

# EXECUTIVE STOCK OPTION EXERCISE WITH FULL AND PARTIAL INFORMATION ON A DRIFT CHANGE POINT

VICKY HENDERSON, KAMIL KLADÍVKO, AND MICHAEL MONOYIOS

**ABSTRACT.** We analyse the valuation and exercise of an American executive call option written on a stock whose drift parameter falls to a lower value at a change point given by an exponential 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 maximising 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. Our setup captures the position of an executive (insider) and employee (outsider), who receive executive stock options. The latter yields a model under the observation filtration  $\widehat{\mathbb{F}}$  where the drift process 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 problem has two spatial dimensions. We analyse and numerically solve to value the option for both agents and illustrate that the additional information of the insider can result in exercise patterns which exploit the information on the change point.

**Keywords:** optimal stopping, executive stock options, insider information, American options, Wonham filter.

**AMS Subject classifications:** 91G80, 93E11, 93E20.

## 1. INTRODUCTION

This paper examines the effect of varying information, concerning a change in the value of the drift parameter of a stock, on valuation and optimal exercise of an American executive call option on the stock. We consider two agents who have the opportunity to exercise a call option of strike  $K \geq 0$  on the stock. The drift of the stock switches from its initial value  $\mu_0$  to a lower value  $\mu_1 < \mu_0$  at an exponential random time (the change point), independent of the Brownian motion  $W$  driving the stock. The first agent (let us call him the *insider*) has “full information”. He can observe the change point process  $Y$  as well as the Brownian motion  $W$ , so his filtration,  $\mathbb{F}$ , is that generated by  $W, Y$ . The second agent, let us call her the *outsider*, has “partial information”. She cannot observe  $Y$  and must therefore filter the change point from stock price observations. The outsider’s filtration is thus the stock price filtration  $\widehat{\mathbb{F}} \subset \mathbb{F}$ . With these two information structures, we examine the valuation and optimal exercise of the option. Each agent is barred from trading the stock, and maximises the discounted expected payoff of the option under the physical measure  $\mathbb{P}$  at the exercise time, which will be a stopping time of the agent’s filtration.

Our motivation for studying this problem is to examine the extent to which possession of privileged information on the performance of a stock can influence the exercise of an executive stock option (ESO). Executives often receive American call options on

---

*Date:* October 29, 2018.

Kamil Kladívko would like to thank the Norwegian School of Economics for providing him a pleasant and stimulating PhD environment. His contribution to the paper is a revised and extended version of Chapter 2 of his PhD thesis.

the stock of their company as part of their compensation package. Several empirical studies examine factors affecting the exercise of ESOs (Carpenter, Stanton and Wallace [11]) and a number consider whether executives time their option exercise based on inside information. Early studies uncover some evidence that executives use private information (Huddart and Lang [31], Carpenter and Remmers [10]) but more recent works that partition exercises based on the exercise strategy employed find much stronger evidence of informed exercise (Cicero [12], Aboody et al [1]). Exercises accompanied by a sale of stock are followed by negative abnormal returns (whilst other exercises are not). Brooks, Chance and Cline [7] also test whether moneyness is a distinguishing factor and find evidence that lower moneyness options show the strongest negative performance after exercise. We are interested in whether empirically observed ESO exercise patterns, such as exercise prior to poor stock performance, can be generated by the differential information in our model.

Executives are constrained in their ability to trade the stock which underlies their options, essentially rendering the stock not tradeable (to them) (Kulatilaka and Marcus [39], Carpenter [9], Detemple and Sundaresan [17] and Hall and Murphy [29]). Thus they face incomplete markets and in this paper we use the simplest potential pricing measure, the physical measure  $\mathbb{P}$  in order to evaluate the worth of the option grant to the executive.

We thus have two separate ESO optimal stopping problems to solve. The full information problem has similarities with papers on American option valuation with regime switching, such as Guo and Zhang [28] (in an infinite horizon case), and Buffington and Elliott [8] and Le and Wang [41] (in the finite horizon case). Numerical methods for such problems have also received a fair amount of attention (see Jobert and Rogers [33] or Khaliq and Liu [36], among others). Our problem is slightly different in that only one switch is allowed, and we characterise the solution via a free boundary problem, using largely known results.

In the partial information scenario, much less is known about the problem. After filtering, we arrive at a model where the stock's drift depends on a second diffusion process, and the problem is considerably more difficult. American option problems with a similar information structure have been studied by Gapeev [25] in an infinite horizon setting, but the finite horizon solution and numerical results are not available to the best of our knowledge. Detemple and Tian [18] have studied American option valuation in a diffusion setting (focusing on an early exercise decomposition result), while Touzi [48] has analysed the problem in a stochastic volatility model via a variational inequality approach to derive the smooth fit condition. Décamps et al [15, 16], Klein [38] and Ekström and Lu [20] have studied 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. These papers are able to reduce the dimensionality of the problem under some circumstances, due to the absence of an explicit change point (their models correspond to the limit that the parameter of the exponential time approaches zero) and due to the rather simpler objective functional they use. Also related is the trading model of Dai et al [14], which also involves a linear reward where dimensionality can be reduced. Such a simplification is not available in our model (we discuss this further in Sections 4 and 5), and the partial information ESO problem with finite horizon is three-dimensional, making it a challenging numerical problem. We give a complete characterisation of the partial information ESO value from a free boundary problem perspective, including a derivation of the smooth pasting property, which in turn requires monotonicity of the value function in the filtered estimate of  $Y$  (this is established using some stochastic flow ideas). This is the first main contribution of the paper.

