.} \tag{3.28}$$

Using the fact that $\tau_1(x - \epsilon)$ will be suboptimal for the starting state $(X_0, Y_0) = (x, 1)$, we have

$$\begin{aligned} &v_1(x) - v_1(x - \epsilon) \\ &\geq \mathbb{E}\left[\mathrm{e}^{-r\tau_1(x-\epsilon)}\left((xH^{(1)}_{\tau_1(x-\epsilon)} - K)^+ - ((x - \epsilon)H^{(1)}_{\tau_1(x-\epsilon)} - K)^+\right)\right] \\ &\geq \mathbb{E}\left[\mathrm{e}^{-r\tau_1(x-\epsilon)}\left((xH^{(1)}_{\tau_1(x-\epsilon)} - K)^+ - ((x - \epsilon)H^{(1)}_{\tau_1(x-\epsilon)} - K)^+\right)\mathbb{1}_{\{(x-\epsilon)H^{(1)}_{\tau_1(x-\epsilon)} \geq K\}}\right] \\ &= \epsilon\mathbb{E}\left[\mathrm{e}^{-r\tau_1(x-\epsilon)}H^{(1)}_{\tau_1(x-\epsilon)}\mathbb{1}_{\{(x-\epsilon)H^{(1)}_{\tau_1(x-\epsilon)} \geq K\}}\right]. \end{aligned} \tag{3.29}$$

We now take the limit as $\epsilon \downarrow 0$. Using (3.28), we almost surely have $\lim_{\epsilon \downarrow 0} H^{(1)}_{\tau_1(x-\epsilon)} = 1$ and, since it is never optimal to exercise below the strike, $\lim_{\epsilon \downarrow 0} \mathbb{1}_{\{(x-\epsilon)H^{(1)}_{\tau_1(x-\epsilon)} \geq K\}} = 1$. Using these properties, along with the uniform integrability of $(H_t^{(1)})_{t \in [0,T]}$, in (3.29), we compute

$$v_1'(x-) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon}(v_1(x) - v_1(x - \epsilon)) \geq 1,$$

which completes the proof in the case $i = 1$.

Now consider the case $i = 0$, so that the stock begins at time zero in the high-drift state with drift μ_0. The early exercise scenarios bifurcate into two possibilities, either (i) before the change point or (ii) at or after the change point. Recall that, given $Y_0 = 0$ and $X_0 = x$, the stock price at $t \in [0, \theta]$ (so up to the change point) is $X_t \equiv X_t^x$, given by

$$X_t = xH_t^{(0)} = x \exp\left[\left(\mu_0 - \frac{1}{2}\sigma^2\right) + \sigma W_t\right], \quad 0 \le t \le \theta,$$

while at or after the change point the stock price is given by

$$X_t = xH_t^{(1)} \exp(\sigma\eta\theta), \quad 0 \le \theta \le t,$$

and observe that for $t = \theta$ the stock price is $xH_\theta^{(1)} \mathrm{e}^{\sigma\eta\theta} = xH_\theta^{(0)}$. The smallest optimal stopping time starting from $(0, X_0, Y_0) = (0, x, 0)$ is then $\tau_0(x)$, given by

$$\tau_0(x) := \inf\{t \in [0, T) : xH_t^{(0)} \ge x_0^*(t)\} \wedge \inf\{t \in [0, T) : xH_t^{(1)} \mathrm{e}^{\sigma\eta\theta} \ge x_1^*(t)\} \wedge T. \tag{3.30}$$

The first time on the right-hand side of (3.30) corresponds to early exercise before the change point if the stock breaches $x_0^*(\cdot)$, while the second time corresponds to early exercise at or after the change point if the stock breaches $x_1^*(\cdot)$. The latter scenario includes the possibility of early exercise at the change point itself, in which case the stock price on exercise is $xH_\theta^{(1)} \mathrm{e}^{\sigma\eta\theta} = xH_\theta^{(0)} \in [x_1^*(\theta), x_0^*(\theta))$.

As we did for the case $i = 1$, set $x = x_0^* \ge K$, fixed for the remainder of the proof, and define

$$\tau_0(x - \epsilon) := \inf\{t \in [0, T) : (x - \epsilon)H_t^{(0)} \ge x_0^*(t)\} \wedge \inf\{t \in [0, T) : (x - \epsilon)H_t^{(1)} \mathrm{e}^{\sigma\eta\theta} \ge x_1^*(t)\} \wedge T$$

for $\epsilon \ge 0$, so that $\tau_0(x) \equiv 0$ and $\tau_0(x - \epsilon)$ is nondecreasing in ϵ. Now, regardless of whether exercise occurs before the change point or not, because the exercise boundaries are nonincreasing and because $x_1^*(t) \le x_0^*(t)$ for all $t \in [0, T)$, we always have

$$\tau_0(x - \epsilon) \le \inf\{t \in [0, T) : (x - \epsilon)H_t^{(0)} \ge x\} \wedge T, \tag{3.31}$$

which is the analogue of (3.27) for the case $i = 0$. With (3.31) in place, the rest of the proof follows the same arguments as in the $i = 1$ case, so we obtain $v_0'(x-) \ge 1$, and the proof of smooth fit is complete. ∎

3.5. Doob–Meyer decomposition of full information Snell envelope.

With the free boundary PDE and smooth pasting condition established for the full information value function, we can now turn to the proof of Proposition 3.4, characterizing the continuity over $[0, T)$ and left limits $x_i^*(T-)$, $i = 0, 1$, of the exercise boundaries as we approach maturity. The key to rigorously establishing this result turns out to be the Doob–Meyer decomposition of the supermartingale that is the full information Snell envelope, in other words, the discounted full information ESO value process. This in turn leads to the decompositions below for the discounted processes $(\mathrm{e}^{-rt} v_i(t, X_t))_{t \in [0,T]}$, $i = 0, 1$, where we recall the representation (3.6) for the ESO value process V in terms of the processes $(v_i(t, X_t))_{t \in [0,T]}$, $i = 0, 1$.

Theorem 3.7 (Doob–Meyer decomposition of full information Snell envelope). *The processes* $(\mathrm{e}^{-rt}v_i(t,X_t))_{t\in[0,T]}$, $i=0,1$, *admit the decomposition*

$$\mathrm{e}^{-rt}v_i(t,X_t) = v_i(0,X_0) + M^i_t - A^i_t, \quad t\in[0,T], \quad i=0,1, \tag{3.32}$$

where

$$M^i_t := \sigma\int_0^t \mathrm{e}^{-rs}X_s\frac{\partial v_i}{\partial x}(s,X_s)\,\mathrm{d}W_s, \quad t\in[0,T], \quad i=0,1,$$

are $(\mathbb{P},\mathbb{F})$*-martingales, and*

$$A^i_t := \int_0^t \mathrm{e}^{-rs}\left((r-\mu_i)X_s - rK\right)\mathbb{1}_{\{X_s\geq x^*_i(s)\}}\,\mathrm{d}s, \quad t\in[0,T], \quad i=0,1,$$

are nondecreasing finite variation processes.

Consequently, the exercise boundaries $x^*_i(\cdot)$, $i=0,1$, *satisfy*

$$(r-\mu_i)x^*_i(t) - rK \geq 0 \quad \textit{for Lebesgue-almost every } t\in[0,T), \quad i=0,1, \tag{3.33}$$

and in particular we have the terminal left-limit lower bounds

$$x^*_i(T-) \geq \left(\frac{r}{r-\mu_i}\right)K, \quad i=0,1. \tag{3.34}$$

Proof. We have identified the full information discounted ESO value process $(\mathrm{e}^{-rt}V_t)_{t\in[0,T]}$ with the Snell envelope of the reward process, the smallest càdlàg $(\mathbb{P},\mathbb{F})$-supermartingale which dominates the reward process. We recall the representation (3.6) of the value process V in terms of the value function processes $(v_i(t,X_t))_{t\in[0,T]}$, $i=0,1$, and also recall that the process Y is equal to either 0 (before the change point) or 1 (from the change point onwards). The smooth fit condition in Theorem 3.6, along with the free boundary PDE system in Proposition 3.5, guarantees that the first partial derivatives $\partial v_i(\cdot,\cdot)/\partial x$, $i=0,1$, are continuous, even across their respective exercise boundaries $x^*_i(\cdot)$. We know also from Proposition 3.5 that the second partial derivatives $\partial^2 v_i(\cdot,\cdot)/\partial x^2$, $i=0,1$, are continuous in their respective continuation regions $\mathcal{C}_i$, $i=0,1$, and equal to zero in their respective stopping regions $\mathcal{S}_i$, $i=0,1$. Though these second derivatives might not be continuous across their respective exercise boundaries, we may nevertheless apply the generalized Itô formula for convex functions (for instance, Karatzas and Shreve [30, Theorem 3.7.1]) to the (discounted) ESO value process. In differential form, we have

$$\begin{aligned}\mathrm{d}(\mathrm{e}^{rt}V_t) = \mathrm{e}^{-rt}&\{(1-Y_t)\left(\mathrm{d}v_0(t,X_t) - rv_0(t,X_t)\,\mathrm{d}t + \lambda(v_1(t,X_t)-v_0(t,X_t))\,\mathrm{d}t\right)\\ &+ Y_t\left(\mathrm{d}v_1(t,X_t) - rv_1(t,X_t)\,\mathrm{d}t\right)\},\end{aligned}$$

where, of course, the term involving λ is due to the possibility of the change point occurring in the next instant. Then, using the generalized Itô rule on the functions $v_i(\cdot,\cdot)$ and integrating over $[0,t]$ for $t \in [0,T]$, we obtain

$$
\begin{aligned}
\mathrm{e}^{-rt}V_t - V_0 = (1-Y_t)\Bigg(&\sigma \int_0^t \mathrm{e}^{-rs} X_s \frac{\partial v_0}{\partial x}(s,X_s)\,\mathrm{d}W_s \\
&- \int_0^t \mathrm{e}^{-rs}\left((r-\mu_0)X_s - rK\right)\mathbb{1}_{\{X_s \geq x_0^*(s)\}}\,\mathrm{d}s\Bigg) \\
&+ Y_t\Bigg(\sigma \int_0^t \mathrm{e}^{-rs} X_s \frac{\partial v_1}{\partial x}(s,X_s)\,\mathrm{d}W_s \\
&- \int_0^t \mathrm{e}^{-rs}\left((r-\mu_1)X_s - rK\right)\mathbb{1}_{\{X_s \geq x_1^*(s)\}}\,\mathrm{d}s\Bigg), \quad t \in [0,T].
\end{aligned} \tag{3.35}
$$

In applying the generalized Itô rule to obtain (3.35), we have used the aforementioned properties of the functions $v_i(\cdot,\cdot)$, $i = 0,1$ (that is, the PDEs satisfied by these functions in the respective continuation regions, along with their analytic forms in the respective stopping regions), with the second derivative of a convex function considered as a measure (see, for example, Karatzas and Shreve [30, equation (3.6.47)]).

Now, in (3.35), the stochastic integral terms are $(\mathbb{P},\mathbb{F})$-martingales, since the discount factor and partial derivative terms are bounded and the stock price process is square-integrable: $\mathbb{E}[X_t^2] < \infty$ for any $t \in [0,T]$. Then, recalling once again the representation (3.6) for the value process V, we have that in both (3.6) and (3.35) above, one either has $Y_t = 0$ or $Y_t = 1$ on a mutually exclusive basis, so only one of the martingales in (3.35) contributes at any particular time. The same also applies to the finite variation terms on the right-hand side of (3.35), which is thus the (unique) Doob–Meyer decomposition of the supermartingale $(\mathrm{e}^{-rt}V_t)_{t\in[0,T]}$ into a martingale minus a nondecreasing process. This establishes the decompositions in (3.32), and also the nondecreasing property of the finite variation processes in (3.35), and thus in (3.32). Since $\mathbb{P}[X_t \geq x_i^*(t)] > 0$ for $i = 0,1$ and for Lebesgue-almost every $t \in [0,T)$, the nondecreasing property implies that the exercise boundaries must satisfy (3.33) and, in particular, (3.34) must hold. ■

We can now establish Proposition 3.4.

Proof of Proposition 3.4. It is clear that at maturity itself, exercise will not occur below the strike, so we must have $x_i^*(T) = K$, $i = 0,1$.

We have established in Corollary 3.3 that the exercise thresholds $x_i^*(\cdot)$, $i = 0,1$, are nonincreasing and right-continuous over $[0,T)$, with lower bounds $x_i^*(T-)$, $i = 0,1$, given in (3.34). With $\mu_i < r$, $i = 0,1$, we first refine this lower bound to be the right-hand side of (3.17), then we show that in fact we have the equality (3.17). For $\mu_i < r$, $i = 0,1$, we can distinguish two cases:

- for $0 \leq \mu_i < r$, we have $x_i^*(T-) \geq (r/(r-\mu_i))K \geq K$;
- for $\mu_i < 0 \leq r$, because it is never optimal to exercise below the strike, we have $x_i^*(T-) \geq K > (r/(r-\mu_i))K$.

We thus have, in all cases, the refined lower bound

$$
x_i^*(T-) \geq \max\left(K, \left(\frac{r}{r-\mu_i}\right)K\right), \quad i = 0,1.
$$

We now show that in fact we have the equality (3.17). Suppose, to the contrary, that we have $x_i^*(T-) > \max\left(K, (r/(r-\mu_i))\,K\right)$, $i = 0, 1$. For each $i = 0, 1$, consider a value $x_i \in \left(\max\left(K, (r/(r-\mu_i))\,K\right), x_i^*(T-)\right)$. Then, for $0 \le t < T$, we have $(t, x_i) \in \mathcal{C}_i$, $i = 0, 1$, so that $v_i(t, x) > (x_i - K)^+ = x_i - K$. Using temporal continuity of $v_i(\cdot,\cdot)$, we thus obtain $v_i(T, x) = \lim_{t\uparrow T} v_i(t, x) > x_i - K$. But, on the other hand, we know that at maturity we have $v_i(T, x) = (x_i - K)^+ = x_i - K$, so we have a contradiction. Thus, (3.17) holds.

Finally, let us show that the exercise thresholds $x_i^*(\cdot)$, $i = 0, 1$, are left-continuous over $[0, T)$, thus establishing the claimed continuity. To prove left-continuity we shall suppose $x_i^*(t_i-) > x_i^*(t_i)$, $i = 0, 1$, for some $t_i \in (0, T)$ and obtain a contradiction. Under this assumption, take $x_i := \frac{1}{2}(x_i^*(t_i-) + x_i^*(t_i)) > x_i^*(t_i) \ge K$, $i = 0, 1$ (of course, not the same x_i as in the previous paragraph). Observe that $(t_i, x_i) \in \mathcal{S}_i$, $i = 0, 1$, but that $(t, x_i) \in \mathcal{C}_i$, $i = 0, 1$, for $t \in (0, t_i)$. For each $i = 0, 1$, let $t \in (0, t_i)$ and $x \in (x_i, x_i^*(t))$ be given, so that (as for x_i) we have $(t_i, x) \in \mathcal{S}_i$, $i = 0, 1$, but $(t, x) \in \mathcal{C}_i$, $i = 0, 1$, for $t \in (0, t_i)$.

Now use the fact that $v_i(\cdot,\cdot)$ solves a given PDE in $\mathcal{C}_i$, as follows: for $v_0(\cdot,\cdot)$, use (3.18) along with the ordering of the value functions and time decay (properties 2 and 3 in Lemma 3.1), while for $v_1(\cdot,\cdot)$, use (3.19) and time decay to conclude that

$$\frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v_i}{\partial x^2}(t, x) \ge r v_i(t, x) - \mu_i x \frac{\partial v_i}{\partial x}(t, x), \quad (t, x) \in \mathcal{C}_i, \quad i = 0, 1.$$

Now consider separately the cases (i) $\mu_i < 0 \le r$ and (ii) $0 \le \mu_i < r$. In case (i) we have $-\mu_i x \frac{\partial v_i}{\partial x}(t, x) > 0$; using this and $v_i(t, x) > x - K$ in $\mathcal{C}_i$, we conclude that $\frac{\partial^2 v_i}{\partial x^2}(t, x) \ge \epsilon > 0$ for some $\epsilon > 0$. In case (ii), using that $x \to v_i(\cdot, x)$ is nondecreasing and convex, so that $0 \le \frac{\partial v_i}{\partial x}(t, x) \le 1$, and once again using $v_i(t, x) > x - K$, we get

$$\frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v_i}{\partial x^2}(t, x) \ge r v_i(t, x) - \mu_i x > r(x - K) - \mu_i x = (r - \mu_i)x - rK, \quad (t, x) \in \mathcal{C}_i, \quad i = 0, 1.$$

But $x > x_i^*(t_i)$ implies that (with $0 \le \mu_i < r$), $(r - \mu_i)x - rK > (r - \mu_i)x_i^*(t_i) - rK \ge 0$, on using (3.33), and so once again we conclude that $\frac{\partial^2 v_i}{\partial x^2}(t, x) \ge \epsilon > 0$ for some $\epsilon > 0$.