A second contribution is a numerical solution of both the full and (in particular) partial information problems in a discrete time setting using a binomial approximation. In the partial information case, the resulting tree for the filtered probability does not recombine, and we develop an approximation in the spirit of the work on Asian options of Hull and White [32], Barraquand and Pudet [4], or Klassen [37]. We compare option values and exercise patterns of the agents. This reveals that the additional information can indeed be exploited, adding substance to the notion that holders of ESOs are able to take advantage of their position inside a firm issuing ESOs as a remuneration tool. We also provide a demonstration of the convergence of our numerical algorithms.

ESOs have received a lot of attention in the literature, usually with a view to investigating the effects of certain contractual features of the ESO or of agent's risk attitudes on their valuation and exercise (see Leung and Sircar [42, 43] or Grasselli and Henderson [27], for example). The effect of inside information has received less attention in the theoretical models. Our third contribution is to provide a theoretical model focussing on the impact of inside information on option exercise and valuation. Monoyios and Ng [45] is an exception, where an insider who had advance knowledge of the future value of the stock was considered. There, the effects of information were indeed manifest in the exercise decision. In this paper, the form of the additional information is quite different, and rather weaker (thus more realistic) than direct knowledge of the future value of the stock.

The rest of the paper is organised as follows. We describe the model in Section 2, introduce the ESO problems under both full and partial information, and carry out a filtering procedure to derive the model dynamics with respect to the stock price filtration. In Section 3 we analyse the full information problem including deriving convexity and monotonicity properties of the value function for the free boundary problem satisfied by the ESO value. In Section 4 we analyse the partial information problem. We again establish convexity and monotonicity properties of the value function, and give a complete analysis of the free boundary problem, including the smooth pasting property. In Section 5 we construct and implement numerical solutions of the full and partial information problems, and perform simulations to compare the exercise patterns of the agents, which reveal that the insider can indeed exercise the ESO in a manner which exploits his privileged information. Depending on the stock price evolution and the change point, the insider can exercise the ESO in situations where the outsider does not, and vice versa. Section 6 concludes the paper.

## 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 (the “insider”) who observes the change point and a less informed agent (the “outsider”) who has to filter the change point from stock price observations. In other words, can the additional information be exploited in the exercise strategy?

We have a complete filtered probability space  $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \in \mathbf{T}}, \mathbb{P})$ . The time set  $\mathbf{T}$  will be the finite interval  $\mathbf{T} = [0, T]$ , for some  $T < \infty$ . The filtration  $\mathbb{F} := (\mathcal{F}_t)_{t \in \mathbf{T}}$  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 non-negative random time, independent of  $W$ , with initial distribution  $\mathbb{P}[\theta = 0] =: y_0 \in [0, 1]$  and

subsequent distribution

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

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

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

so that  $Y_0 = \mathbb{1}_{\{\theta=0\}}$  with  $\mathbb{E}[Y_0] = y_0$ . We may take  $\mathbb{F}$  to be the  $\mathbb{P}$ -augmentation of  $\mathbb{F}^{W,Y}$ , the filtration generated by the pair  $W, Y$ .

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

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

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  $\mu_0$  to the lower value  $\mu_1$  at the change point. Define the constant  $\eta > 0$  by

$$(2.3) \quad \eta := \frac{\mu_0 - \mu_1}{\sigma}.$$

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

$$(2.4) \quad dX_t = (\mu_0 - \sigma\eta Y_t)X_t dt + \sigma X_t dW_t.$$

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

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

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 modelled, but we do not do so for simplicity. For example, the simplest is a constant dividend  $q$  which could be included with minor adjustments by re-interpreting the drifts as being net of dividends.

We may write the stock price evolution as

$$(2.5) \quad dX_t = \sigma X_t d\xi_t,$$

where  $\xi$  is the volatility-scaled return process given by

$$(2.6) \quad \xi_t := \frac{1}{\sigma} \int_0^t \frac{dX_s}{X_s} = \left(\frac{\mu_0}{\sigma}\right)t - \eta \int_0^t Y_s ds + W_t =: \int_0^t h_s ds + W_t, \quad t \in \mathbf{T},$$

with the process  $h$  defined by

$$(2.7) \quad h_t := \frac{\mu_0}{\sigma} - \eta Y_t, \quad t \in \mathbf{T},$$

so  $h$  and  $W$  are independent. The process  $\xi$  will be used as an observation process in a filtering algorithm in Section 2.2.

Define the *observation filtration*  $\widehat{\mathbb{F}} = (\widehat{\mathcal{F}}_t)_{t \in \mathbf{T}}$  by

$$\widehat{\mathcal{F}}_t \equiv \mathcal{F}_t^X := \sigma(X_s : 0 \leq s \leq t), \quad t \in \mathbf{T},$$

so that  $\widehat{\mathbb{F}}$  is the filtration generated by the stock price, or equivalently by the process  $\xi$  in (2.6), and we have  $\widehat{\mathbb{F}} \subset \mathbb{F}$ .

An executive stock option (ESO) on  $X$  is an American call option with strike  $K \geq 0$  and maturity  $T$ , so has payoff  $(X_t - K)^+$  if exercised at  $t \in \mathbf{T}$ . We assume the executive receives the cash payoff on exercise as this is both the most common type and the most relevant for the study of private information.

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. Both agents are prohibited from trading  $X$  (think of them as employees of the company whose share price is  $X$ ). For simplicity, we shall assume there are no other trading opportunities for these agents, and there are no other contractual features of the ESO, such as a vesting period or partial exercise opportunities. This is so we can focus exclusively on the effect of the different information sets of the two agents.

The first agent (the *insider*) 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 (the *outsider*) 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 outsider's filtration is therefore the observation filtration  $\widehat{\mathbb{F}}$ , and the only difference between the agents is that the outsider does not know the value of the process  $Y$ , which she will filter from stock price observations.