Thus, in either case we have

$$\frac{\partial^2 v_i}{\partial x^2}(t, x) \ge \epsilon > 0 \quad \text{for all } t \in (0, t_i), \quad x \in (x_i, x_i^*(t)), \quad i = 0, 1.$$

Then, with $\varphi(\xi) := (\xi - K)^+ = \xi - K$ (in the region of interest) and $x \in (x_i, x_i^*(t_i-))$ (so that $(t, x) \in \mathcal{C}_i$ for $t \in (0, t_i)$ but $(t_i, x) \in \mathcal{S}_i$), we compute

$$v_i(t, x) - \varphi(x) = \int_{x_i^*(t)}^{x} \int_{x_i^*(t)}^{u} \left(\frac{\partial^2 v_i}{\partial x^2}(t, \xi) - \varphi''(\xi)\right) \mathrm{d}\xi\, \mathrm{d}u \ge \frac{1}{2}\epsilon(x - x_i^*(t))^2, \quad i = 0, 1,$$

where we have used the value-matching and smooth pasting relations $v_i(t, x_i^*(t)) = \varphi(x_i^*(t))$ and $\frac{\partial v_i}{\partial x}(t, x_i^*(t)) = \varphi'(x_i^*(t))$. Finally, letting $t \uparrow t_i$ and using the continuity of $v_i(\cdot,\cdot)$, we get $v_i(t_i, x) \ge x_i - K + \frac{1}{2}\epsilon(x - x_i^*(t_i-))^2 > x_i - K$, which implies that $(t_i, x) \in \mathcal{C}_i$, $i = 0, 1$. But this contradicts our earlier assertion that $(t_i, x) \in \mathcal{S}_i$, $i = 0, 1$, and the proof is complete. ■

4. The partial information ESO problem. We now turn to the partial information problem (2.10), over $\widehat{\mathbb{F}}$-stopping times, with model dynamics given by Lemma 2.2. In particular, the stock price drift is $\mu(\widehat{Y})$, defined by

$$\mu(\widehat{Y}_t) := \mu_0 - \sigma\eta\widehat{Y}_t, \quad t \in [0,T],$$

which we see is the partial information analogue of the full information drift in (2.5).

The partial information value function $u : [0,T]\times\mathbb{R}_+\times[0,1] \to \mathbb{R}_+$ is defined by

$$u(t,x,y) := \sup_{\tau\in\widehat{\mathcal{T}}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(X_\tau - K)^+ \middle| X_t = x, \widehat{Y}_t = y\right], \quad t\in[0,T], \tag{4.1}$$

subject to the $(\mathbb{P},\widehat{\mathbb{F}})$-dynamics of the two-dimensional diffusion $(X,\widehat{Y})$ as given in (2.11) and (2.13), and the ESO value process U in (2.10) is given as

$$U_t = u(t, X_t, \widehat{Y}_t), \quad t\in[0,T].$$

For $0\le s\le t\le T$, write $(X_t,\widehat{Y}_t) \equiv (X_t^{s,x,y}, \widehat{Y}_t^{s,y})$ for the value of this diffusion given $(X_s,\widehat{Y}_s) = (x,y)$. Define

$$G_t^{s,y} := \exp\left\{\left(\mu_0 - \frac{1}{2}\sigma^2\right)(t-s) - \sigma\eta\int_s^t \widehat{Y}_u^{s,y}\,\mathrm{d}u + \sigma(\widehat{W}_t - \widehat{W}_s)\right\}, \quad 0\le s\le t\le T,$$

so we have

$$X_t^{s,x,y} = xG_t^{s,y}, \quad 0\le s\le t\le T. \tag{4.2}$$

When $s=0$, write $(X_t^{x,y},\widehat{Y}_t^y) \equiv (X_t^{0,x,y},\widehat{Y}_t^{0,y})$ and $G_t^y \equiv G_t^{0,y}$ for $t\in[0,T]$, so that

$$X_t^{x,y} = xG_t^y, \quad t\in[0,T].$$

The partial information value function in (4.1) is thus

$$u(t,x,y) = \sup_{\tau\in\widehat{\mathcal{T}}_{t,T}} \mathbb{E}\left[\mathrm{e}^{-r(\tau-t)}(xG_\tau^{t,y} - K)^+\right], \quad (t,x,y)\in[0,T]\times\mathbb{R}_+\times[0,1].$$

Using the time-homogeneity of the diffusion $(X,\widehat{Y})$, optimizing over $\widehat{\mathcal{T}}_{t,T}$ is equivalent to optimizing over $\widehat{\mathcal{T}}_{0,T-t}$, so the value function can be recast into the form

$$u(t,x,y) = \sup_{\tau\in\widehat{\mathcal{T}}_{0,T-t}} \mathbb{E}\left[\mathrm{e}^{-r\tau}(xG_\tau^y - K)^+\right]. \tag{4.3}$$

From this representation, elementary properties of the ESO partial information value function can be derived, largely in a similar manner to the proof of Lemma 3.1 in the full information case (but proving monotonicity in y is more involved, as we shall see).

Remark 4.1 (minimal conditions for early exercise: partial information). Similarly to the full information case, if the drift process $\mu(\widehat{Y})$ of the stock satisfies $\mu(\widehat{Y}) \geq r$ almost surely, then the reward process is a $(\mathbb{P}, \widehat{\mathbb{F}})$-submartingale, so no early exercise is optimal, and the American ESO value coincides with that of its European counterpart.

Lemma 4.2 (convexity, monotonicity, time decay: partial information). *The function* $u : [0,T] \times \mathbb{R}_+ \times [0,1]$ *in* (4.1) *characterizing the partial information ESO value function has the following properties:*

1. *For* $(t,y) \in [0,T] \times [0,1]$*, the map* $x \to u(t,x,y)$ *is convex and nondecreasing.*
2. *For* $(t,x) \in [0,T] \times \mathbb{R}_+$*, the map* $y \to u(t,x,y)$ *is nonincreasing.*
3. *For* $(x,y) \in \mathbb{R}_+ \times [0,1]$*, the map* $t \to u(t,x,y)$ *is nonincreasing.*

Proof. The proofs of the first and third properties are virtually identical to the proofs of the corresponding properties for the full information case in Lemma 3.1: that is, convexity and monotonicity of $x \to u(t,x,y)$ follow directly from the corresponding properties of the payoff map $x \to (x-K)^+$, while the time decay property that $t \to u(t,x,y)$ is nonincreasing follows directly from the fact that the exercise opportunities at an earlier time contain all the exercise opportunities available at a later time, given the time-homogeneity of the diffusion $(X, \widehat{Y})$. That is, in (4.1) we have $\widehat{\mathcal{T}}_{t,T} \supseteq \widehat{\mathcal{T}}_{t',T}$ for $t' \geq t$ (equivalently, in (4.3), we have $\widehat{\mathcal{T}}_{0,T-t} \supseteq \widehat{\mathcal{T}}_{0,T-t'}$).

Let us focus therefore on the second claim, that the map $y \to u(t,x,y)$ is nonincreasing. In (4.3), the quantity G^y_τ is the value at $\tau \in \widehat{\mathcal{T}}_{0,T-t}$ of the process G^y given by

$$G^y_t := \exp\left(\left(\mu_0 - \frac{1}{2}\sigma^2\right)t + \sigma\widehat{W}_t - \sigma\eta\int_0^t \widehat{Y}^y_s\,\mathrm{d}s\right), \quad t \in [0,T]. \tag{4.4}$$

From (4.4) and (4.3), the desired monotonicity of the map $y \to u(t,x,y)$ will follow if we can show that the process $\widehat{Y}^y \equiv \widehat{Y}(y)$, seen as a function of the initial value y, that is, as a *stochastic flow*, is nondecreasing with respect to y:

$$\frac{\partial \widehat{Y}_t}{\partial y}(y) \geq 0 \quad \text{a.s.,} \quad t \in [0,T]. \tag{4.5}$$

The meaning of (4.5) is that for almost all $\omega \in \Omega$, we consider the process $\widehat{Y}$ with initial value $y \in [0,1)$ as a function of y, so we have $Y_t(y) \equiv Y_t(y,\omega)$, and the theory of stochastic flows (for example, Kunita [33, Chapter 4]) guarantees that we may choose versions of $\widehat{Y}(y)$ which, for each $t \in [0,T]$ and almost all $\omega \in \Omega$, are diffeomorphisms in y from $[0,1) \to [0,1]$. In other words, the map $y \to \widehat{Y}(\omega, y)$ is smooth, and one can compute the derivative of $\widehat{Y}(\omega,y)$ with respect to y for almost all $\omega \in \Omega$. We do this in Proposition 4.3 below, to give (4.5), and this completes the proof. ∎

4.1. The filtered change point stochastic flow.

Consider the solution to the SDE (2.13) for $\widehat{Y}$ for some initial condition $\widehat{Y}_0 = y \in [0,1)$. Write $\widehat{Y}(y) = (\widehat{Y}_t(y))_{t\in[0,T]}$ for this process. Using the theory of stochastic flows (see, for instance, Kunita [33, Chapter 4]), we may choose versions of $\widehat{Y}(y)$ which, for each $t \in [0,T]$ and almost all $\omega \in \Omega$, are diffeomorphisms in y from $[0,1) \to [0,1]$. In other words, the map $y \to \widehat{Y}(y)$ is smooth. (See El Karoui, Jeanblanc-Picqué, and Shreve [21] and Monoyios and Ng [40] for other applications of these ideas to American claims and ESOs, respectively.)

We wish to show the property (4.5). To achieve this, we shall look at the flow of the so-called likelihood ratio Φ, defined for $\hat{Y} \in [0,1)$ by

$$\Phi_t := \frac{\widehat{Y}_t}{1-\widehat{Y}_t}, \quad t \in [0,T]. \tag{4.6}$$

To examine the flow of Φ, it turns out to be helpful to define the measure $\mathbb{P}^* \sim \mathbb{P}$ on $\widehat{\mathcal{F}}_T$ by

$$\Gamma_t := \left.\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}\right|_{\widehat{\mathcal{F}}_t} = \mathcal{E}(\eta\widehat{Y}\cdot\widehat{W})_t, \quad t \in [0,T], \tag{4.7}$$

where $\mathcal{E}(\cdot)$ denotes the stochastic exponential, and $(\widehat{Y}\cdot\widehat{W}) \equiv \int_0^\cdot \widehat{Y}_s\,\mathrm{d}\widehat{W}_s$ denotes the stochastic integral. Since $\widehat{Y}$ is bounded, the Novikov condition is satisfied and $\mathbb{P}^*$ is indeed a probability measure equivalent to $\mathbb{P}$.

By Girsanov's Theorem the process

$$W_t^* := \widehat{W}_t - \eta\int_0^t \widehat{Y}_s\,\mathrm{d}s, \quad t \in [0,T],$$

is a $(\mathbb{P}^*,\widehat{\mathbb{F}})$ Brownian motion. Using this along with the Itô formula, the dynamics of (X,Φ) with respect to $(\mathbb{P}^*,\widehat{\mathbb{F}})$ are given by

$$\mathrm{d}X_t = \mu_0 X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t^*, \tag{4.8}$$

$$\mathrm{d}\Phi_t = \lambda(1+\Phi_t)\,\mathrm{d}t - \eta\Phi_t\,\mathrm{d}W_t^*. \tag{4.9}$$

Equations (4.8) and (4.9) exhibit an interesting feature in that X and Φ become decoupled under $\mathbb{P}^*$. Similar measure changes have been employed by Décamps, Mariotti, and Villeneuve [14, 15], Klein [32], and Ekström and Lu [19] for related optimal stopping problems involving an investment timing decision or an optimal liquidation decision when a drift parameter is assumed to take on one of two values, but the agent is unsure which value pertains in reality. This corresponds to $\lambda \downarrow 0$ in our setup, and both X and Φ become geometric Brownian motions with respect to $(\mathbb{P}^*,\widehat{\mathbb{F}})$, yielding an easier problem in that Φ becomes a deterministic function of X. This property, when combined with the linear payoff function in these papers, allows for a reduction in dimension under some circumstances in those works. In our problem, Φ depends on the entire history of the Brownian paths, as exhibited in (4.10) below, and hence on the history of the stock price, given that we are in the observation filtration with driving Brownian motion $\widehat{W}$. This, combined with the nonlinear call payoff, makes the aforementioned dimension reduction impossible, and the numerical solution of the partial information ESO problem is made more complex.

With $\Phi_0 = \phi$, here is the result which quantifies the derivative of $\Phi(\phi)$ and hence of $\widehat{Y}(y)$ with respect to their respective initial conditions, a property which was used in the proof of Lemma 4.2.

Proposition 4.3. *Define Φ by (4.6), and define the exponential $(\mathbb{P}^*,\widehat{\mathbb{F}})$-martingale Λ by*

$$\Lambda_t := \mathcal{E}(-\eta W^*)_t, \quad t \in [0,T].$$

Let $\Phi(\phi)$ denote the solution of the SDE (4.9) with initial condition $\Phi_0 = \phi \in \mathbb{R}_+$. Then $\Phi(\phi)$ has the representation

$$\Phi_t(\phi) = \mathrm{e}^{\lambda t}\Lambda_t\left(\phi + \lambda\int_0^t \frac{\mathrm{e}^{-\lambda s}}{\Lambda_s}\,\mathrm{d}s\right), \quad t \in [0,T], \tag{4.10}$$

so that

$$\frac{\partial \Phi_t}{\partial \phi}(\phi) = \mathrm{e}^{\lambda t}\Lambda_t, \quad t \in [0,T]. \tag{4.11}$$

Consequently, if $\widehat{Y}(y)$ denotes the solution to (2.13) with initial condition $\widehat{Y}_0 = y \neq 1$, then

$$\frac{\partial \widehat{Y}_t}{\partial y}(y) = \mathrm{e}^{\lambda t}\Lambda_t\left(\frac{1-\widehat{Y}_t(y)}{1-y}\right)^2 \geq 0, \quad t \in [0,T]. \tag{4.12}$$

Proof. It is straightforward to show that $\Phi(\phi)$ as given in (4.10) solves the SDE (4.9) with initial condition $\Phi_0 = \phi$, and the formula (4.11) follows immediately. Then, using

$$\widehat{Y}_t(y) = \frac{\Phi_t(\phi)}{1+\Phi_t(\phi)}, \quad y = \frac{\phi}{1+\phi}, \quad t \in [0,T],$$

an exercise in differentiation yields (4.12). ∎

Observe that the second term on the right-hand side of (4.10) depends on the whole history of $(\Lambda_s)_{s\in[0,t]}$ over the time interval $[0,t]$, so that Φ (and hence $\widehat{Y}$) are path-dependent. As we are working in the observation filtration, these processes depend on the history of the stock price itself. This can be made explicit in some circumstances, as we show for Φ in (4.30) of section 4.5, where the integral term is written in terms of the stock price path. This path-dependence is a consequence of the filtering algorithm, and in particular that we are continuously computing an updated version at each time of the conditional expectation of a process given observations of the stock up to that time. It is not uncommon for this updating to generate path-dependence. This is the "learning" aspect of the filtering algorithm. For some special parameter values, the path-dependence can sometimes disappear. In this example, for $\lambda = 0$ we lose the history-dependent term in (4.10), reducing to the uncertain two-value drift model alluded to after (4.9).

Remark 4.4 (completing the proof of Lemma 4.2). Equation (4.12) as derived in the above proof is a $\mathbb{P}^*$-almost sure relation, and so also holds under $\mathbb{P}$ since these measures are equivalent. This is enough to complete the proof of Lemma 4.2, as claimed earlier.