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 outsider 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 outsider could be rendered ignorant of the values  $\mu_0, \mu_1$ , so these could be modelled (for example) as random variables whose values would be filtered from price observations. However, our constant volatility model is the simplest one can envisage with differential information on a change point.

**2.1. The ESO optimal stopping problems.** Each agent will maximise, over stopping times of their respective filtration, the discounted expectation of the ESO payoff under the physical measure  $\mathbb{P}$ . 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 a general starting time  $t \in [0, T]$ , the insider's ESO value process is  $V$ , an  $\mathbb{F}$ -adapted process defined by

$$(2.8) \quad V_t := \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[ e^{-r(\tau-t)} (X_\tau - K)^+ \middle| \mathcal{F}_t \right], \quad t \in [0, T].$$

We shall call (2.8) the full information problem.

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

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

We shall call (2.9) the partial information problem.

Naturally, the salient distinction between (2.8) and (2.9) is the filtration with respect to which the stopping time and essential supremum are defined. For the full information problem (2.8) the stock dynamics will be (2.4). For the partial information problem (2.9) we must derive the model dynamics under the observation filtration. This is done in Section 2.2 below.

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 [46] (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 will be transmitted through the prism of the ESO exercise decision. In contrast, the classical change point detection problem has some explicit objective functional which directly penalises 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  $\xi$  in (2.6), with the filtration generated by  $\xi$  equivalent to the 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  $\phi$ . 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 2.1** (Observation filtration dynamics). *With respect to the observation filtration  $\widehat{\mathbb{F}}$  the stock price follows*

$$(2.10) \quad dX_t = (\mu_0 - \sigma\eta\widehat{Y}_t)X_t dt + \sigma X_t d\widehat{W}_t,$$

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

$$(2.11) \quad \widehat{W}_t := \xi_t - \int_0^t \widehat{h}_s ds = \xi_t - \frac{\mu_0}{\sigma}t + \eta \int_0^t \widehat{Y}_s ds, \quad t \in \mathbf{T},$$

where analogously to (2.7),  $\widehat{h}_t := \frac{\mu_0}{\sigma} - \eta\widehat{Y}_t$ ,  $t \in \mathbf{T}$ , and  $\widehat{W}$  is a  $(\mathbb{P}, \widehat{\mathbb{F}})$ -Brownian motion.

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

$$(2.12) \quad d\widehat{Y}_t = \lambda(1 - \widehat{Y}_t) dt - \eta\widehat{Y}_t(1 - \widehat{Y}_t) d\widehat{W}_t, \quad \widehat{Y}_0 = \mathbb{E}[Y_0] = y_0 \in [0, 1].$$

*Proof.* We use the innovations approach to filtering, as discussed in Rogers and Williams [47], Chapter VI.8 or Bain and Crisan [2], Chapter 3, for instance.

By Theorem VI.8.4 in [47], the innovations process  $\widehat{W}$ , defined by (2.11), is a  $(\mathbb{P}, \widehat{\mathbb{F}})$ -Brownian motion. Using (2.11) in the stock price SDE (2.5) then yields (2.10).

It remains to prove (2.12). 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  $\mathbb{E} \left[ \int_0^t |\mathcal{G}f_s|^2 ds \right] < \infty$  for all  $t \in \mathbf{T}$ , such that

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

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)

$$(2.13) \quad \widehat{f}_t = \widehat{f}_0 + \int_0^t \widehat{\mathcal{G}}f_s ds + \int_0^t (\widehat{f_s h_s} - \widehat{f_s} \widehat{h_s}) d\widehat{W}_s, \quad t \in \mathbf{T}.$$

Take  $f(y) = y$ . Then the martingale  $M^{(f)} = M$ , as defined in (2.2), so that  $\mathcal{G}f = \lambda(1 - Y)$  and the filtering equation (2.13) reads as

$$(2.14) \quad \widehat{Y}_t = y_0 + \lambda \int_0^t (1 - \widehat{Y}_s) ds + \int_0^t (\widehat{Y_s h_s} - \widehat{Y_s} \widehat{h_s}) d\widehat{W}_s, \quad t \in \mathbf{T},$$

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

Now,

$$(2.15) \quad \widehat{Y}_t \widehat{h}_t = \mathbb{E} \left[ Y_t \left( \frac{\mu_0}{\sigma} - \eta Y_t \right) \middle| \widehat{\mathcal{F}}_t \right] = \left( \frac{\mu_0}{\sigma} \right) \widehat{Y}_t - \eta \mathbb{E}[Y_t^2 | \widehat{\mathcal{F}}_t] = \left( \frac{\mu_0}{\sigma} - \eta \right) \widehat{Y}_t, \quad t \in \mathbf{T},$$

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

On the other hand,

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

Using (2.15) and (2.16) in (2.14) then yields the integral form of (2.12).  $\square$

Note that  $\widehat{Y}$  in (2.12) is a diffusion in  $[0, 1]$  with an absorbing state at  $\widehat{Y} = 1$ .

### 3. THE FULL INFORMATION ESO PROBLEM

In this section we focus on the full information problem defined in (2.8). Define the reward process  $R$  as the discounted payoff process:

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

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

$$(3.2) \quad \widetilde{V}_t := e^{-rt} V_t = \operatorname{ess\,sup}_{t \in \mathcal{T}_{t,T}} \mathbb{E}[R_\tau | \mathcal{F}_t], \quad t \in \mathbf{T}.$$

Classical optimal stopping theory (see for example Appendix D of Karatzas and Shreve [35]) characterises the solution to the problem (2.8) in terms of the Snell envelope of  $R$ , the smallest non-negative càdlàg  $(\mathbb{P}, \mathbb{F})$ -super-martingale  $\widetilde{V}$  that dominates  $R$ , with  $\widetilde{V}_T = R_T$  almost surely, and hence  $V_T = (X_T - K)^+$ . A stopping time  $\bar{\tau} \in \mathcal{T}$  is optimal for the problem (2.8) starting at time zero if and only if  $\widetilde{V}_{\bar{\tau}} = R_{\bar{\tau}}$  almost surely (so  $V_{\bar{\tau}} = (X_{\bar{\tau}} - K)^+$ ), and if and only if the stopped super-martingale  $\widetilde{V}^{\bar{\tau}}$  defined by  $\widetilde{V}_t^{\bar{\tau}} := \widetilde{V}_{\bar{\tau} \wedge t}$ ,  $t \in \mathbf{T}$ , is a  $(\mathbb{P}, \mathbb{F})$ -martingale. The smallest optimal stopping time in  $\mathcal{T}_{t,T}$  for the problem (2.8) is  $\bar{\tau}(t)$ , the first time that the discounted ESO value process coincides with the reward, so is given by

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

**3.1. Full information value function.** Introduce the value function  $v : [0, T] \times \mathbb{R}_+ \times \{0, 1\} \rightarrow \mathbb{R}_+$  for the full information optimal stopping problem (2.8) as

$$(3.3) \quad v(t, x, i) := \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[ e^{-r(\tau-t)} (X_\tau - K)^+ \middle| X_t = x, Y_t = i \right], \quad i = 0, 1, \quad t \in [0, T],$$

and write  $v_i(t, x) \equiv v(t, x, i)$ ,  $i = 0, 1$ . 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 (2.8) has the representation

$$(3.4) \quad V_t = v(t, X_t, Y_t) = (1 - Y_t)v_0(t, X_t) + Y_t v_1(t, X_t), \quad t \in [0, T].$$

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

American option problems with regime-switching parameters have been studied fairly extensively, usually in a classical set-up where the optimal stopping problem is formulated under a martingale measure for the stock, and with the Markov switching process allowed to switch back and forth between regimes. In our case, only one switch is allowed, but where this does not materially affect the proofs of certain properties given in the literature, we may (and shall) take the resulting features of the value function as given. Although our problem is formulated under the physical measure  $\mathbb{P}$ , one can formally map our case to the conventional scenario under a martingale measure by setting the stock drift  $\mu(\cdot)$  equal to the interest rate minus a fictitious “dividend yield”  $q(\cdot)$ , so  $\mu(\cdot) \equiv r - q(\cdot)$ . An infinite horizon problem with multiple regime switching was studied by Guo and Zhang [28], who obtained closed form solutions, and also by Gapeev [25] (whose primary focus was the partial information case) in a slightly different context, with a dividend rate switching between different values. The finite horizon problem was treated by Buffington and Elliott [8], who derived approximate solutions in the manner of Barone-Adesi and Elliott [3], and by Le and Wang [41], who gave a rigorous treatment of the smooth pasting property that was absent from [8]. We can therefore take some properties of the value function as given, where they have been proven in earlier work. We shall establish some elementary properties in Lemma 3.1 below that pertain to our situation, with a call payoff as opposed to a put, and with only one switch.

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

$$H_{s,t} := \exp \left\{ \left( \mu_0 - \frac{1}{2} \sigma^2 \right) (t - s) - \sigma \eta \int_s^t Y_u du + \sigma (W_t - W_s) \right\}, \quad 0 \leq s \leq t \leq T,$$

so that given  $X_s = x \in \mathbb{R}_+$ , the stock price at  $t \in [s, T]$  is

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

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

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

For use further below, also define

$$H_{s,t}^{(i)} := \exp \left\{ \left( \mu_i - \frac{1}{2} \sigma^2 \right) (t - s) + \sigma (W_t - W_s) \right\}, \quad 0 \leq s \leq t \leq T, \quad i = 0, 1.$$

The value function in (3.3) is then given as

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

Using stationarity of the Brownian increments and the absence of memory property of the exponential distribution, optimising over  $\mathcal{T}_{t,T}$  is equivalent to optimising over  $\mathcal{T}_{0,T-t}$ , so the value function may be re-cast into the form

$$(3.5) \quad v_i(t, x) = \sup_{\tau \in \mathcal{T}_{0,T-t}} \mathbb{E} \left[ e^{-r\tau} (x H_\tau - K)^+ \mid Y_0 = i \right], \quad (t, x) \in [0, T] \times \mathbb{R}_+, \quad i = 0, 1.$$

The following lemma gives the elementary properties of the 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 (3.3) characterising the full information ESO value function (and the ESO value process via (3.4)) have the following properties:*

- (1) *For  $i = 0, 1$  and  $t \in [0, T]$ , the map  $x \rightarrow v_i(t, x)$  is convex and non-decreasing.*
- (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  $t \rightarrow v_i(t, x)$  is non-increasing.*



- Proof.* (1) Convexity and monotonicity of the map  $x \rightarrow v_i(t, x)$  are immediate from the representation (3.5), the linearity of the map  $x \rightarrow X_\tau^x = xH_\tau$ , along with the convexity and monotonicity properties of the payoff function  $x \rightarrow (x - K)^+$ .
- (2) If  $Y_t = 0$  (so  $\theta > t$ ) then for any  $\tau \in \mathcal{T}_{t,T}$  we have  $\int_t^\tau Y_s ds = (\tau - \theta)\mathbb{1}_{\{\tau \geq \theta\}} \leq \tau - t$ , and hence

$$\begin{aligned}
v_0(t, x) &= \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[ e^{-r(\tau-t)} \left( xH_{t,\tau}^{(0)} \exp \left( -\sigma\eta \int_t^\tau Y_s ds \right) - K \right)^+ \middle| Y_t = 0 \right] \\
&\geq \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[ e^{-r(\tau-t)} \left( xH_{t,\tau}^{(0)} \exp(-\sigma\eta(\tau-t)) - K \right)^+ \right] \\
&= \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[ e^{-r(\tau-t)} (xH_{t,\tau}^{(1)} - K)^+ \right] \\
&= v_1(t, x).
\end{aligned}$$

- (3) This is immediate from the representation (3.5) and the fact that  $\mathcal{T}_{0,T-t'} \subseteq \mathcal{T}_{0,T-t}$  for  $t' \geq t$ .  $\square$

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

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

Since the functions  $v_i(\cdot, \cdot)$  are continuous, the continuation regions  $\mathcal{C}_i$  are open sets.