4.2. Partial information free boundary problem. The properties in Lemma 4.2 imply that there exists a function $x^* : [0,T]\times[0,1] \to [K,\infty)$, the optimal exercise boundary, which is decreasing in time and also in y, such that it is optimal to exercise the ESO as soon as the stock price exceeds the threshold $x^*(t,y)$. Thus, the optimal exercise boundary in the finite horizon ESO problem under partial information is a surface, and the continuation and stopping regions $\widehat{\mathcal{C}}, \widehat{\mathcal{S}}$ for the partial information problem are given by

$$\begin{aligned}\widehat{\mathcal{C}} &:= \{(t,x,y) \in [0,T] \times \mathbb{R}_+ \times [0,1] : u(t,x,y) > (x-K)^+\} \\ &= \{(t,x,y) \in [0,T] \times \mathbb{R}_+ \times [0,1] : x < x^*(t,y)\}, \\ \widehat{\mathcal{S}} &:= \{(t,x,y) \in [0,T] \times \mathbb{R}_+ \times [0,1] : u(t,x,y) = (x-K)^+\} \\ &= \{(t,x,y) \in [0,T] \times \mathbb{R}_+ \times [0,1] : x \geq x^*(t,y)\}.\end{aligned}$$

The following lemma gives the left-limiting terminal value $x^*(T-,y)$ of the exercise surface. As in the full information case, this requires for its proof the free boundary characterization of the value function along with a smooth pasting property and also the Doob–Meyer decomposition of the (partial information) Snell envelope, so the proof of the lemma will be given in section 4.4, once the required preparation is in place.

Lemma 4.5. *The partial information exercise surface $x^*(\cdot,\cdot)$ has left-limiting value as we approach maturity, given by*

$$x^*(T-,y) = \max\left(K, \left(\frac{r}{r-(\mu_0 - \sigma\eta y)}\right)K\right), \quad y \in [0,1], \quad \mu_0 - \sigma\eta y < r. \tag{4.13}$$

Observe that, since the drift of the stock under the observation filtration is $\mu(\widehat{Y}) := \mu_0 - \sigma\eta\widehat{Y}$, the limiting value in (4.13) is

$$x^*(T-,y) = \max\left(K, \left(\frac{r}{r-\mu(y)}\right)K\right),$$

where $\mu(y)$ is the $\widehat{\mathbb{F}}$-drift of the stock when the filtered change point is equal to $y \in [0,1]$. The last condition in (4.13) therefore corresponds to the region of the state space where the filtered stock drift is less than the interest rate, and Lemma 4.5 is in a similar spirit to the full information result in Proposition 3.4, where we replace the distinct values $i = 0,1$ of the change point process by the continuum of values in $[0,1]$ for filtered change point process.

Also, by Remark 2.1, if we invoke a fictitious "dividend yield" $\delta(\cdot) := r - \mu(\cdot)$, then we have $x^*(T-y) = \max(K, (r/\delta(y))K)$, so the classical result for the exercise boundary value at $(T-)$ for no-arbitrage call valuation extends to the scenario with a random dividend yield $\delta(\widehat{Y})$, the same pattern we saw in the full information problem with random drift $\mu(Y)$.

We now turn to the free boundary characterization of the partial information value function. Let $\mathcal{L}_{X,\widehat{Y}}$ denote the generator under $\mathbb{P}$ of the two-dimensional process $(X,\widehat{Y})$ with respect to the observation filtration $\widehat{\mathbb{F}}$, with dynamics given by (2.11) and (2.13). Thus, $\mathcal{L}_{X,\widehat{Y}}$ is defined by

$$\mathcal{L}_{X,\widehat{Y}}f(t,x,y) := (\mu_0 - \sigma\eta y)xf_x + \frac{1}{2}\sigma^2x^2f_{xx} + \lambda(1-y)f_y + \frac{1}{2}\eta^2y^2(1-y)^2f_{yy} - \sigma\eta xy(1-y)f_{xy},$$

acting on any sufficiently smooth function $f : [0,T] \times \mathbb{R}_+ \times [0,1]$. Define the operator $\mathcal{L}$ by

$$\mathcal{L} := \frac{\partial}{\partial t} + \mathcal{L}_{X,\widehat{Y}} - r.$$

The partial information free boundary problem for the ESO is then as follows.

Proposition 4.6 (free boundary problem: partial information). *The partial information ESO value function $u(\cdot,\cdot,\cdot)$ defined in (4.1) is the unique solution in $[0,T]\times\mathbb{R}_+\times[0,1]$ of the free boundary problem*

$$\mathcal{L}u(t,x,y)=0,\quad 0\le x<x^*(t,y),\quad t\in[0,T),\quad y\in[0,1], \tag{4.14}$$

$$u(t,x,y)=x-K,\quad x\ge x^*(t,y),\quad t\in[0,T),\quad y\in[0,1], \tag{4.15}$$

$$u(T,x,y)=(x-K)^+,\quad x\in\mathbb{R}_+,\quad y\in[0,1], \tag{4.16}$$

$$\lim_{x\downarrow 0}u(t,x,y)=0,\quad t\in[0,T),\quad y\in[0,1]. \tag{4.17}$$

Proof. It is clear that u satisfies the boundary conditions (4.15), (4.16), and (4.17). To verify (4.14), take a point $(t,x,y)\in\widehat{\mathcal{C}}$ (so that $x<x^*(t,y)$) and a rectangular cuboid $\mathcal{R}=(t_{\min},t_{\max})\times(x_{\min},x_{\max})\times(y_{\min},y_{\max})$, with $(t,x,y)\in\mathcal{R}\subset\widehat{\mathcal{C}}$. Let $\partial\mathcal{R}$ denote the boundary of this region, and let $\partial_0\mathcal{R}:=\partial\mathcal{R}\setminus(\{t_{\min}\}\times(x_{\min},x_{\max})\times(y_{\min},y_{\max}))$ denote the so-called parabolic boundary of $\mathcal{R}$. Consider the terminal-boundary value problem

$$\mathcal{L}f=0\quad\text{in }\mathcal{R},\qquad f=u\quad\text{on }\partial_0\mathcal{R}. \tag{4.18}$$

Classical theory for parabolic PDEs (for instance, Friedman [24, Chapter 3]) guarantees the existence of a unique solution to (4.18), with all derivatives appearing in $\mathcal{L}$ being continuous. We wish to show that f and u agree on $\mathcal{R}$.

With $(t,x,y)\in\mathcal{R}$ given, define the stopping time $\tau\in\widehat{\mathcal{T}}_{0,t_{\max}-t}$ by

$$\tau:=\inf\{\rho\in[0,t_{\max}-t):(t+\rho,xG^y_\rho,\widehat{Y}^y_\rho)\in\partial_0\mathcal{R}\}\wedge(t_{\max}-t),$$

where the process G^y is defined in (4.4), and define the process N by

$$N_\rho:=\mathrm{e}^{-r\rho}f(t+\rho,xG^y_\rho,\widehat{Y}^y_\rho),\quad 0\le\rho\le t_{\max}-t.$$

The stopped process $(N_{\rho\wedge\tau})_{0\le\rho\le t_{\max}-t}$ is a $(\mathbb{P},\widehat{\mathbb{F}})$-martingale by virtue of the Itô formula and the system (4.18) satisfied by f, and therefore

$$f(t,x,y)=N_t=\mathbb{E}[N_\tau]=\mathbb{E}[\mathrm{e}^{-r\tau}u(t+\tau,xG^y_\tau,\widehat{Y}^y_\tau)], \tag{4.19}$$

where we have used the boundary condition in (4.18) to obtain the last equality.

Since $\mathcal{R}\subset\widehat{\mathcal{C}}$, $(t+\tau,xG^y_\tau,\widehat{Y}^y_\tau)\in\widehat{\mathcal{C}}$, so τ must satisfy

$$\tau\le\tau^*(t,x,y):=\inf\{\rho\in[0,T-t):u(t+\rho,xG^y_\rho,\widehat{Y}^y_\rho)=(xG^y_\rho-K)^+\}\wedge(T-t).$$

In other words, τ must be less than or equal to the smallest optimal stopping time $\tau^*(t,x,y)$ for the starting state (t,x,y). Now, the stopped process

$$\mathrm{e}^{-r(\rho\wedge\tau^*(t,x,y))}u\left(t+(\rho\wedge\tau^*(t,x,y)),xG^y_{\rho\wedge\tau^*(t,x,y)},\widehat{Y}^y_{\rho\wedge\tau^*(t,x,y)}\right),\quad 0\le\rho\le T-t,$$

is a martingale, so this and the optional sampling theorem yield that

$$\mathbb{E}\left[\mathrm{e}^{-r\tau}u(t+\tau,xG^y_\tau,\widehat{Y}^y_\tau)\right]=u(t,x,y). \tag{4.20}$$

Then (4.19) and (4.20) show that f and u agree on $\mathcal{R}$ (and hence also on $\widehat{\mathcal{C}}$ since $\mathcal{R} \subset \widehat{\mathcal{C}}$ and $(t, x, y) \in \mathcal{R}$ were arbitrary). Thus, u satisfies (4.14).

Finally, to show uniqueness, let g defined on the closure of $\widehat{\mathcal{C}}$ be a solution to the system (4.14)–(4.17). For starting state $(0, x, y)$ such that $x < x^*(0, y)$ define

$$L_t := \mathrm{e}^{-rt} g(t, xG_t^y, \widehat{Y}_t^y), \quad t \in [0, T],$$

as well as the optimal stopping time for $u(0, x, y)$, given by

$$\tau^*(x, y) := \inf\{t \in [0, T) : xG_t^y \geq x^*(t, \widehat{Y}_t^y)\} \wedge T.$$

The Itô formula yields that $(L_{t\wedge\tau^*(x,y)})_{t\in[0,T]}$ is a martingale. Then optional sampling, along with the fact that $\tau^*(x, y)$ attains the supremum in (4.3) starting at time zero, yields that

$$\begin{aligned} g(0, x, y) = L_0 &= \mathbb{E}[L_{\tau^*(x,y)}] \\ &= \mathbb{E}\left[\mathrm{e}^{-r\tau^*(x,y)} g(\tau^*(x, y), xG^y_{\tau^*(x,y)}, \widehat{Y}^y_{\tau^*(x,y)})\right] \\ &= \mathbb{E}\left[\mathrm{e}^{-r\tau^*(x,y)} (xG^y_{\tau^*(x,y)} - K)^+\right] \\ &= u(0, x, y), \end{aligned}$$

so that the solution is unique. ■

4.3. Partial information smooth fit condition. We establish, in Theorem 4.7 below, a smooth pasting property for the partial information value function. This is a natural property one might expect to hold but to the best of our knowledge has not been established before in a diffusion model such as our partial information model. In stochastic volatility models, Touzi [46] has used variational inequality techniques to show the smooth pasting property. It may be that this method could be adapted to our setting.

We shall employ a method more akin to the classical proof of smooth fit in American option problems, in a similar spirit to Karatzas and Shreve [31, Lemma 2.7.8] (for the case of the Black–Scholes put) or Monoyios and Ng [40, Theorem 3.4] (in a model with inside information). The proof of Theorem 4.7 is simplified by using the measure $\mathbb{P}^* \sim \mathbb{P}$ defined in (4.7). Because the proof involves analyzing the first time the stock almost surely breaches a surface, and as we are working in the observation filtration, any early exercise crossing point must ultimately depend only on the stock price path, so moving to a measure where X has constant drift (equal to μ_0 under $\mathbb{P}^*$, recall the SDE (4.8)) simplifies matters.

Put explicitly, any optimal early exercise time will be the first time $t \in [0, T)$ that we have $X_t \geq x^*(t, \widehat{Y}_t)$. In this relation, the process $\widehat{Y}$ depends on the history of the stock price, through the history-dependence of the process $\Phi \equiv \widehat{Y}/(1 - \widehat{Y})$ in (4.10) (see also (4.30) in section 4.5, where we make explicit the dependence of Φ on the history of the stock price), so the early exercise crossing point is indeed dependent only on the stock price (albeit in a path-dependent manner), and this makes our method of proof work. This in turn can ultimately be traced to the fact that, under the observation filtration, both the stock X and the filtered change point process $\widehat{Y}$ are driven by the *same* one-dimensional Brownian motion. Put yet another way, the full information incomplete model with an observed but unhedgeable change point has been rendered into a complete model with two diffusion processes driven by one Brownian motion. This is a not uncommon feature in filtering models. The price one pays for this induced market completeness is that the second factor $\widehat{Y}$ depends on the entire history of the stock price, also a not uncommon feature of models with filtering—this is the "learning" aspect of filtering coming to the fore.

Theorem 4.7 (smooth pasting: partial information value function). *The partial information value function defined in* (4.1) *satisfies the smooth pasting property*

$$\frac{\partial u}{\partial x}(t, x^*(t,y), y) = 1, \quad t \in [0,T), \quad y \in [0,1],$$

at the optimal exercise threshold $x^*(t,y)$*.*

Proof. In this proof it entails no loss of generality if we set $r = 0$ and $t = 0$, but this considerably simplifies notation, so let us proceed in this way. Write $u(x,y) \equiv u(0,x,y)$ and $x^*(y) \equiv x^*(0,y)$ for brevity.

The map $x \to u(x,y)$ is convex and nondecreasing, so we have $u_x(x,y) \leq 1$ in the continuation region $\widehat{\mathcal{C}} = \{(x,y) \in \mathbb{R}_+ \times [0,1] : x < x^*(y)\}$, and thus $u_x(x^*(y)-, y) \leq 1$. We also have $u_x(x,y) = 1$ in the stopping region $\widehat{\mathcal{S}} = \{(x,y) \in \mathbb{R}_+ \times [0,1] : x \geq x^*(y)\}$, and thus $u_x(x^*(y)+, y) = 1$. Hence, the proof will be complete if we can show that $u_x(x^*(y)-, y) \geq 1$. Recall the measure $\mathbb{P}^*$ defined in (4.7) and the $(\mathbb{P}^*, \widehat{\mathbb{F}})$-dynamics of the stock in (4.8). Given $X_0 = x$, the stock price at time $t \in [0,T]$ is

$$X_t = xG_t := x \exp\left(\left(\mu_0 - \frac{1}{2}\sigma^2\right)t + \sigma W_t^*\right), \quad t \in [0,T].$$

For $(x,y) \in \mathbb{R}_+ \times (0,1)$, denote by $\tau(x,y)$ the optimal $\widehat{\mathbb{F}}$-stopping time for $u(x,y)$, given by the first time the stock breaches the exercise surface at the prevailing value of $\widehat{Y}$. Working under $\mathbb{P}^*$, we thus have

$$\tau(x,y) = \inf\{t \in [0,T) : xG_t \geq x^*(t, \widehat{Y}_t^y)\} \wedge T,$$

where $\widehat{Y}^y$ denotes the filtered change point process with initial condition $Y_0 = y$.

Set $x = x^*(y) \geq K$, which will be fixed for the remainder of the proof, and define

$$\tau(x-\epsilon, y) := \inf\{t \in [0,T) : (x-\epsilon)G_t \geq x\} \wedge T$$

for $\epsilon \geq 0$, and the dependence on y on the right-hand side is of course suppressed in $x \equiv x^*(y)$. We have $\tau(x,y) \equiv 0$ and that $\tau(x-\epsilon,y)$ is nondecreasing in ϵ. Moreover, because the exercise surface is nonincreasing in time and in y, we have

$$\tau(x-\epsilon, y) \leq \inf\{t \in [0,T) : (x-\epsilon)G_t \geq x\} \wedge T. \tag{4.21}$$

The Law of the Iterated Logarithm for the Brownian motion W^* (Karatzas and Shreve [30, Theorem 2.9.23]) implies that

$$\sup_{0 \leq t \leq a} G_t > 1, \quad \mathbb{P}^*\text{-almost surely}$$

for every $a > 0$, so there will exist a sufficiently small $\epsilon > 0$ such that

$$\sup_{0\le t\le a} (x-\epsilon)G_t^y \ge x, \quad \mathbb{P}^*\text{-almost surely}$$

for every $a > 0$. Thus, the right-hand side of (4.21) tends to zero as $\epsilon \downarrow 0$, and therefore $\tau(x-\epsilon, y) \downarrow 0$ as $\epsilon \downarrow 0$, $\mathbb{P}^*$-almost surely and, since $\mathbb{P}^* \sim \mathbb{P}$, this is also true $\mathbb{P}$-almost surely:

$$\tau(x-\epsilon, y) \downarrow 0 \quad \text{as} \quad \epsilon \downarrow 0, \quad \mathbb{P}\text{-almost surely.} \tag{4.22}$$

Using the fact that $\tau(x-\epsilon, y)$ will be suboptimal for the starting state $(X_0, \widehat{Y}_0) = (x, y)$, we have

$$\begin{aligned} &u(x,y) - u(x-\epsilon, y) \\ &\ge \mathbb{E}\left[\left((xG_{\tau(x-\epsilon,y)} - K)^+ - ((x-\epsilon)G_{\tau(x-\epsilon,y)} - K)^+\right)\right] \\ &\ge \mathbb{E}\left[\left((xG_{\tau(x-\epsilon,y)} - K)^+ - ((x-\epsilon)G_{\tau(x-\epsilon,y)} - K)^+\right)\mathbb{1}_{\{(x-\epsilon)G_{\tau(x-\epsilon,y)}\ge K\}}\right] \\ &= \epsilon\mathbb{E}\left[G_{\tau(x-\epsilon,y)}\mathbb{1}_{\{(x-\epsilon)G_{\tau(x-\epsilon,y)}\ge K\}}\right]. \end{aligned} \tag{4.23}$$

We now take the limit as $\epsilon \downarrow 0$. Using (4.22) we almost surely have $\lim_{\epsilon\downarrow 0} G_{\tau(x-\epsilon,y)} = 1$ and, since it is never optimal to exercise below the strike, $\lim_{\epsilon\downarrow 0} \mathbb{1}_{\{(x-\epsilon)G_{\tau(x-\epsilon,y)}\ge K\}} = 1$. Using these properties, along with the uniform integrability of $(G_t)_{t\in[0,T]}$, in (4.23), we compute

$$u_x(x-, y) = \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}(u(x,y) - u(x-\epsilon, y)) \ge 1,$$

which completes the proof. ■

4.4. Doob–Meyer decomposition of partial information Snell envelope. As was done in the full information case, with the free boundary PDE and smooth pasting condition established for the partial information value function, we can now derive a Doob–Meyer decomposition for the partial information Snell envelope of the reward process, and this allows us to prove Lemma 4.5 on the left-limiting value of the partial information exercise surface as we approach maturity.