*Remark 3.2* (Minimal conditions for early exercise: full information). If the drift process  $\mu \equiv \mu(Y)$  of the stock satisfies  $\mu \geq r$  almost surely, then the reward process is a  $(\mathbb{P}, \mathbb{F})$ -sub-martingale, so no early exercise is optimal, and the American ESO value coincides with that of its European counterpart. In particular, if  $\mu_0 \geq r$ , then we expect no early exercise when  $Y = 0$  (so 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 non-increasing critical stock price function (or exercise boundary)  $x_i : [0, T] \rightarrow [K, \infty)$  where  $x_0(t) > x_1(t) > K$  for all  $t \in [0, T]$ . The terminal values  $x_i(T)$  can be found via standard PDE arguments and are stated in the corollary. 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 stock price satisfies  $x_1(\theta) \leq X_\theta < x_0(\theta)$ , 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. We formalise these properties in the corollary below.

**Corollary 3.3.** *There exist two non-increasing functions  $x_i : [0, T] \rightarrow [K, \infty)$ ,  $i = 0, 1$ , satisfying*

$$x_i(T) = \max \left( K; \frac{r}{r - \mu_i} K \right)$$

as well as

$$x_1(t) < x_0(t), \quad t \in [0, T],$$

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

$$\begin{aligned}
\mathcal{C}_i &= \{(t, x) \in [0, T] \times \mathbb{R}_+ : x < x_i(t)\}, \quad i = 0, 1, \\
\mathcal{S}_i &= \{(t, x) \in [0, T] \times \mathbb{R}_+ : x \geq x_i(t)\}, \quad i = 0, 1.
\end{aligned}$$

The smallest optimal stopping time  $\bar{\tau}$  for the full information problem (2.8) starting at time zero is  $\bar{\tau}(0) \equiv \bar{\tau}$ , given by

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

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.

When  $\mu_i < r$ ,  $i = 0, 1$ , so that bounded exercise thresholds exist prior to maturity, it is not hard to proceed along fairly classical lines to show that the exercise boundaries are in fact continuous, using methods similar to those in Karatzas and Shreve [35], Section 2.7, or Peskir and Shiryaev [46], Section VII.25.2, but we shall not pursue this here, in the interests of brevity. We thus move directly to the free boundary characterisation 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 the proposition below. This is essentially well-known, due to Guo and Zhang [28] in the infinite horizon case, and to Le and Wang [41] in the finite horizon case. These works used a multiple regime switching model, but the proofs go through largely unaltered. In the partial information case, however, we shall give a full proof (see the proof of Proposition 4.5) as we have not found a rigorous demonstration in earlier papers.

**Proposition 3.4** (Free boundary problem: full information). *The full information value function  $v(t, x, i) \equiv v_i(t, x)$ ,  $i = 0, 1$ , defined in (3.3) is the unique solution in  $[0, T] \times \mathbb{R}_+ \times \{0, 1\}$  of the free boundary problem*

$$\begin{aligned} \mathcal{L}_0 v_0(t, x) &= -\lambda(v_1(t, x) - v_0(t, x)), \quad 0 \leq x < x_0(t), \quad t \in [0, T), \\ \mathcal{L}_1 v_1(t, x) &= 0, \quad 0 \leq x < x_1(t), \quad t \in [0, T), \\ v_i(t, x) &= x - K, \quad x \geq x_i(t), \quad t \in [0, T), \quad i = 0, 1, \\ v_i(T, x) &= (x - K)^+, \quad x \in \mathbb{R}_+, \quad i = 0, 1, \\ \lim_{x \downarrow 0} v_i(t, x) &= 0, \quad t \in [0, T), \quad i = 0, 1. \end{aligned}$$

*Proof.* This follows similar reasoning to the proof of Theorem 3.1 of Guo and Zhang [28] (modified to take into account the time-dependence of the functions  $v_i(\cdot, \cdot)$ ,  $i = 0, 1$ ). Alternatively, one can adapt the proof of Proposition 1 in Le and Wang [41].  $\square$

Proposition 3.4 shows that for  $i = 0, 1$ , each  $v_i(t, x)$  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  $x_i(t)$ . This is settled by the smooth pasting property below. This property has been established in Le and Wang [41] for a put option in a model with multiple regime switching.