Recall that the partial information Snell envelope is the càdlàg supermartingale identified with the discounted ESO value process $(\mathrm{e}^{-rt}U_t)_{t\in[0,T]}$, with $U_t = u(t, X_t, \widehat{Y}_t)$.

Lemma 4.8 (Doob–Meyer decomposition of partial information Snell envelope). *The process* $(\mathrm{e}^{-rt}u(t, X_t, \widehat{Y}_t))_{t\in[0,T]}$ *admits the decomposition*

$$\mathrm{e}^{-rt}u(t, X_t, \widehat{Y}_t) = u(0, X_0, \widehat{Y}_0) + M_t - A_t, \quad t \in [0,T], \tag{4.24}$$

where

$$M_t := \int_0^t \mathrm{e}^{-rs}\left(\sigma X_s u_x(s, X_s, \widehat{Y}_s) - \eta\widehat{Y}_s(1-\widehat{Y}_s)u_y(s, X_s, \widehat{Y}_s)\right)\mathrm{d}\widehat{W}_s, \quad t \in [0,T],$$

is a $(\mathbb{P}, \widehat{\mathbb{F}})$*-martingale, and*

$$A_t := \int_0^t \mathrm{e}^{-rs} \left((r - \mu_0 + \sigma\eta\widehat{Y}_s)X_s - rK \right) \mathbb{1}_{\{X_s \geq x^*(s,\widehat{Y}_s)\}} \, \mathrm{d}s, \quad t \in [0,T],$$

is a nondecreasing finite variation process.

Proof. The proof is similar to the corresponding proof of Theorem 3.7 in the full information scenario, so we shall be more brief here. Using the generalized Itô formula for convex functions, the PDE (4.14) satisfied by $u(\cdot,\cdot,\cdot)$ in the continuation region $\widehat{\mathcal{C}}$, and the fact that $u(t,x,y) = x - K$ in the stopping region, we obtain the decomposition (4.24). The square-integrability of the stock price and bounded nature of the derivatives u_x, u_y in M imply that M is indeed a martingale. Since the Snell envelope is a supermartingale with a unique Doob–Meyer decomposition into a martingale minus a nondecreasing process of finite variation, we conclude that A is a nondecreasing process. ∎

Some observations on the parameter values for which we obtain a bounded exercise surface are in order. With $\mu(\widehat{Y}) \equiv \mu_0 - \sigma\eta\widehat{Y}$ the partial information stock price drift, the nondecreasing property of the process A in Lemma 4.8 means that we have $((r - \mu(\widehat{Y}_t))X_t - rK)\mathbb{1}_{\{X_t \geq x^*(t,\widehat{Y}_t)\}} \geq 0$ almost surely for all $t \in [0,T]$, and hence we also have $(r - \mu(\widehat{Y}_t))x^*(t,\widehat{Y}_t) - rK \geq 0$. Now, suppose we have $\mu(\widehat{Y}_t) \geq r$ almost surely for all $t \in [0,T]$. We then compute that $x^*(t,\widehat{Y}_t) \leq -(r/(\mu(\widehat{Y}_t) - r))K$, which is impossible, since the exercise surface cannot lie below the strike. We conclude that when the stock drift exceeds the interest rate, the finite variation process in the Doob–Meyer decomposition will be zero, and the ESO value process is a martingale. This is of course exactly in line with Remark 4.1, that early exercise will not occur if the stock drift dominates the interest rate, in which case the ESO value process is a martingale and equal to the European version of the ESO.

We are now ready to prove Lemma 4.5.

Proof of Lemma 4.5. From the nondecreasing property of the process A in Lemma 4.8 we have $((r - \mu(\widehat{Y}_t))X_t - rK)\mathbb{1}_{\{X_t \geq x^*(t,\widehat{Y}_t)\}} \geq 0$ almost surely for all $t \in [0,T]$, and hence we also have $(r - \mu(\widehat{Y}_t))x^*(t,\widehat{Y}_t) - rK \geq 0$.

Suppose that $\mu(\widehat{Y}_t) < r$. In this case, we conclude that $x^*(t,\widehat{Y}_t) \geq (r/(r - \mu(\widehat{Y}_t)))K$. From the fact that the exercise surface is nonincreasing in time, we conclude that we have the terminal left-limit lower bound

$$x_i^*(T-,y) \geq \left(\frac{r}{r - \mu_0 + \sigma\eta y} \right) K$$

for all values of $y \in [0,1]$ satisfying $\mu_0 - \sigma\eta y < r$. There are now two cases to consider separately, which lead to a refinement of this lower bound:

- for $0 \leq \mu_0 - \sigma\eta y < r$, we obtain $x^*(T-,y) \geq (r/(r - \mu_0 + \sigma\eta y))\, K \geq K$;
- for $\mu_0 - \sigma\eta y \leq 0 < r$, because it is never optimal to exercise below the strike, we have $x^*(T-,y) \geq K > (r/(r - \mu_0 + \sigma\eta y))K$.

We thus have, in all cases, the refined lower bound

$$x^*(T-,y) \geq \max\left(K, \left(\frac{r}{r-\mu_0+\sigma\eta y}\right)K\right), \quad \mu_0 - \sigma\eta y < r.$$

We now show that in fact we have equality here, thus establishing (4.13). Suppose, to the contrary, that we have $x^*(T-,y) > \max\left(K, (r/(r-\mu_0+\sigma\eta y))K\right)$. Fixing $y \in [0,1]$, consider a value $x \in (\max\left(K, (r/(r-\mu_0+\sigma\eta y))K\right), x^*(T-,y))$. Then, for $0 \leq t < T$, we have $(t,x,y) \in \widehat{\mathcal{C}}$, so that $u(t,x,y) > (x-K)^+ = x-K$. Using temporal continuity of $u(\cdot,\cdot,\cdot)$, we thus obtain $u(T,x,y) = \lim_{t\uparrow T} u(t,x,y) > x-K$. But, on the other hand, we know that at maturity we have $u(T,x,y) = (x-K)^+ = x-K$, so we have a contradiction. Thus, (4.13) holds. ■

4.5. A comment on a change of state variable. In this section, we illustrate the inherent complexity of the partial information case, due to its path-dependent structure. Consider the partial information problem (2.10). We shall change measure to $\mathbb{P}^*$ defined in (4.7), and this naturally leads to a change of state variable from $(X, \widehat{Y})$ to (X, Φ), with Φ defined in (4.6). This leads to the following lemma.

Lemma 4.9. *Let Φ be the likelihood ratio process defined in (4.6). The partial information ESO value process U in (2.10) satisfies*

$$\mathrm{e}^{-(r+\lambda)t}(1+\Phi_t)U_t = \operatorname*{ess\,sup}_{\tau\in\widehat{\mathcal{T}}_{t,T}} \mathbb{E}^*\left[\mathrm{e}^{-(r+\lambda)\tau}(1+\Phi_\tau)(X_\tau-K)^+|\widehat{\mathcal{F}}_t\right], \quad t\in[0,T], \tag{4.25}$$

where $\mathbb{E}^*[\cdot]$ denotes expectation with respect to $\mathbb{P}^*$ in (4.7), and the $(\mathbb{P}^*, \widehat{\mathbb{F}})$-dynamics of X, Φ are given in (4.8) and (4.9).

Proof. Let Z denote the change of measure martingale defined by

$$Z_t := \frac{1}{\Gamma_t} = \left.\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^*}\right|_{\widehat{\mathcal{F}}_t} = \mathcal{E}(-\eta\widehat{Y}\cdot W^*)_t, \quad t\in[0,T], \tag{4.26}$$

satisfying

$$\mathrm{d}Z_t = -\eta\widehat{Y}_t Z_t\,\mathrm{d}W_t^*, \quad Z_0 = 1. \tag{4.27}$$

The Itô formula along with the dynamics of Φ in (4.9) yields that Z is given in terms of Φ as

$$Z_t = \mathrm{e}^{-\lambda t}\left(\frac{1+\Phi_t}{1+\Phi_0}\right), \quad t\in[0,T], \tag{4.28}$$

because the right-hand side of (4.28) satisfies the SDE (4.27). Then an application of the Bayes formula to the definition of U in (2.10) yields the result. ■

The point of (4.25) is that the state variables in the objective function have decoupled dynamics under $\mathbb{P}^*$ (recall (4.8) and (4.9)). However, the problematic feature of the history dependence of Φ remains, as exhibited in (4.10), inheriting this feature from the filtered change-point process $\widehat{Y}$. Indeed, using the solution of the stock price SDE (4.8), the representation (4.10) may be converted to one involving the stock price and its history, as follows.

With $X_0 = x$, from (4.8) we have $X_t = x\exp\left(\left(\mu_0 - \frac{1}{2}\sigma^2\right)t + \sigma W_t^*\right)$, $t \geq 0$, so that

$$\exp(\sigma W_t^*) = \left(\frac{X_t}{x}\right)\exp\left(\mu_0 - \frac{1}{2}\sigma^2\right)t, \quad t \geq 0.$$

Using this relation to compute the process $\Lambda = \mathcal{E}(-\eta W^*)$ we get

$$\Lambda_t = \exp\left(-\eta W_t^* - \frac{1}{2}\eta^2 t\right) = \left(\frac{X_t}{x}\right)^{-\eta/\sigma}\exp\left(\eta\nu_0 - \frac{1}{2}\eta^2\right)t, \quad t \geq 0, \tag{4.29}$$

where

$$\nu_0 := \frac{\mu_0}{\sigma} - \frac{1}{2}\sigma.$$

Then, with $\Phi_0 = \phi$, substituting (4.29) into (4.10), we obtain

$$\Phi_t(\phi) = \phi\mathrm{e}^{\kappa t}\left(\frac{X_t}{x}\right)^{-\eta/\sigma} + \lambda\int_0^t \mathrm{e}^{\kappa(t-s)}\left(\frac{X_t}{X_s}\right)^{-\eta/\sigma}\mathrm{d}s, \quad t \in [0,T], \tag{4.30}$$

where κ is a constant given by

$$\kappa := \lambda + \eta\nu_0 - \frac{1}{2}\eta^2.$$

The second term on the right-hand side of (4.30) is the awkward history-dependent term which makes numerical solution of the partial information ESO problem difficult. For $\lambda = 0$, we see that Φ becomes a deterministic function of the current stock price, and this limit corresponds to a simpler model in which an unknown drift is assumed to take one of two values, but the agent is unsure which value pertains in reality, and so filtering is used to estimate the drift. A number of papers have used such a model and exploited the absence of path-dependence to reduce the dimension of the problem (see Décamps, Mariotti, and Villeneuve [14, 15], Klein [32], and Ekström and coauthors [19, 18, 20]). This simplification is not available to us, so the partial information problem is potentially more challenging to solve numerically.

5. On the effect of a vesting period on ESO exercise. ESOs often include a contractual feature called a vesting period, a period of time during which option exercise is not permitted. In this section, we briefly describe the effect of a vesting period on the exercise of ESOs in the full and partial information models. In section 7 we shall also demonstrate the impact of vesting on the ESO value.

Suppose there is a vesting period $[0, t_v)$, so that the ESO can only be exercised in the time interval $[t_v, T]$. Then we seek optimal stopping times, with respect to the appropriate filtration, lying in the exercise interval $[t_v, T]$. Thus, for $0 \leq t < t_v$, the discounted full information ESO value process is

$$\mathrm{e}^{-rt}\check{V}_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t_v,T}}\mathbb{E}[R_\tau|\mathcal{F}_t], \quad t \in [0, t_v), \tag{5.1}$$

while for $t \in [t_v, T]$, the vesting period is over, and we have reverted back to our original problem without a vesting period with value process $(V_t)_{t\in[t_v,T]}$, given by

$$\mathrm{e}^{-rt}V_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}[R_\tau|\mathcal{F}_t], \quad t \in [t_v, T]. \tag{5.2}$$

Note that $\check{V}_t \leq V_t$ for $t < t_v$ (the value with vesting is clearly dominated by the one without vesting, due to the extra exercise opportunities).

Similarly, for $t \in [0, t_v)$, the discounted partial information value process is

$$\mathrm{e}^{-rt}\check{U}_t = \operatorname*{ess\,sup}_{\tau\in\widehat{\mathcal{T}}_{t_v,T}} \mathbb{E}[R_\tau|\widehat{\mathcal{F}}_t], \quad t \in [0, t_v),$$

satisfying $\check{U}_t \leq U_t$ for $t < t_v$, while for $t \in [t_v, T]$ we are back to our original problem without a vesting period:

$$\mathrm{e}^{-rt}U_t = \operatorname*{ess\,sup}_{\tau\in\widehat{\mathcal{T}}_{t,T}} \mathbb{E}[R_\tau|\widehat{\mathcal{F}}_t], \quad t \in [t_v, T].$$

The key overall idea is well expressed by Leung and Sircar [36, section 5.1.1], as follows: "When a vesting period of t_v years is imposed, the employee cannot exercise the ESO during $[0, t_v)$, but the post-vesting exercising strategy will be unaffected."

In what follows, we examine the situation where we have $y_0 = 0$, $\mu_0 > r$, $\mu_1 < r$. Thus, for the full information problem, no exercise will occur before the strictly positive change point $\theta \sim \mathrm{Exp}(\lambda)$, as the reward process $(R_t)_{0,\theta}$ over the time interval up to the change point is a submartingale.

5.1. The full information case. First, consider the case that the change point occurs after the vesting period has elapsed, that is, $\theta \geq t_v$. For $t < t_v$, no exercise can occur, and at $t = t_v$ we revert back to our original problem, the vesting period having elapsed. The postvesting exercise strategy will then be as in the no-vesting case.

Next, consider the case $\theta < t_v$, that is, the change point occurs during the vesting period. For $t \in [0, t_v)$ there is no exercise as we are still in the vesting period. At $t = t_v$, we are now in the low-drift state, so the stock is a geometric Brownian motion with drift $\mu_1 < r$. There will now be an exercise boundary $(x_1^*(t))_{t_v\leq t\leq T}$. If $X_{t_v} \geq x_1^*(t_v)$, then we are in the exercise region as soon as the vesting period has elapsed, and immediate exercise occurs at $t = t_v$. If, on the other hand, $X_{t_v} < x_1^*(t_v)$, then there is no immediate exercise at t_v, and exercise occurs the first time that the stock breaches the boundary from below, at time $\bar{\tau} = \inf\{t \in [t_v, T) : X_t \geq x_1^*(t)\} \wedge T$.

Thus, the overall conclusion is that the exercise boundary is infinite over $[0, t_v)$, regardless of when the change point occurs. If the change point has occurred by time t_v, then immediate exercise occurs at time t_v if the prevailing stock price at t_v is higher than or equal to the exercise boundary $x_1^*(t_v)$ at that point. If the change point has not occurred by time t_v, we are back to our original problem over the interval $[t_v, T]$.

5.2. The partial information case. Regardless of when the change point occurs, if we are in the vesting period $[0, t_v)$, no exercise can occur, so the partially informed agent's exercise surface is infinite.

At $t = t_v$ we revert back to our original problem, the vesting period having elapsed. Again, this is regardless of whether the change point has occurred or not (the partially informed agent is not aware of the change point having occurred or not and is therefore filtering it from stock price observations). We now have an optimal exercise surface $(x^*(t, y))_{t_v \le t \le T, 0 \le y \le 1}$, and exercise occurs the first time that the stock breaches the exercise surface evaluated at the prevailing value of $\widehat{Y}$, that is, at $\tau^* = \inf\{t \in [t_v, T) : X_t \ge x^*(t, \widehat{Y}_t)\} \wedge T$.

In other words, the postvesting exercise strategy will then be as in the no-vesting case, with the prevesting boundary set to infinity.

6. Numerical scheme and convergence tests.

In this section, we describe numerical schemes for the PDEs in the full and partial information case and present numerical studies to illustrate the convergence and computational complexity. We present our novel algorithm for the two-dimensional, degenerate free boundary value problem in the partial information case in some detail and analyze its convergence properties, while we only state the simple scheme for the full information case. Note that alternative numerical methods could be employed, for example, a binomial scheme (nonrecombining for the partial information case) or a Longstaff–Schwartz Monte Carlo approach. However, the finite difference schemes we propose are far superior in terms of speed and accuracy.

6.1. The partial information case.

We begin by noting that the partial information ESO value function $u(\cdot, \cdot, \cdot)$ satisfying (4.14)–(4.17) is also the unique solution in $[0, T] \times \mathbb{R}_+ \times [0, 1]$ of the equivalent linear complementarity problem

$$\min\left(-\mathcal{L}u(t, x, y), u - (x - K)^+\right) = 0, \ t \in [0, T), \quad x \in \mathbb{R}_+, \ y \in [0, 1], \tag{6.1}$$

$$u(T, x, y) = (x - K)^+, \quad x \in \mathbb{R}_+, \quad y \in [0, 1], \tag{6.2}$$

where we repeat for convenience that

$$\mathcal{L} = \frac{\partial}{\partial t} + \mathcal{L}_{X, \widehat{Y}} - r \tag{6.3}$$

with

$$\mathcal{L}_{X,\widehat{Y}} f(t, x, y) = (\mu_0 - \sigma\eta y) x f_x + \frac{1}{2}\sigma^2 x^2 f_{xx} + \lambda(1 - y) f_y + \frac{1}{2}\eta^2 y^2 (1 - y)^2 f_{yy} - \sigma\eta x y(1 - y) f_{xy}$$

for any sufficiently smooth function $f : [0, T] \times \mathbb{R}_+ \times [0, 1]$.

The degeneracy of the equation requires the notion of viscosity solutions for a rigorous analysis. A general framework of so-called monotone schemes for the approximation of viscosity solutions to nonlinear PDEs was first introduced and analyzed in Barles and Souganidis [3]. It is well documented in the literature that the monotone approximation of degenerate diffusion problems in multiple dimensions generally requires complicated, so-called wide stencil schemes (see, for example, Debrabant and Jakobsen [13] and Ma and Forsyth [38]). The analysis in Reisinger [42] demonstrates clearly that the construction becomes more difficult when the correlation approaches ± 1, the above case being such a singular limit of perfect negative correlation between the driver of X and Y. Moreover, all schemes known to us which are monotone for general, possibly degenerate multidimensional equations have convergence order no larger than 1 in the mesh size and time step.

Initial numerical experiments with standard, nonmonotone finite difference schemes for the above PDE, in particular the 7-point and 9-point stencils for the diffusion term, exhibited severe instabilities for small mesh sizes.

In the following construction, we take advantage of a problem-specific coordinate transformation which allows us to define a monotone, second order accurate approximation to the second order terms. This will be supplemented with either monotone and first order, or nonmonotone and second order, backward differentiation formulae (BDF) for the first order derivative terms.

The second order version of the method is not theoretically guaranteed to converge to the viscosity solution in the degenerate case; however, recent results in Bokanowski and Debrabant [5] and Bokanowski, Picarelli, and Reisinger [6] show stability of BDF schemes in more regular cases, and we will demonstrate excellent empirical properties of the scheme below.

6.1.1. Mesh construction and diffusion approximation. We begin by simultaneously constructing a computational domain $[K^2/x_{\max}, x_{\max}] \times [y_{\min}, 1-y_{\min}] \subset \mathbb{R}_+ \times (0,1)$ and a nonuniform tensor-product mesh on that domain, where $x_{\max}$ and $y_{\min}$ will be chosen so as to make negligible the impact that imposing approximate data at the boundary has on the quantities of interest.

We first fix $x_{\max}$ and a positive integer N to define the x-coordinates of the mesh nodes by

$$x_i = K\exp(\sigma(i-N/2)h), \quad 0 \le i \le N, \tag{6.4}$$

so that $x_{N/2} = K$ for even N and h is chosen such that $x_N = x_{\max}$. This nonuniform mesh is motivated by the observation that the log transform $X \to \log X/\sigma$ leads to a standard Brownian motion with stochastic drift, i.e., satisfying the SDE

$$\mathrm{d}\left(\frac{1}{\sigma}\log X_t\right) = \mathrm{d}\widehat{W}_t + \left(\frac{1}{\sigma}\left(\mu_0 - \sigma\eta\widehat{Y}_t\right) - \frac{1}{2}\sigma\right)\,\mathrm{d}t, \tag{6.5}$$

and turns the differential operator $\mathcal{L}_{X,\widehat{Y}}$ into one with constant coefficients in x.

By a similar application of Itô's formula, one can further derive that, for $\hat{Y} \neq 0$ or 1,

$$\mathrm{d}\left(\frac{1}{\eta}\log\left(\frac{\widehat{Y}_t}{1-\widehat{Y}_t}\right)\right) = -\,\mathrm{d}\widehat{W}_t + \left(\frac{1}{2}\eta(2\widehat{Y}_t - 1) + \lambda\frac{1}{\eta\widehat{Y}_t}\right)\,\mathrm{d}t. \tag{6.6}$$

Inverting the map on the left-hand side, we define a mesh for the y-coordinate by

$$y_j = \frac{\exp(\eta(j-L/2)h)}{1+\exp(\eta(j-L/2)h)}, \quad 0 \le j \le L, \tag{6.7}$$

where L is chosen such that $y_0 = y_{\min}$ (and hence $y_L = 1 - y_{\min}$), a sufficiently small value, and centered at $y_{L/2} = 1/2$ for even L.

The purpose of these transformations is to fix the principal component of the diffusion matrix to $(-1,1)$ and facilitate the construction of a monotone, second order, narrow (i.e., using only neighboring mesh points) scheme. More concretely, combining the identities above, we obtain the following by simple Taylor expansion for smooth f:

$$
\begin{aligned}
(D^2 f)(t, x_i, y_j) &:= \frac{f(t, x_{i-1}, y_{j+1}) - 2f(t, x_i, y_j) + f(t, x_{i+1}, y_{j-1})}{h^2} \\
&= \frac{1}{2}\sigma^2 x_i^2 f_{xx} + \frac{1}{2}\eta^2 y_j^2 (1 - y_j)^2 f_{yy} - \sigma\eta x_i y_j (1 - y_j) f_{xy} \\
&\quad + \frac{1}{2}\sigma^2 x_i f_x + \frac{1}{2}\eta^2 y_j (1 - y_j)(1 - 2y_j) f_y + O(h^2),
\end{aligned} \tag{6.8}
$$

where the derivatives on the right-hand side are evaluated at (t, x_i, y_j).

The important feature of (6.8) is that the second-order part of the operator is approximated up to order two in h by a one-dimensional finite difference in a diagonal direction, plus some first order terms.

6.1.2. Drift approximation. We define the drift coefficients in (6.5) and (6.6) by

$$
\mu_x(t, x, y) := \frac{1}{\sigma}(\mu_0 - \sigma\eta y) - \frac{1}{2}\sigma, \qquad \mu_y(t, x, y) := \frac{\lambda}{\eta}\frac{1}{y} - \frac{1}{2}\eta(1 - 2y)
$$

(with the subscripts on μ_x and μ_y not denoting partial derivatives). These are precisely the the drifts of X and $\widehat{Y}$ minus the "correction terms" from (6.8) which have to be subtracted from D^2 for a consistent discretization of the second order terms in the PDE.

We approximate the first derivative in x, with coefficient μ_x, by an "upwinding" approximation

$$
(\mu_x D_x f)(t, x_i, y_j) = (\mu_x(t, x_i, y_j))^+ (D_x^+ f)(t, x_i, y_j) + (\mu_x(t, x_i, y_j))^- (D_x^- f)(t, x_i, y_j),
$$

where $(\cdot)^\pm$ denotes the positive and negative part, respectively, and $D_x^\pm$ is either the one-sided first order BDF1 approximation defined by

$$
(\overline{D}_x^\pm f)(t, x_i, y_j) := \mp \frac{f(t, x_i, y_j) - f(t, x_{i\pm 1}, y_j)}{h} = \sigma x f_x(t, x_i, y_j) + O(h),
$$

or the one-sided second order BDF2 approximation

$$
(\widehat{D}_x^\pm f)(t, x_i, y_j) := \mp \frac{3f(t, x_i, y_j) - 4f(t, x_{i\pm 1}, y_j) + f(t, x_{i\pm 2}, y_j)}{2h} = \sigma x f_x(t, x_i, y_j) + O(h^2).
$$

Two approximations to the first y-derivative are defined analogously.

6.1.3. Timestepping and overall scheme. Combining the approximations above, for all points (t, x_i, y_j) where f is smooth we have

$$
\begin{aligned}
\overline{L} f &:= D^2 f + \mu_x \overline{D}_x f + \mu_y \overline{D}_y f = \mathcal{L}_{X,\widehat{Y}} f + O(h), \\
\widehat{L} f &:= D^2 f + \mu_x \widehat{D}_x f + \mu_y \widehat{D}_y f = \mathcal{L}_{X,\widehat{Y}} f + O(h^2).
\end{aligned}
$$

For the time discretization, we follow Forsyth and Vetzal [23] and Reisinger and Whitley [44] to define a nonuniform time mesh of $M + 1$ points $t_m = T - (\sqrt{T} - mk)^2$, $m = 0, \ldots, M$, for $k = \sqrt{T}/M$. This transformation is motivated by the square-root behavior of both the exercise boundary and the value function at the strike close to maturity. The limited regularity prevents second order convergence of uniform timestepping schemes (see Forsyth and Vetzal [23]).

Taking into account this time transformation, we introduce either the BDF1 scheme (implicit Euler scheme)

$$\begin{aligned}\frac{f(t_{m+1},x_i,y_j)-f(t_m,x_i,y_j)}{k} &+ 2mk(\overline{L}f-rf)(t_m,x_i,y_j)\\ &= \left(\frac{\partial}{\partial t}+\mathcal{L}_{X,\widehat{Y}}-r\right)f(t_m,x_i,y_j)+O(k)+O(h),\end{aligned}$$

where $\overline{L}$ uses the BDF1 scheme for the drift also, or the BDF2 scheme

$$\begin{aligned}\frac{-f(t_{m+2},x_i,y_j)+4f(t_{m+1},x_i,y_j)-3f(t_m,x_i,y_j)}{2k} &+ 2mk(\widehat{L}f-rf)(t_m,x_i,y_j)\\ &= \left(\frac{\partial}{\partial t}+\mathcal{L}_{X,\widehat{Y}}-r\right)f(t_m,x_i,y_j)+O(k^2)+O(h^2),\end{aligned}$$

where $\widehat{L}$ uses the BDF2 scheme for the drift. The finite difference approximations are therefore consistent with $\mathcal{L}$ in (6.3) of order 1 and 2, respectively.

We can hence define a scheme for the numerical approximation $U^m = (U^m_{i,j})_{i,j}$ to the ESO value function u in the partial information case in the interior of the mesh by

$$\min\left(\frac{U^m_{i,j}-U^{m+1}_{i,j}}{k}-2mk\left((\overline{L}-rI)U^m\right)_{i,j}, U^m_{i,j}-\max(x_i-K,0)\right)=0, \quad 0\le m<M,\ 0<i<N,\ 0<j<L, \tag{6.9}$$

in the case of BDF1, and similarly in the case of BDF2.

From the construction of $\overline{L}$, the left-hand side of (6.9) is increasing in $U^m_{i,j}$, and decreasing in $U^{m'}_{i',j'}$ for all $(m',i',j')\neq(m,i,j)$, and therefore satisfies the definition of monotonicity in Barles and Souganidis [3]. The monotonicity is violated for the BDF2 scheme due to the alternating signs in the approximations to the first time and space derivatives. It is shown in Bokanowski and Debrabant [5] that such schemes still have good stability properties for American options under Black–Scholes. Although this analysis is not applicable here due to the degeneracy of the diffusion operator, we observe no stability issues in the numerical tests. We emphasize that the judicious choice of mesh and discretization of the second derivative terms is crucial for the stability of the scheme, due again to the degeneracy.

Summarizing, we obtain the following properties of the schemes.

Proposition 6.1. *The BDF1 scheme (6.9) is monotone and consistent with (6.1) in the interior $(-K^2/x^2_{\max}, x_{\max})\times(y_{\min}, 1-y_{\min})\times(0,T)$, of first order in both h and k. The BDF2 scheme is nonmonotone and consistent of second order in both h and k.*

6.1.4. Boundary and terminal conditions. We have four spatial boundaries with different characteristics as a result of the degeneracy of the drift and diffusion coefficients at some of the boundaries. The appropriate approximation of the boundary conditions is therefore essential for convergence to the correct solution of the initial boundary value problem. We discuss the boundaries in some detail in turn.

For $x = 0$, we set

$$U_{0,j}^m = 0, \qquad 0 \le m < M, 0 \le j \le L.$$

For $x = x_{\max}$, we set

$$U_{N,j}^m = \max(x_N - K, C(t_m, x_N, y_j)), \qquad 0 \le m < M, 0 < j < L,$$

where $C(t,x,y)$ is the Black–Scholes price of a European call option at time t and for underlying asset price $X_0 = x$, with constant interest rate r and dividend yield $r - (\mu_0 - \eta\sigma y)$, volatility σ, strike K, and maturity T. For those y where we can choose $x_{\max}$ such that $x^\star(T,y) \le x_{\max}$, the assumed boundary value coincides with the value function exactly. Generally, if $x^\star(T,y) > x_{\max}$ for some y, but with $x_{\max}$ several standard deviations away from K, the approximation error in the region of interest will be small.

For $y \to 0$, we have

$$\mathcal{L}_{X,\widehat{Y}} f \to \mu_0 x f_x + \frac{1}{2}\sigma^2 x^2 f_{xx} + \lambda f_y,$$

which we approximate at $(t_m, x_i, y_0) = (t_m, x_i, y_{\min})$ for $0 < i < N$ by

$$\left(\frac{\mu_0}{\sigma} - \sigma\eta y_{\min} - \frac{\sigma}{2}\right) D_x f + \frac{1}{2} D_x^2 f + \frac{\lambda}{\eta}\frac{1}{y_{\min}} D_y^+ f,$$

where $D_x^2 f(t_m, x_i, y_{\min}) = (f(t_m, x_{i+1}, y_{\min}) - 2f(t_m, x_i, y_{\min}) + f(t_m, x_{i-1}, y_{\min}))/h^2$. As the coefficient of the first y-derivative is positive, a right-sided difference (i.e., using only points in the interior of the domain) is appropriate and preserves monotonicity of the scheme.

For $y \to 1$, we have

$$\mathcal{L}_{X,\widehat{Y}} f \to \mu_0 x f_x + \frac{1}{2}\sigma^2 x^2 f_{xx},$$

which we approximate at $(t_m, x_i, y_L) = (t_m, x_i, 1 - y_{\min})$ for $0 < i < N$ by

$$\left(\frac{\mu_0}{\sigma} - \sigma\eta y_{\min} - \frac{\sigma}{2}\right) D_x f + \frac{1}{2} D_x^2 f,$$

using only boundary points.

As $y_{\min} \to 0$, the above approximations are consistent with the equation at $y = 0$ and $y = 1$, respectively. For fixed $y_{\min}$, to compute the solution at time t_m at a spatial point $(x_i, y) \in \{x_i\} \times [0, y_{\min})$, i.e., outside the computational domain, we extrapolate linearly from $y_0 = y_{\min}$ by $U_{i,0}^m + (y - y_0)(U_{i,1}^m - U_{i,0}^m)/(y_1 - y_0)$. This is of second order accuracy in $y_{\min}$ as the solution is smooth in this region. In particular, this is how the value in the regime $Y = 0$ is computed.

Lastly, the numerical terminal condition at $t = T$ is

$$U_{i,j}^M = \max(x_i - K, 0), \qquad 0 \le i \le N, 0 \le j \le L.$$

6.1.5. Penalization and Newton iteration. We now consider the penalty approximation

$$\frac{V_{i,j}^{m+1} - V_{i,j}^{m}}{k} + 2mk\left((\overline{L} - rI)V^m\right)_{i,j} + \rho \max\left(\max(x_i - K, 0) - V_{i,j}^m, 0\right) = 0 \tag{6.10}$$

for a penalty parameter $\rho > 0$ in the case of BDF1, and similarly in the case of BDF2.