**Lemma 3.5** (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.* This can be established along similar lines to the proof of Lemma 8 in Le and Wang [41]. A more direct demonstration along the lines of the proof of Lemma 2.7.8 in Karatzas and Shreve [35] is possible. We do not present it here, but it is similar in spirit to the proof of the smooth fit condition that we give for the partial information problem (see the proof of Theorem 4.6).  $\square$

#### 4. THE PARTIAL INFORMATION ESO PROBLEM

We now turn to the outsider's partial information problem (2.9), over  $\widehat{\mathbb{F}}$ -stopping times, with model dynamics given by Lemma 2.1. The partial information value function  $u : [0, T] \times \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}_+$  is defined by

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

and the ESO value process  $U$  in (2.9) is given as

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

With respect to  $\widehat{\mathbb{F}}$ , the dynamics of the two-dimensional diffusion  $(X, \widehat{Y})$  are given in (2.10) and (2.12). For  $0 \leq s \leq t \leq T$ , 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

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

so we have

$$(4.2) \quad X_t^{s,x,y} = x G_t^{s,y}, \quad 0 \leq s \leq t \leq T.$$

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

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

The partial information value function in (4.1) is thus

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

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

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

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  $\mu \equiv \mu(\widehat{Y})$  of the stock satisfies  $\mu \geq r$  almost surely, then the reward process is a  $(\mathbb{P}, \widehat{\mathbb{F}})$ -sub-martingale, 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) characterising the partial information ESO value function has the following properties.*

- (1) For  $(t, y) \in [0, T] \times [0, 1]$ , the map  $x \rightarrow u(t, x, y)$  is convex and non-decreasing.
- (2) For  $(t, x) \in [0, T] \times \mathbb{R}_+$ , the map  $y \rightarrow u(t, x, y)$  is non-increasing.

(3) For  $(x, y) \in \mathbb{R}_+ \times [0, 1]$ , the map  $t \rightarrow u(t, x, y)$  is non-increasing.

*Proof.* The proofs of the first and third properties are similar to the proofs of the corresponding properties for the full information case in Lemma 3.1, so are omitted. Let us focus therefore on the second claim.

In (4.3), the quantity  $G_\tau^y$  is the value at  $\tau \in \widehat{\mathcal{T}}_{0, T-t}$  of the process  $G^y$  given by

$$(4.4) \quad G_t^y := \exp \left( \left( \mu_0 - \frac{1}{2} \sigma^2 \right) t + \sigma \widehat{W}_t - \sigma \eta \int_0^t \widehat{Y}_s^y ds \right), \quad t \in [0, T].$$

From (4.4) and (4.3), the desired monotonicity of the map  $y \rightarrow u(t, x, y)$  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 non-decreasing with respect to  $y$ :

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

This property is shown in Proposition 4.3 further below, and this completes the proof.  $\square$

**4.1. The stochastic flow  $\widehat{Y}(y)$ .** Let us consider the solution to the SDE (2.12) for  $\widehat{Y}$  with initial condition  $\widehat{Y}_0 = y_0 \in [0, 1]$ . 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 [40], Chapter 4), we may choose versions of  $\widehat{Y}(y)$  which, for each  $t \in [0, T]$  and each  $\omega \in \Omega$ , are diffeomorphisms in  $y$  from  $[0, 1) \rightarrow [0, 1]$ . In other words, the map  $y \rightarrow \widehat{Y}(y)$  is smooth.

We wish to show the property (4.5). To achieve this, we shall look at the flow of the so-called likelihood ratio  $\Phi$ , defined by

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

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

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

where  $\mathcal{E}(\cdot)$  denotes the stochastic exponential, and  $(\widehat{Y} \cdot \widehat{W}) \equiv \int_0^\cdot \widehat{Y}_s d\widehat{W}_s$  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

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

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

$$(4.8) \quad dX_t = \mu_0 X_t dt + \sigma X_t dW_t^*,$$

$$(4.9) \quad d\Phi_t = \lambda(1 + \Phi_t) dt - \eta \Phi_t dW_t^*.$$

Equations (4.8) and (4.9) exhibit an interesting feature in that  $X$  and  $\Phi$  become decoupled under  $\mathbb{P}^*$ . Similar measure changes have been employed by Décamps et al [15, 16], Klein [38] and Ekström and Lu [20] 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 set-up, and both  $X$  and  $\Phi$  become geometric

Brownian motions with respect to  $(\mathbb{P}^*, \widehat{\mathbb{F}})$ , yielding an easier problem, in that  $\Phi$  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,  $\Phi$  depends on the entire history of the Brownian paths, as exhibited in equation (4.10) below. This, combined with the non-linear call payoff makes the aforementioned dimension reduction impossible, and this makes the numerical solution of the partial information ESO problem much more complex, involving a recombining stock price tree.

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  $\Phi$  by (4.6), and define the exponential  $(\mathbb{P}^*, \widehat{\mathbb{F}})$ -martingale  $\Lambda$  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*

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

*so that*

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

*Consequently, if  $\widehat{Y}(y)$  denotes the solution to (2.12) with initial condition  $\widehat{Y}_0 = y \neq 1$ , then*

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

*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). □

*Remark 4.4.* 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  $x^* : [0, T] \times [0, 1] \rightarrow [K, \infty)$ , 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

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

PDE arguments based on the free boundary characterisation of the value function in Proposition 4.5 below, can be employed to obtain the terminal value  $x^*(T, y)$  as:

$$x^*(T, y) = \max \left( K; \frac{r}{r - \mu_0 + (\mu_0 - \mu_1)y} K \right).$$

This terminal value can also be obtained by analogy with the standard American call terminal exercise boundary, where we set the stock drift  $\mu(\cdot)$  equal to the interest rate minus a fictitious “dividend yield”  $q(\cdot)$ , so  $\mu(\cdot) \equiv r - q(\cdot)$  as mentioned in Section 3.1 in the context of the full information problem.

American option valuation for systems governed by two-dimensional diffusion processes have been considered by Detemple and Tian [18], whose Proposition 2 shows that the continuation and stopping regions are indeed characterised by a stock price threshold. The additional feature here is some indication of the shape of these regions due to the monotonicity with respect to  $y$  and time. It is possible to go further and establish that the exercise boundary is continuous, using ideas similar to those in Karatzas and Shreve [35], Section 2.7, or Peskir and Shiryaev [46], Section VII.25.2. As in the full information case, we shall not pursue this. We thus move directly to the free boundary characterisation of the partial information value function.