Defining P as the $(N+1) \times (L+1)$ vector with $P_{i,j} = \max(x_i - K, 0)$ and $D(V)$ as the $((N+1)\times(L+1))^2$ diagonal matrix with $D_{(i,j),(i,j)}(V) = 1$ if $V_{i,j} < P_{i,j}$ and 0 otherwise, this can be rewritten as

$$\left((1 + rmk)I - 2k(mk)\overline{L} + \rho k D(V^m)\right) V^m = kV^{m+1} + D(V^m)P.$$

The solution of this type of equation by semismooth Newton iterations is discussed in [23]. In the case of the BDF1 scheme, $-\overline{L}$ is an M-matrix and hence $(1 + rmk)I - 2k(mk)\overline{L}$ is a strictly diagonally dominant M-matrix. This guarantees on the one hand convergence of the solution of the penalized solution $V = V(\rho)$ of (6.10) to U from (6.9) as $\rho \to \infty$, and on the other hand convergence of the Newton iteration in finitely many steps. In practice, we can choose the penalty parameter very large (e.g., 10^{10}) to make the difference between V and U negligible, without a negative impact on other properties of the scheme.

We end by stating without detailed proof the convergence result for the first order scheme.

Proposition 6.2. *The solution V of the penalized BDF1 scheme (6.10) converges to the solution u of (6.1) uniformly on compact subsets of $(0, T) \times (0, \infty) \times (0, 1)$ as k, h, $y_{\min} \to 0$ and $x_{\max}$, $\rho \to \infty$.*

Below, we report the number of required Newton iterations, alongside the empirically observed convergence order.

6.2. The full information case. We begin by observing that the full information ESO value function $v(t, x, i) \equiv v_i(t, x)$, $i = 0, 1$, satisfying (3.18)–(3.22), is also the unique solution in $[0, T] \times \mathbb{R}_+ \times \{0, 1\}$ of the equivalent linear complementarity problem (LCP)

$$\begin{aligned}
\min\left(-\mathcal{L}_0 v_0(t,x) + \lambda\left(v_0(t,x) - v_1(t,x)\right), v_0 - (x-K)^+\right) &= 0, \quad x \in \mathbb{R}_+,\ t \in [0,T),\\
\min\left(-\mathcal{L}_1 v_1(t,x), v_1 - (x-K)^+\right) &= 0, \quad x \in \mathbb{R}_+,\ t \in [0,T),\\
v_i(T,x) = (x-K)^+, &\quad x \in \mathbb{R}_+,\ i = 0,1,
\end{aligned}$$

where we repeat for convenience

$$\mathcal{L}_i f(t,x) = \left(\frac{\partial}{\partial t} + \mu_i x \frac{\partial}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2} - r\right) f(t,x), \quad i = 0, 1.$$

We approximate this LCP by

$$\begin{aligned}
\min\left(\frac{V_i^{0,m} - V_i^{0,m+1}}{k} - (LV^{0,m})_i + \lambda(V_i^{0,m} - V_i^{1,m}), V_i^{0,m} - \max(x_i - K, 0)\right) &= 0,\\
\min\left(\frac{V_i^{1,m} - V_i^{1,m+1}}{k} - (LV^{1,m})_i, V_i^{1,m} - \max(x_i - K, 0)\right) &= 0,\\
0 \le m < M,\ 0 < i < N&,
\end{aligned}$$

Table 1

For a sequence of meshes, given are the estimated pointwise errors of the BDF1 and BDF2 schemes, the resulting convergence orders, the average number of Newton iterations, and the run time.

		BDF2				BDF1	
$N = L$	M	Error	Order	Avg. iter.	CPU (s)	Error	Order
24	23	$1.61 \cdot 10^{-1}$	-	2.4	0.38	$1.52 \cdot 10^{0}$	-
34	33	$7.95 \cdot 10^{-2}$	2.03	2.5	1.2	$1.05 \cdot 10^{0}$	1.06
46	46	$3.71 \cdot 10^{-2}$	2.20	2.6	2.9	$6.79 \cdot 10^{-1}$	1.26
66	65	$2.00 \cdot 10^{-2}$	1.79	2.6	7.6	$4.94 \cdot 10^{-1}$	0.92
92	91	$9.77 \cdot 10^{-3}$	2.06	2.7	34	$3.39 \cdot 10^{-1}$	1.08

where x_i is as in (6.4) and

$$(LV^{j,m})_i = \left(\mu_j - \frac{1}{2}\sigma^2\right) \frac{V_{i+1}^{j,m} - V_{i-1}^{j,m}}{2h} + \frac{1}{2} \frac{V_{i+1}^{j,m} - 2V_i^{j,m} + V_{i-1}^{j,m}}{h^2} - rV_i^{j,m}.$$

Consistency and monotonicity, and hence convergence, follow directly in this case. The scheme is of first order in k and of second order in h. The computational complexity is smaller than in the two-dimensional case though, and we therefore do not propose a second order version. Penalization is now applied separately to the two components, and a Newton iteration can be applied in a natural way to the system of equations.

6.3. Numerical tests. We discuss here some tests for the numerical performance of the partial information algorithm. The full information case is straightforward and we do not report our test results here. In this section, we test in detail the convergence of the finite difference scheme with respect to the discretization parameters. The financial parameters chosen are $\sigma = 0.3$, $\lambda = 0.1$, $\mu_0 = 0.08$, $\mu_1 = -0.05$, $r = 0.025$, $T = 10$, $K = 100$. The truncation parameters were $y_{\min} = 0.02$, $x_{\max} = 8K$, and the mesh parameters h and k varied as detailed below.

We list in Table 1 various quantities of interest for different mesh refinements, for both the BDF1 and BDF2 scheme, where N and L are (as above) the number of mesh intervals in the x and y directions, and M the number of time steps. The numbers for N and M are arrived at by the rule $N = 2\lceil N_0\sqrt{2}^n\rceil$, $n \geq 0$, with $N_0 = 8$, and $M = \lceil M_0\sqrt{2}^n\rceil$, $n \geq 0$, with $M_0 = 16$. This is motivated by the identical convergence order in h and k for each of the schemes. Then L is determined as explained below (6.7) and is also proportional to N and M. We ensure, moreover, that N is even for the mesh construction above. Here, N_0 and M_0 are chosen empirically so that the errors from the time and space discretization are similar. The fact that we arrived at $N = L \approx M$ for these particular model parameters is coincidental.

The numerical solution is evaluated at $(t, x, y) = (0, K, 1/2)$ and then the error (third and seventh columns) estimated by extrapolation from the solutions for subsequent mesh refinements; the order (fourth and eighth columns) is then estimated from the errors for consecutive meshes. The numbers clearly demonstrate first order and second order convergence for the BDF1 and BDF2 scheme, respectively. This behavior is further illustrated in Figure 1. The error on the finest level is smaller than 0.01 absolutely, or 1 basis point given a strike of 100.

We also report in Table 1 the number of Newton iterations needed to solve the nonlinear system, averaged over all time points. For nonuniform meshes, the number is typically higher

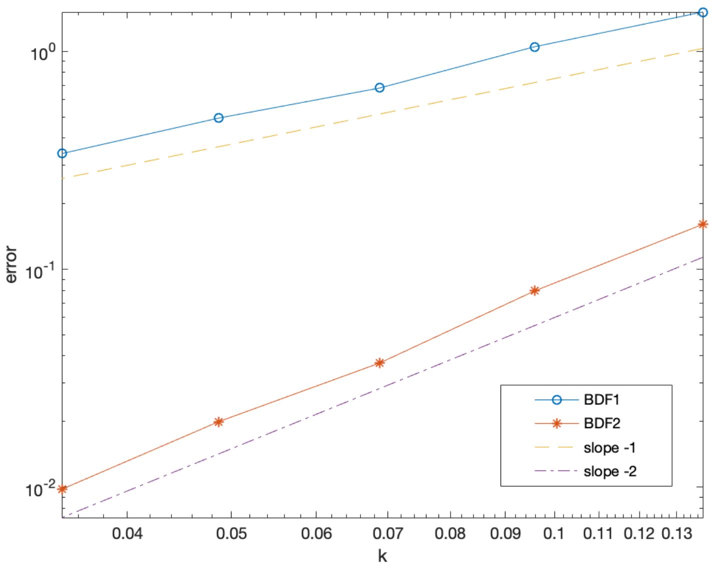

Figure 1. *Estimated pointwise errors for decreasing time steps as in Table 1. The comparison with lines of slope −1 and −2 in the loglog plot demonstrates first and second order convergence of the BDF1 and BDF2 scheme, respectively.*

close to maturity due to the singular behavior of the exercise boundary, but this effect is alleviated by the local refinement.

The total number of unknowns increases by a factor of $\sqrt{2}^3 \approx 2.8$ upon refinement, and this is a lower bound for the asymptotic increase in computational complexity. In practice, the cost of solving each linear system within the Newton iteration, involving a sparse block-tridiagonal matrix, using the default sparse equation solver in MATLAB, increases superlinearly. For optimized performance a multigrid solver as in Reisinger and Rotaetxe Arto [43] could be used. Both the iteration count and computational time are very similar between the two schemes, and we only report the BDF2 ones.

7. Numerical results: ESO exercise and valuation.

This section demonstrates numerically the exercise policies of the agents in section 7.1. In section 7.2, we undertake a study of postexercise stock returns which supports the approach taken in the empirical literature on private information. We consider the impact of the information differential on ESO valuation in section 7.3.

7.1. Difference in exercise policies due to information differential.

We are primarily interested in the difference between the exercise policies for the agents due to the information differential they have. To illustrate exercise patterns for both agents, we numerically solve for the thresholds of both types of agents and simulate the stock price to demonstrate exercise behavior. A set of outputs with various parameter values is plotted in Figure 2. In each panel we display the stock price, the exercise boundary for the agent with full information, $x_i^*(t); i = 0, 1; t \in [0, T]$, and the partially informed agent's exercise boundary, $x^*(t, .); t \in [0, T]$

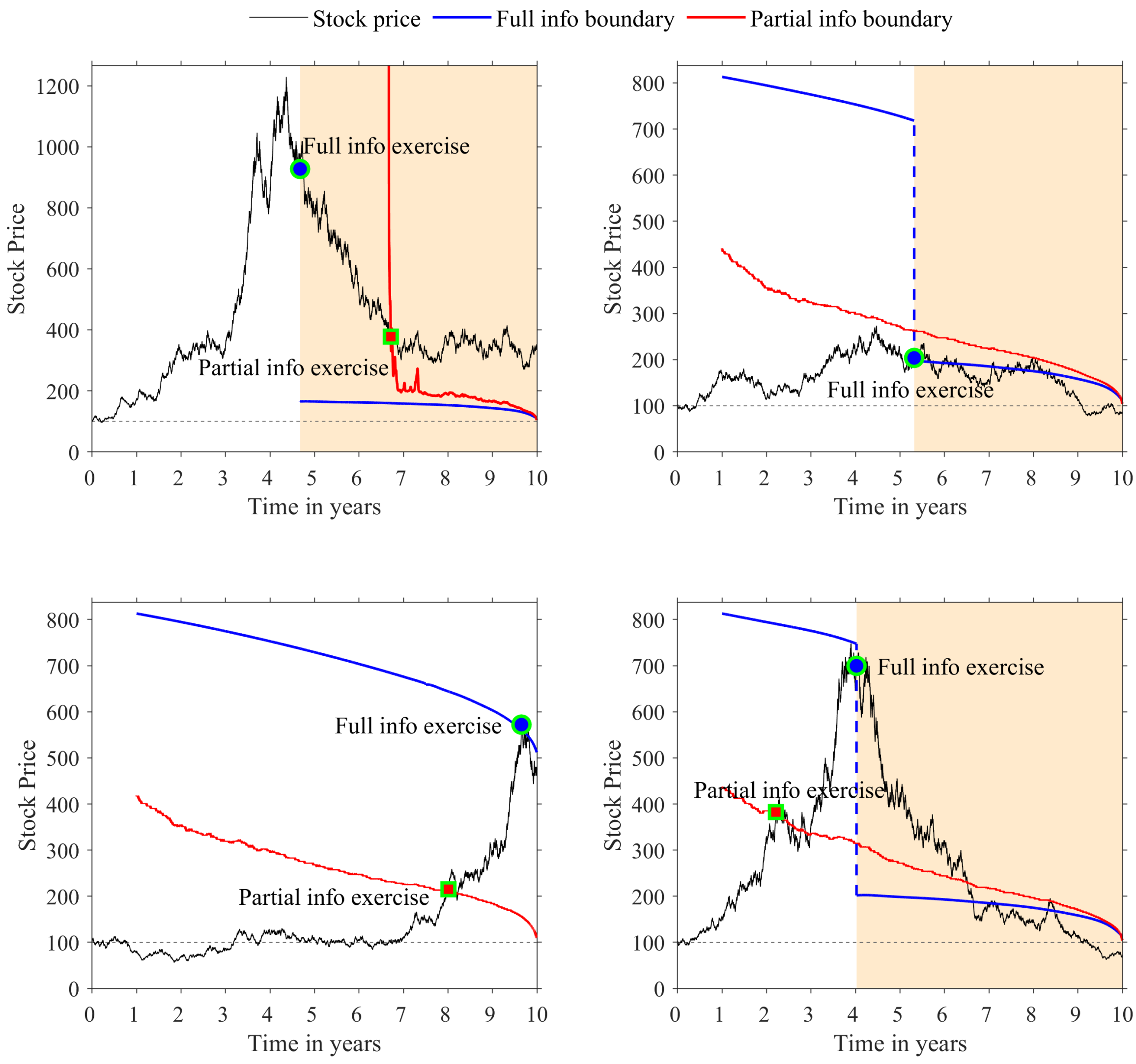

Figure 2. Monte Carlo simulations of the stock price, thresholds, and exercise decisions of the agents with full and partial information. *In each panel we display the stock price, the exercise boundary for the full information case, and the exercise boundary for the partial information model, with* $\hat{Y}_0 = \mathbb{E}[Y_0] = y_0 = 0$*. Exercise decisions of the full information agent and partial information agent with* $y_0 = 0$ *are marked with circles and squares, respectively. The option maturity is ten years with a one-year vesting period* $t_v = 1$*, and granted at-the-money with* $X_0 = K = 100$*. In each panel, the shaded background indicates the switch in drift regime to* $\mu_1 < \mu_0$*. In all panels, we take parameter values for the transition intensity* $\lambda = 10\%$ *and volatility* $\sigma = 30\%$*, and the risk-free rate is* $r = 2.5\%$*. In the top left panel, expected returns are given by* $\mu_0 = 8\%, \mu_1 = -5\%$*, so that* $\mu_0 > r > \mu_1$ *holds. In all other panels, expected returns in the two regimes are* $\mu_0 = 2\%$*,* $\mu_1 = -2\%$*, so that* $r > \mu_0 > \mu_1$*.*

with $\hat{Y}_0 = \mathbb{E}[Y_0] = y_0 = 0$. We set the switch intensity to be $\lambda = 10\%$, which implies a probability of 63% of μ_0 switching to μ_1 during the option's life. Given that the "vast majority of options are granted at-the-money" with maturities of ten years (Carpenter, Stanton, and Wallace [11]) we consider an ESO granted at-the-money with $X_0 = K = 100$ and maturity

$T = 10$ years. We include a vesting period of one year, $t_v = 1$. The shaded area in each panel denotes the time after the change point has occurred, i.e., the drift has switched from μ_0 to μ_1. Exercise decisions are recorded on each plot for both the partial information agent (with a square) and the fully informed agent (with a circle).

In the top-left panel, we observe, since $\mu_0 > r$, $x_0^*(\cdot) = \infty$ and no exercise occurs before the change point. The agent with full information exercises on the change point. The threshold of the partially informed agent, $x^*(\cdot, \cdot)$, rapidly drops from infinity following the change point, as the filtering puts higher weight on the switch having occurred. The agent with partial information exercises as the stock price reaches the threshold. However, the fully informed agent has obtained a far larger option payoff in this scenario.

The remaining three panels consider the case $r > \mu_0 > \mu_1$. The upper-right panel demonstrates a scenario where the stock price is not performing as well as in the left panel, and the agent with partial information never exercises. The agent with full information exercises on the change point, although the stock price does go slightly higher after that. The agent with full information has obtained a higher option payoff than the agent with partial information, as the latter never exercises and the option is out-of-the-money at maturity.

In the lower-left panel, where no change point occurs before option maturity, consistent with Proposition 3.4, $x_0^*(T-) = \max(K, \frac{r}{r-\mu_0}K) = 500$. In this panel, the stock does very well. The stock price first reaches the boundary of the partially informed agent and, finally, the much higher boundary of the agent with full information. Under this scenario, the fully informed agent has benefited from the additional information (the knowledge that the switch has not occurred) and has secured a much higher payoff than the agent with partial information.

Finally, the lower-right panel demonstrates a scenario where the agent with full information exercises in direct response to the switch and benefits from the additional information. In this panel, the partial information agent has already exercised as the stock price crosses their boundary. The agent with full information continues to wait as he knows the switch has not occurred. He then benefits with a larger exercise payoff by exercising exactly at the change point.

In all panels, we observe that the boundaries respect the mathematical results of sections 3 and 4. The full information boundaries are in accordance with Corollary 3.3 since we can observe the ordering $x_0^*(t) \geq x_1^*(t) \geq K$ for the three panels where $r > \mu_0 > \mu_1$, and, when $\mu_0 > r$, we see $x_0^*(t) = \infty$. For any μ_i, we have $x_i^*(T) = K$, and $x_i^*(T-) = \max(K, \frac{r}{r-\mu_i}K)$ for $\mu_i < r$ from Proposition 3.4 is also satisfied. In the top-left panel with $\mu_0 > r$, consistent with Remark 4.1, we have no early exercise for the agent with partial information. The exercise boundary for the agent with partial information, $x^*(t, .)$, is indeed decreasing in t, in accordance with Lemma 4.2, and the boundaries respect Lemma 4.5.

In Figure 3, we illustrate the complete exercise surfaces generated by the model for the agents with full and partial information. We plot the full information thresholds, $x_0^*(t), x_1^*(t)$, $t \in [0, T]$, and the partial information surface, $x^*(t, y)$, $t \in [0, T], y \in [0, 1]$. The behavior with the full and partial information thresholds with respect to time is consistent with that displayed in Figure 2. For example, consistent with Proposition 3.4, we have for the full information boundaries, $x_0^*(10-) = 500, x_1^*(10-) = 100$. Turning to the behavior of the thresholds with respect to varying $\hat{Y}$, the exercise surface for the agent with partial information, $x^*(t, y)$, is indeed decreasing in y, in accordance with Lemma 4.2.

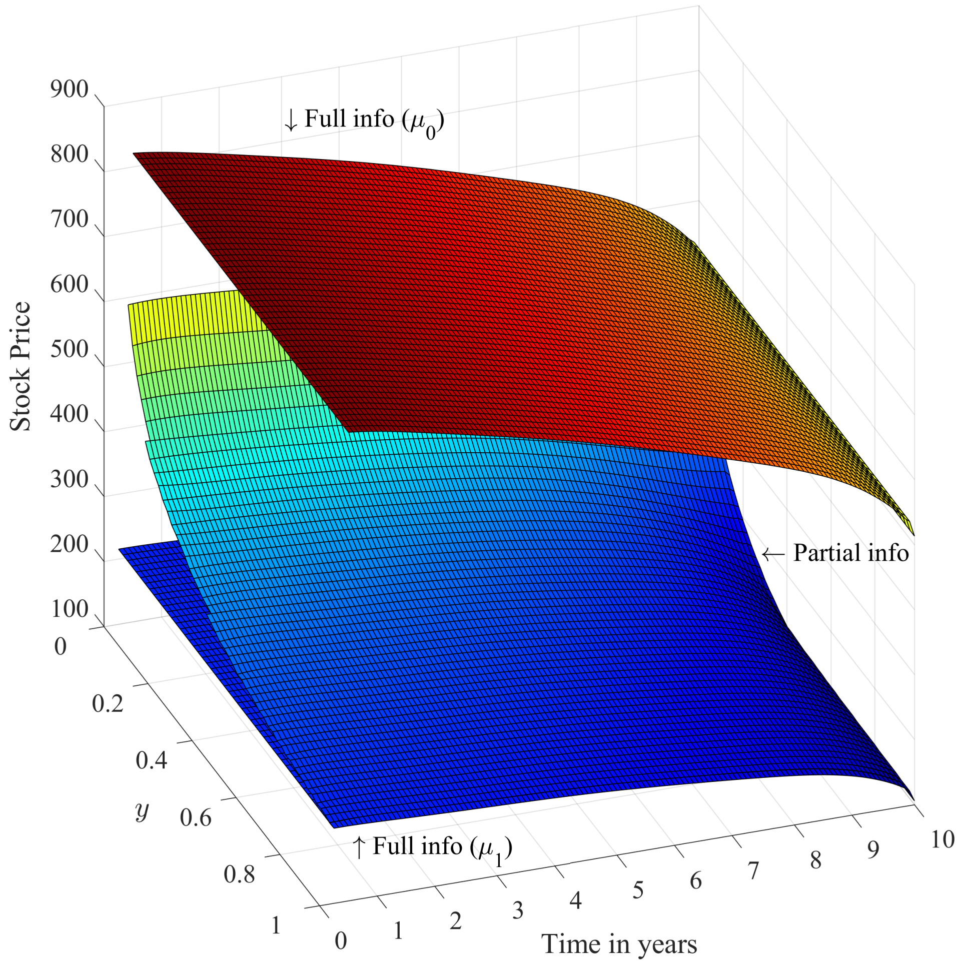

Figure 3. Exercise surfaces under full and partial information against time and $y \in [0,1]$ the spatial dependence arising from the filtered process $\hat{Y}$. *The uppermost and lowermost surfaces are those of the agent with full information: the uppermost surface* $x_0^*(.)$ *in regime* 0 *with* μ_0*, and the lowermost surface* $x_1^*(.)$ *in regime* 1 *with* μ_1*. These do not depend upon* y*, so each surface for the full information agent is constant in the* y *direction, and has been plotted for comparison with the surface of the agent with partial information. The exercise surface* $x^*(t, y)$*,* $t \in [0, T], y \in [0, 1]$*, for the agent with partial information lies between the two surfaces from the full information problem. The option maturity is ten years and granted at-the-money with* $X_0 = K = 100$*. Expected returns in the two regimes are* $\mu_0 = 2\%$*,* $\mu_1 = -2\%$*, transition intensity* $\lambda = 10\%$*, and volatility* $\sigma = 30\%$*, and the risk-free rate is* $r = 2.5\%$*.*

7.2. An application to postexercise returns. In this section, we demonstrate how our model can be linked to the empirical finance literature on private information and the exercise of ESOs. In fact, our model provides a consistent theoretical foundation for the empirical tests conducted in this literature. A body of papers (Aboody et al. [1], Brooks, Chance, and Cline [7], and Cicero [12]) aim to identify and evidence that executives use private information when exercising their company ESOs. (Note that these papers, and ours, do not take any stance on the legality of such exercises.) The idea is “if the executive has negative information, the stock (owned by them) would almost surely be sold, and in all likelihood the stock would perform poorly for a period of time thereafter." (Brooks, Chance, and Cline [7, p. 733]). These studies examine ESO exercise data in which the stock is sold upon exercise. The general approach is then to examine the long-term abnormal returns after the exercise of ESOs. If the abnormal returns are significantly negative following exercise, there is support for the explanation of private information being a factor in exercise decisions. Brooks, Chance, and Cline [7] match firms with ESO exercises of top executives, believed to hold private information, to firms with no record of top executive ESO exercises, but with similar firm characteristics. They observe one year of stock data following each top executive option exercise and compute the BHAR (buy-and-hold-abnormal returns) to be the so-called insider returns minus the matched returns. Brooks, Chance, and Cline [7] find strong evidence of ESO exercise due to insider information, via significant negative differences in the returns. Insider exercises are linked to significantly negative postexercise returns over the following year.

We use our model of differential information to generate postexercise returns over the following year and compare any difference between returns following exercises by our agent with full information versus our agent with partial information. Our full information agent knows the change point in the stock price process, when the expected return of the stock drops. Then, if our model is to be consistent with the approach of Brooks, Chance, and Cline [7], we need to demonstrate that the difference between the average postexercise returns from fully and partially informed agents is also negative. To be in line with the literature, we consider simulated returns for one year following each option exercise, and we only include exercises which are more than one year before option maturity (exercises closer to maturity are considered less likely to be information related).

In Figure 4 we display the results of the simulations. The left-hand panel uses volatility 20%, while the right-hand panel uses 30%. We keep the expected return after a change point fixed at $\mu_1 = -10\%$, but take three values for the expected return μ_0. We first observe that the mean cumulative log-returns postexercise for the case of full information do not vary much from the different values of initial expected return μ_0. Recall from Corollary 3.3, with full information, and with $\mu_0 = 8\%, 18\% > r$, exercises occur only in the bad state. Thus the one-year log-returns are $\mu_1 - 0.5\sigma^2$ (for the left panel, -12%, and for the right panel, -14.5%). With $\mu_0 = 2\%$, there are some early exercises in the good state, and their occurrence increases with volatility, as shown by the plots. With only partial information, the cumulative log-returns postexercise vary much more with the value of μ_0. We see the postexercise returns are worse the higher the expected return μ_0. The one-year log-returns for the partial information case vary between about -2.2% to -8.2% when volatility is 20%, and -5.3% to -8.4% for volatility 30%.

Overall, the simulations support our conjecture that, indeed, exercises by the agent with full information are followed by significantly negative stock returns, and the difference between average postexercise returns for fully and partially informed agents is significantly negative. For our simulations, this difference between mean postexercise returns for fully and partially informed agents varies between about -3.8% and -9.7%, depending on the expected stock return μ_0 and volatility, covering the range of values reported by Brooks, Chance, and Cline [7]. Our model thus provides theoretical support for the tests conducted in the empirical literature to evidence so-called insider exercises.

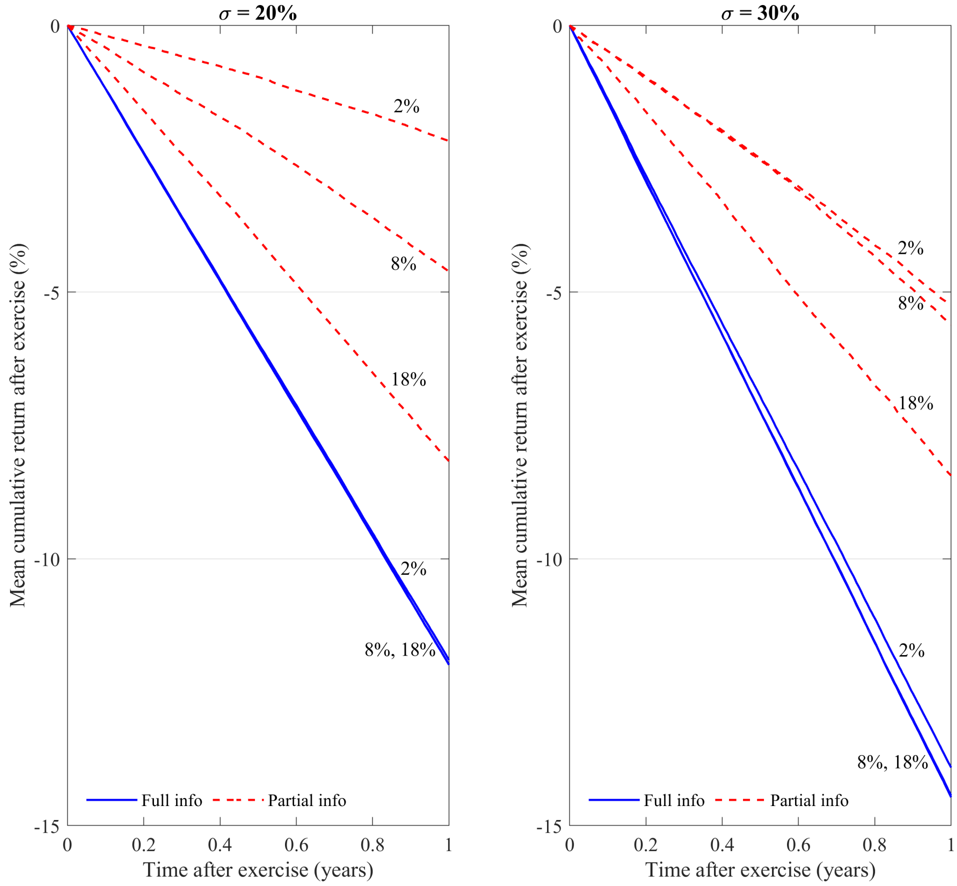

Figure 4. Mean cumulative postexercise returns with full and partial information over one year. *In the left panel, volatility is* $\sigma = 20\%$*, and in the right panel, volatility is* $\sigma = 30\%$*. The expected return* $\mu_1 = -10\%$ *is fixed, and expected return* $\mu_0 = 2, 8, 18\%$*. The transition intensity is* $\lambda = 10\%$*, and the risk-free rate is* $r = 2.5\%$*. The option maturity is ten years and granted at-the-money with* $X_0 = K = 100$*. Simulations use* 1 *million price paths.*

7.3. ESO valuation. We now turn to the impact of differential information about the stock price on ESO valuation by the agents themselves. We emphasize that the ESO values we report represent the value to the individual agent, often termed subjective value in the literature on ESO compensation (see Carpenter [9]). It is the value under the $\mathbb{P}$ measure.

Table 2 reports the time-zero ESO values for the agent with full information, $V = V_0$, and for the agent with partial information, $U = U_0$. The table also gives a breakdown of each ESO value into its European (labeled E_V and E_U) and American (labeled A_V and A_U) early exercise components. This breakdown shows the value differential arises entirely from the American early exercise component of the ESO values. As the simulations demonstrate in section 7.1, the agent with full information uses this knowledge to time his option exercise advantageously.

The additional value that the agent with full information places on the ESO is significant in magnitude. Consider the American early exercise value as a proportion of total ESO value

Table 2

Comparative statics for the full and partial information option values. *Each subpanel of six numbers contains the option values for full information in the left column and partial information in the right column. Each column contains (from top to bottom) the American component, the European component, and the total ESO value (sum of European and American). We have, for the full information case, $V = E_V + A_V$, and for the partial information model $U = E_U + A_U$. The option maturity is ten years and granted at-the-money with $X_0 = K = 100$. Parameter values considered are $\mu_0 = 2\%, 8\%, 18\%$, $\mu_1 = -2\%, -5\%, -10\%$, transition intensity $\lambda = 10\%, 20\%$, and volatility $\sigma = 20\%, 30\%, 40\%$, and the risk-free rate is fixed at $r = 2.5\%$. We fix $y_0 = 0$.*

Read:		$\lambda = 10\%$						$\lambda = 20\%$					
A_V E_V V	A_U E_U U	$\sigma = 20\%$, μ_1						$\sigma = 20\%$, μ_1					
		-2%		-5%		-10%		-2%		-5%		-10%	
μ_0	2%	2.9	1.5	5.3	2.7	7.7	3.8	4.2	3.1	7.4	5.3	10.5	7.1
		23.1	23.1	19.4	19.4	16.0	16.0	18.9	18.9	13.5	13.5	8.9	8.9
		26.0	24.6	24.7	22.1	23.7	19.8	23.0	21.9	20.9	18.8	19.4	16.0
	8%	5.5	0.0	9.8	0.3	14.6	1.3	7.3	0.7	13.0	2.5	19.0	5.0
		60.0	60.0	54.5	54.5	48.8	48.8	40.6	40.6	32.9	32.9	25.5	25.5
		65.4	60.0	64.2	54.8	63.4	50.0	47.9	41.3	45.9	35.5	44.5	30.5
	18%	12.8	0.4	21.9	1.8	33.3	4.8	16.1	1.8	27.8	5.2	41.5	10.5
		210.6	210.6	200.5	200.5	188.6	188.6	122.8	122.8	109.6	109.6	94.9	94.9
		223.3	211.0	222.4	202.3	221.9	193.4	138.9	124.6	137.4	114.8	136.4	105.4
		$\sigma = 30\%$, μ_1						$\sigma = 30\%$, μ_1					
		-2%		-5%		-10%		-2%		-5%		-10%	
μ_0	2%	3.6	2.5	6.2	4.1	9.4	5.9	5.1	4.2	8.8	7.1	13.0	9.9
		32.2	32.2	27.8	27.8	23.2	23.2	27.7	27.7	21.2	21.2	14.7	14.7
		35.8	34.7	34.1	32.0	32.6	29.1	32.8	31.9	30.1	28.3	27.7	24.6
	8%	5.7	0.0	10.0	0.2	15.3	0.9	7.8	1.0	13.5	3.2	20.3	6.1
		68.7	68.7	62.7	62.7	55.9	55.9	49.5	49.5	41.0	41.0	31.8	31.8
		74.4	68.7	72.7	62.9	71.3	56.8	57.3	50.5	54.5	44.2	52.2	37.9
	18%	12.2	0.0	20.9	0.2	32.2	1.2	15.6	0.5	26.5	2.0	40.4	5.5
		216.0	216.0	205.8	205.8	193.3	193.3	130.0	130.0	116.4	116.4	100.6	100.6
		228.2	216.0	226.7	206.0	225.5	194.5	145.5	130.4	143.0	118.5	141.0	106.0
		$\sigma = 40\%$, μ_1						$\sigma = 40\%$, μ_1					
		-2%		-5%		-10%		-2%		-5%		-10%	
μ_0	2%	4.3	3.3	7.3	5.4	11.0	7.7	6.1	5.3	10.3	8.7	15.4	12.5
		40.8	40.8	35.9	35.9	30.3	30.3	36.0	36.0	28.7	28.7	20.6	20.6
		45.1	44.1	43.2	41.3	41.3	38.0	42.1	41.3	39.0	37.4	36.0	33.1
	8%	6.3	0.0	10.7	0.2	16.4	1.0	8.6	1.5	14.6	4.2	22.1	7.9
		77.9	77.9	71.4	71.4	63.8	63.8	58.4	58.4	49.1	49.1	38.6	38.6
		84.2	77.9	82.2	71.6	80.3	64.8	67.0	59.8	63.8	53.4	60.7	46.5
	18%	12.3	0.0	20.8	0.0	32.1	0.4	15.8	0.2	26.6	1.1	40.6	3.7
		223.5	223.5	213.0	213.0	199.9	199.9	138.3	138.3	124.3	124.3	107.3	107.3
		235.8	223.5	233.8	213.1	232.0	200.3	154.1	138.4	150.9	125.4	148.0	111.0

for each of the full and partial information cases. For example, with $\lambda = 10\%$, $\mu_0 = 8\%$, $\mu_1 = -5\%$, $\sigma = 30\%$, the American early exercise value represents 13.8% (10/72.7) of the ESO value for full information, and 0.32% (0.2/62.9) of value for partial information. If we compare these American-as-proportion-of-total values for the full and partial information agents, we see that the magnitude is much larger for the agent with full information. In our example, we see the 13.8% is about 43 times larger than the 0.32%. This ratio varies between around 1.2, up to values as high as 69. There are also some zero values for the American early exercise value under partial information, which tend to be for high μ_0 and the best case of -2% for μ_1, indicating no early exercises take place. In these scenarios, the agent with full information gains significantly as he uses his additional information on the change point to time exercise advantageously.