Let  $\mathcal{L}_{X, \hat{Y}}$  denote the generator under  $\mathbb{P}$  of the two-dimensional process  $(X, \hat{Y})$  with respect to the observation filtration  $\hat{\mathbb{F}}$ , with dynamics given by (2.10) and (2.12). Thus,  $\mathcal{L}_{X, \hat{Y}}$  is defined by

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

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

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

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

**Proposition 4.5** (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*

$$(4.13) \quad \mathcal{L}u(t, x, y) = 0, \quad 0 \leq x < x^*(t, y), \quad t \in [0, T], \quad y \in [0, 1],$$

$$(4.14) \quad u(t, x, y) = x - K, \quad x \geq x^*(t, y), \quad t \in [0, T], \quad y \in [0, 1],$$

$$(4.15) \quad u(T, x, y) = (x - K)^+, \quad x \in \mathbb{R}_+, \quad y \in [0, 1],$$

$$(4.16) \quad \lim_{x \downarrow 0} u(t, x, y) = 0, \quad t \in [0, T], \quad y \in [0, 1].$$

*Proof.* It is clear that  $u$  satisfies the boundary conditions (4.14), (4.15) and (4.16). To verify (4.13), take a point  $(t, x, y) \in \hat{\mathcal{C}}$  (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  $(t, x, y) \in \mathcal{R} \subset \hat{\mathcal{C}}$ . Let  $\partial \mathcal{R}$  denote the boundary of this region, and let  $\partial_0 \mathcal{R} := \partial \mathcal{R} \setminus (\{t_{\max}\} \times (x_{\min}, x_{\max}) \times (y_{\min}, y_{\max}))$  denote the so-called parabolic boundary of  $\mathcal{R}$ . Consider the terminal-boundary value problem

$$(4.17) \quad \mathcal{L}f = 0 \quad \text{in } \mathcal{R}, \quad f = u \quad \text{on } \partial_0 \mathcal{R}.$$

Classical theory for parabolic PDEs (for instance, Friedman [24]) guarantees the existence of a unique solution to (4.17) 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

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

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

$$N_{\rho} := e^{-r\rho} f(t + \rho, xG_{\rho}^y, \widehat{Y}_{\rho}^y), \quad 0 \leq \rho \leq t_{\max} - t.$$

The stopped process  $(N_{\rho \wedge \tau})_{0 \leq \rho \leq t_{\max}-t}$  is a martingale by virtue of the Itô formula and the system (4.17) satisfied by  $f$ , and hence

$$(4.18) \quad f(t, x, y) = N_t = \mathbb{E}[N_{\tau}] = \mathbb{E}[e^{-r\tau} u(t + \tau, xG_{\tau}^y, \widehat{Y}_{\tau}^y)],$$

where we have used the boundary condition in (4.17) to obtain the last equality.

Since  $\mathcal{R} \subset \widehat{\mathcal{C}}$ ,  $(t + \tau, xG_{\tau}^y, \widehat{Y}_{\tau}^y) \in \widehat{\mathcal{C}}$ , so  $\tau$  must satisfy

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

In other words,  $\tau$  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

$$e^{-r(\rho \wedge \tau^*(t, x, y))} u\left(t + (\rho \wedge \tau^*(t, x, y)), xG_{\rho \wedge \tau^*(t, x, y)}^y, \widehat{Y}_{\rho \wedge \tau^*(t, x, y)}^y\right), \quad 0 \leq \rho \leq T - t,$$

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

$$(4.19) \quad \mathbb{E}\left[e^{-r\tau} u(t + \tau, xG_{\tau}^y, \widehat{Y}_{\tau}^y)\right] = u(t, x, y),$$

and (4.18) and (4.19) 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 (4.13).

Finally, to show uniqueness, let  $g$  defined on the closure of  $\widehat{\mathcal{C}}$  be a solution to the system (4.13)–(4.16). For starting state  $(0, x, y)$  such that  $x < x^*(0, y)$  define

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

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

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

so that the solution is unique. □

**4.2.1. Smooth fit condition.** We have the smooth pasting property below. It is natural to expect this property to hold, but to the best of our knowledge has not been established before in a model such as ours. In stochastic volatility models, Touzi [48] has used variational inequality techniques to show the smooth pasting property. This method can probably be adapted to our setting, but we shall employ a method more akin to the classical proof of smooth fit in American option problems, along similar lines to Lemma 2.7.8 in Karatzas and Shreve [35] or Theorem 3.4 in Monoyios and Ng [45]. We will utilise the measure  $\mathbb{P}^*$  defined in (4.7) to simplify the problem.

**Theorem 4.6** (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  $x \rightarrow u(x, y)$  is convex and non-decreasing, so we have  $u_x(x, y) \leq 1$  in the continuation region  $\hat{\mathcal{C}} = \{(x, y) \in \mathbb{R}_+ \times [0, 1] : x < x^*(y)\}$ , and thus  $u_x(x^*(y)-, y) \leq 1$ . We also have  $u_x(x, y) = 1$  in the stopping region  $\hat{\mathcal{S}} = \{(x, y) \in \mathbb{R}_+ \times [0, 1] : x \geq x^*(y)\}$ , and thus  $u_x(x^*(y)+, y) = 1$ . Hence, the proof will be complete if we can show that  $u_x(x^*(y)-, y) \geq 1$ .