The table documents how the full and partial information ESO values vary with changes in stock-specific parameters μ_0, μ_1, and σ and the transition intensity λ. The option values under full and partial information increase with the value of expected return μ_0. Under the partial information model, the American component of value often drops with μ_0, consistent with there being relatively few exercises for high values of μ_0.

Under both full and partial information, option values decrease as μ_1 decreases. However, the American component of value increases with $|\mu_1|$, for both full and partial information, indicating that the ability to time the exercise of the option is more valuable when the expected return following a change point is worse. For example, scenarios with a low μ_0 of 2%, the worst case for μ_1 of -10%, and the transition probability $\lambda = 0.2$, the American component of option value can be as high as 40–50% of ESO value.

Volatility increases the full and partial information option values. The European component is increasing in volatility, but the American component can increase or decrease. If μ_0 is sufficiently high, volatility can reduce the American component of value in both full and partial information scenarios.

A higher probability of a downward jump in expected return (higher λ) reduces the full and partial information ESO values. The European component of value is reduced, as a higher λ simply means a greater chance of switching to the bad regime. However, the American component of value increases with λ because the ability to time the exercise becomes more important when the chance of the bad state is increased. This is true for both the agent with full and the agent with partial information.

We now turn to briefly examine the impact of vesting on ESO valuation. Section 5 described the effect of a vesting period $[0, t_v)$ on option exercise. Table 3 documents the ESO values for both a three- and a five-year vesting period for a representative subset of market parameters from Table 2 and fixing volatility at $\sigma = 30\%$. Hence the ESO values should be compared to the middle panel of Table 2, where the same volatility is used but no vesting period.

As we anticipate, the American early exercise values are nonincreasing as t_v increases, as the option becomes unexercisable for a larger share of the life of the option. For example, when $\mu_0 = 2\%, \mu_1 = -5\%, \sigma = 30\%$, and $\lambda = 10\%$, the early exercise value for full information falls from 6.2 to 6 to 5 as t_v increases from 0 to 3 years to 5 years. Corresponding early exercise values in the partial information setting are 4.1, 4.0, 3.8. For some parameters, say when μ_0 is high, the early exercise value in the case with partial information did not vary with t_v, as

Table 3

The effect of a vesting period of 3 and 5 years on ESO valuation by agents with full and partial information. *We take $\sigma = 30\%$, and thus values should be compared with the middle panels of Table 2. Each subpanel of six numbers contains the option values for full information in the left column and partial information in the right column. Each column contains (from top to bottom) the American component, the European component, and the total ESO value (sum of European and American). We have, for the full information case, $V = E_V + A_V$, and for the partial information model $U = E_U + A_U$. The option maturity is ten years and granted at-the-money with $X_0 = K = 100$. We consider vesting periods of $t_v = 3$ years and $t_v = 5$ years. Parameter values considered are $\mu_0 = 2\%, 8\%, 18\%$, $\mu_1 = -2\%, -5\%, -10\%$, and transition intensity $\lambda = 10\%, 20\%$, and the risk-free rate is fixed at $r = 2.5\%$. We fix $y_0 = 0$.*

Read:		$\lambda = 10\%$						$\lambda = 20\%$					
A_V	A_U	$t_v = 3$ years						$t_v = 3$ years					
E_V	E_U	μ_1						μ_1					
V	U	-2%		-5%		-10%		-2%		-5%		-10%	
μ_0	2%	3.5	2.5	6.0	4.1	8.6	5.8	4.9	4.2	8.3	7.0	11.5	9.5
		32.2	32.2	27.8	27.8	23.2	23.2	27.7	27.7	21.2	21.2	14.7	14.7
		35.7	34.7	33.8	31.9	31.8	29.0	32.6	31.9	29.5	28.2	26.2	24.2
	8%	5.6	0.0	9.6	0.2	14.2	0.9	7.5	1.0	12.7	3.2	18.4	6.1
		68.7	68.7	62.7	62.7	55.9	55.9	49.5	49.5	41.0	41.0	31.8	31.8
		74.3	68.7	72.3	62.9	70.1	56.8	57.0	50.5	53.7	44.2	50.2	37.9
	18%	11.9	0.0	20.1	0.2	30.1	1.2	15.0	0.4	25.3	2.0	37.4	5.4
		216.0	216.0	205.8	205.8	193.3	193.3	130.0	130.0	116.4	116.4	100.6	100.6
		227.9	216.0	225.9	206.0	223.4	194.5	145.0	130.4	141.7	118.4	138.0	106.0
		$t_v = 5$ years						$t_v = 5$ years					
		μ_1						μ_1					
		-2%		-5%		-10%		-2%		-5%		-10%	
μ_0	2%	3.3	1.3	5.0	3.8	6.8	5.1	4.3	3.8	6.8	6.0	8.8	7.7
		32.2	32.2	27.8	27.8	23.2	23.2	27.7	27.7	21.2	21.2	14.7	14.7
		35.5	34.5	32.8	31.6	30.0	28.3	32.0	31.5	28.0	27.2	23.5	22.4
	8%	5.0	0.0	8.2	0.2	11.7	0.9	6.5	1.0	10.6	3.1	14.5	5.9
		68.7	68.7	62.7	62.7	55.9	55.9	49.5	49.5	41.0	41.0	31.8	31.8
		73.7	68.7	70.9	62.9	67.6	56.8	56.0	50.5	51.6	44.1	46.3	37.7
	18%	10.7	0.0	17.6	0.2	26.0	1.2	13.2	0.4	21.4	2.0	30.6	5.4
		216.0	216.0	205.8	205.8	193.3	193.3	130.0	130.0	116.4	116.4	100.6	100.6
		226.7	216.0	223.4	206.0	219.3	194.5	143.2	130.4	137.8	118.4	131.2	106.0

these are situations where there are no exercises taking place when there is no vesting period, and thus additional exercise restrictions via vesting do not alter the agent's value.

REFERENCES

[1] D. Aboody, J. Hughes, J. Liu, and W. Su, *Are executive stock option exercises driven by private information?*, Rev. Account. Stud., 14 (2008), pp. 551–571.

[2] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Stoch. Model. Appl. Probab. 60, Springer, New York, 2009, https://doi.org/10.1007/978-0-387-76896-0.

[3] G. Barles and P. E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptot. Anal., 4 (1991), pp. 271–283.

[4] E. Bayraktar, *A proof of the smoothness of the finite time horizon American put option for jump diffusions*, SIAM J. Control Optim., 48 (2009), pp. 551–572, https://doi.org/10.1137/070686494.
[5] O. Bokanowski and K. Debrabant, *High Order Finite Difference Schemes for Some Nonlinear Diffusion Equations with an Obstacle Term*, preprint, https://arxiv.org/abs/1802.05681, 2018.
[6] O. Bokanowski, A. Picarelli, and C. Reisinger, *Stability and Convergence of Second Order Backward Differentiation Schemes for Parabolic Hamilton-Jacobi-Bellman Equations*, preprint, https://arxiv.org/abs/1802.07146, 2018.
[7] R. Brooks, D. Chance, and B. Cline, *Private information and the exercise of executive stock options*, Financial Management, 41 (2012), pp. 733–764.
[8] J. Buffington and R. J. Elliott, *American options with regime switching*, Int. J. Theor. Appl. Finance, 5 (2002), pp. 497–514, https://doi.org/10.1142/S0219024902001523.
[9] J. Carpenter, *The exercise and valuation of executive stock options*, J. Financial Econom., 48 (1998), pp. 127–158.
[10] J. Carpenter and B. Remmers, *Executive stock option exercises and inside information*, J. Business, 74 (2001), pp. 513–534.
[11] J. Carpenter, R. Stanton, and N. Wallace, *Employee stock option exercise and firm cost*, J. Finance, 74 (2019), pp. 1175–1216.
[12] D. Cicero, *The manipulation of executive stock option exercise strategies: Information timing and backdating*, J. Finance, 64 (2009), pp. 2627–2663.
[13] K. Debrabant and E. R. Jakobsen, *Semi-Lagrangian schemes for linear and fully non-linear diffusion equations*, Math. Comp., 82 (2013), pp. 1433–1462.
[14] J.-P. Décamps, T. Mariotti, and S. Villeneuve, *Investment timing under incomplete information*, Math. Oper. Res., 30 (2005), pp. 472–500, https://doi.org/10.1287/moor.1040.0132.
[15] J.-P. Décamps, T. Mariotti, and S. Villeneuve, *Investment timing under incomplete information: Erratum*, Math. Oper. Res., 34 (2009), pp. 255–256, https://doi.org/10.1287/moor.1080.0357.
[16] J. Detemple, *American-Style Derivatives: Valuation and Computation*, Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2006.
[17] E. Ekström, *Properties of American option prices*, Stochastic Process. Appl., 114 (2004), pp. 265–278, https://doi.org/10.1016/j.spa.2004.05.002.
[18] E. Ekström and C. Lindberg, *Optimal closing of a momentum trade*, J. Appl. Probab., 50 (2013), pp. 374–387, https://doi.org/10.1239/jap/1371648947.
[19] E. Ekström and B. Lu, *Optimal selling of an asset under incomplete information*, Int. J. Stoch. Anal., (2011), 543590.
[20] E. Ekström and M. Vannestål, *American options and incomplete information*, Int. J. Theor. Appl. Finance, 22 (2019), 1950035, https://doi.org/10.1142/s0219024919500353.
[21] N. El Karoui, M. Jeanblanc-Picqué, and S. E. Shreve, *Robustness of the Black and Scholes formula*, Math. Finance, 8 (1998), pp. 93–126, https://doi.org/10.1111/1467-9965.00047.
[22] N. El Karoui, J.-P. Lepeltier, and A. Millet, *A probabilistic approach to the reduite in optimal stopping*, Probab. Math. Statist., 13 (1992), pp. 97–121.
[23] P. A. Forsyth and K. R. Vetzal, *Quadratic convergence for valuing American options using a penalty method*, SIAM J. Sci. Comput., 23 (2002), pp. 2095–2122, https://doi.org/10.1137/S1064827500382324.
[24] A. Friedman, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, NJ, 1964.
[25] P. V. Gapeev, *Pricing of perpetual American options in a model with partial information*, Int. J. Theor. Appl. Finance, 15 (2012), pp. 1–22, https://doi.org/10.1142/S0219024911006450.
[26] P. V. Gapeev and A. N. Shiryaev, *Bayesian quickest detection problems for some diffusion processes*, Adv. in Appl. Probab., 45 (2013), pp. 164–185, http://projecteuclid.org/euclid.aap/1363354107.
[27] M. Grasselli and V. Henderson, *Risk aversion and block exercise of executive stock options*, J. Econom. Dynam. Control, 33 (2009), pp. 109–127, https://doi.org/10.1016/j.jedc.2008.05.005.
[28] X. Guo and Q. Zhang, *Closed-form solutions for perpetual American put options with regime switching*, SIAM J. Appl. Math., 64 (2004), pp. 2034–2049, https://doi.org/10.1137/S0036139903426083.
[29] S. Huddart and M. Lang, *Information distribution within firms: Evidence from stock option exercises*, J. Accounting Econom., 34 (2003), pp. 3–31.

[30] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Grad. Texts in Math. 113, Springer-Verlag, New York, 1991, https://doi.org/10.1007/978-1-4612-0949-2.
[31] I. Karatzas and S. E. Shreve, *Methods of Mathematical Finance*, Appl. Math. (N.Y.) 39, Springer-Verlag, New York, 1998, https://doi.org/10.1007/b98840.
[32] M. Klein, *Comment on "Investment timing under incomplete information,"* Math. Oper. Res., 34 (2009), pp. 249–254, https://doi.org/10.1287/moor.1080.0349.
[33] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge Stud. Adv. Math. 24, Cambridge University Press, Cambridge, UK, 1997; reprint of the 1990 original.
[34] H. Le and C. Wang, *A finite time horizon optimal stopping problem with regime switching*, SIAM J. Control Optim., 48 (2010), pp. 5193–5213, https://doi.org/10.1137/080739550.
[35] T. Leung and R. Sircar, *Accounting for risk aversion, vesting, job termination risk and multiple exercises in valuation of employee stock options*, Math. Finance, 19 (2009), pp. 99–128, https://doi.org/10.1111/j.1467-9965.2008.00359.x.
[36] T. Leung and R. Sircar, *Exponential hedging with optimal stopping and application to employee stock option valuation*, SIAM J. Control Optim., 48 (2009), pp. 1422–1451, https://doi.org/10.1137/080718930.
[37] T. Leung, R. Sircar, and T. Zariphopoulou, *Forward indifference valuation of American options*, Stochastics, 84 (2012), pp. 741–770, https://doi.org/10.1080/17442508.2012.694438.
[38] K. Ma and P. A. Forsyth, *An unconditionally monotone numerical scheme for the two-factor uncertain volatility model*, IMA J. Numer. Anal., 37 (2016), pp. 905–944.
[39] M. Monoyios, *Utility-based valuation and hedging of basis risk with partial information*, Appl. Math. Finance, 17 (2010), pp. 519–551, https://doi.org/10.1080/13504861003650883.
[40] M. Monoyios and A. Ng, *Optimal exercise of an executive stock option by an insider*, Int. J. Theor. Appl. Finance, 14 (2011), pp. 83–106, https://doi.org/10.1142/S0219024911006279.
[41] G. Peskir and A. Shiryaev, *Optimal Stopping and Free-Boundary Problems*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2006.
[42] C. Reisinger, *The non-locality of Markov chain approximations to two-dimensional diffusions*, Math. Comp. Simul., 143 (2018), pp. 176–185.
[43] C. Reisinger and J. Rotaetxe Arto, *Boundary treatment and multigrid preconditioning for semi-Lagrangian schemes applied to Hamilton–Jacobi–Bellman equations*, J. Sci. Comput., 72 (2017), pp. 198–230.
[44] C. Reisinger and A. Whitley, *The impact of a natural time change on the convergence of the Crank–Nicolson scheme*, IMA J. Numer. Anal., 34 (2014), pp. 1156–1192.
[45] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and Martingales. Vol. 2: Itô Calculus*, Cambridge University Press, Cambridge, UK, 2000, https://doi.org/10.1017/CBO9781107590120; reprint of the second (1994) edition.
[46] N. Touzi, *American options exercise boundary when the volatility changes randomly*, Appl. Math. Optim., 39 (1999), pp. 411–422, https://doi.org/10.1007/s002459900112.