Under measure  $\mathbb{P}^*$  defined in (4.7), the dynamics of  $X$  with respect to  $(\mathbb{P}^*, \hat{\mathbb{F}})$  are given by

$$dX_t = \mu_0 X_t dt + \sigma X_t dW_t^*,$$

and equivalently, the process  $G^y$  is given by

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

For any  $(x, y) \in \mathbb{R}_+ \times (0, 1)$ , let  $\tau^*(0, x, y) \equiv \tau^*(x, y)$  denote the optimal stopping time for  $u(x, y)$ , so that

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

Set  $x = x^*(y)$ , which will be fixed for the remainder of the proof. For  $\epsilon > 0$ , since the exercise boundary is non-increasing in time and in  $y$ , we have

$$(4.20) \quad \tau^*(x - \epsilon, y) \leq \inf \{t \in [0, T] : (x - \epsilon) G_t^y \geq x\} \wedge T.$$

The Law of the Iterated Logarithm for the Brownian motion  $W^*$  (Karatzas and Shreve [34], Theorem 2.9.23) implies that

$$\sup_{0 \leq t \leq \rho} G_t^y > 1, \quad \mathbb{P}^*\text{-a.s.}$$

for every  $\rho > 0$ . Hence there exists a sufficiently small  $\epsilon > 0$  such that

$$\sup_{0 \leq t \leq \rho} (x - \epsilon) G_t^y \geq x, \quad \mathbb{P}^*\text{-a.s.}$$

for every  $\rho > 0$ . Thus the right-hand-side of (4.20) tends to zero as  $\epsilon \downarrow 0$  and we have

$$(4.21) \quad \lim_{\epsilon \downarrow 0} \tau^*(x - \epsilon, y) = 0, \quad \mathbb{P}^*\text{-a.s.}$$

Since  $\mathbb{P}^* \sim \mathbb{P}$ , this is also true  $\mathbb{P}$ -a.s. Using the fact that  $\tau^*(x - \epsilon, y)$  will be sub-optimal for starting state  $(x, y)$ , we have

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



We now take the limit as  $\epsilon \downarrow 0$ . Using (4.21) and the fact that it is never optimal to exercise when the stock price is below the strike, we have

$$\lim_{\epsilon \downarrow 0} \mathbb{1}_{\{(x-\epsilon)G_{\tau^*}^y(x-\epsilon, y) \geq K\}} = 1, \quad \text{a.s.}$$

and we also have

$$\lim_{\epsilon \downarrow 0} G_{\tau^*}^y(x-\epsilon, y) = 1, \quad \text{a.s.}$$

Using these properties as well as the uniform integrability of  $(G_t^y)_{t \in [0, T]}$ , we obtain

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

which completes the proof.  $\square$

With the smooth pasting property in place, it is feasible to apply the Itô formula to the ESO value function and derive an early exercise decomposition for the value function, and an associated integral equation for the exercise boundary, in the manner of Theorem 2.7.9 and Corollary 2.7.11 of Karatzas and Shreve [35], or Theorem 3.5 and Corollary 3.1 of Monoyios and Ng [45]. We shall not go down this route here, instead we shall solve the ESO free boundary problem directly via a numerical scheme, in Section 5.

## 5. NUMERICAL ALGORITHM AND SIMULATIONS

This section is devoted to numerical solution of the ESO problems. In principle one could resort to finite difference solutions of the differential equations of Sections 3 and 4 (the latter would be intensive due to the three-dimensional free boundary problem of Proposition 4.5). Instead, we shall develop a binomial scheme and (in the partial information case) an associated discrete time filter (which is interesting in its own right). First we illustrate the inherent complexity of the partial information case, due to its path-dependent structure.

**5.1. A change of state variable.** Consider the partial information problem (2.9). We shall change measure to  $\mathbb{P}^*$  defined in (4.7), and this naturally leads to a change of state variable from  $(X, \hat{Y})$  to  $(X, \Phi)$ , with  $\Phi$  defined in (4.6). This leads to the following lemma.

**Lemma 5.1.** *Let  $\Phi$  be the likelihood ratio process defined in (4.6). The partial information ESO value process  $U$  in (2.9) satisfies*

$$(5.1) \quad e^{-(r+\lambda)t} (1 + \Phi_t) U_t = \operatorname{ess\,sup}_{\tau \in \hat{\mathcal{T}}_{t,T}} \mathbb{E}^* \left[ e^{-(r+\lambda)\tau} (1 + \Phi_\tau) (X_\tau - K)^+ | \hat{\mathcal{F}}_t \right], \quad t \in [0, T],$$

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

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

$$(5.2) \quad Z_t := \frac{1}{\Gamma_t} = \frac{d\mathbb{P}}{d\mathbb{P}^*} \Big|_{\hat{\mathcal{F}}_t} = \mathcal{E}(-\eta \hat{Y} \cdot W^*)_t, \quad t \in [0, T],$$

satisfying

$$(5.3) \quad dZ_t = -\eta \hat{Y}_t Z_t dW_t^*, \quad Z_0 = 1.$$

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

$$(5.4) \quad Z_t = e^{-\lambda t} \left( \frac{1 + \Phi_t}{1 + \Phi_0} \right), \quad t \in [0, T],$$

because the right-hand-side of (5.4) satisfies the SDE (5.3). Then an application of the Bayes formula to the definition of  $U$  in (2.9) yields the result.  $\square$

The point of (5.1) 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  $\Phi$  remains, as exhibited in (4.10), inheriting this feature from the filtered switching process  $\hat{Y}$ . Indeed, using the solution of the stock price SDE (4.8), the representation (4.10) can be converted to one involving the stock price and its history: with  $\Phi_0 = \phi$  and  $X_0 = x$ , we have

$$(5.5) \quad \Phi_t(\phi) = \phi e^{\kappa t} \left( \frac{X_t}{x} \right)^{-\eta/\sigma} + \lambda \int_0^t e^{\kappa(t-s)} \left( \frac{X_t}{X_s} \right)^{-\eta/\sigma} ds, \quad t \in [0, T],$$

where  $\kappa$  is a constant given by

$$\kappa := \lambda + \eta\nu_0 - \frac{1}{2}\eta^2, \quad \text{with} \quad \nu_0 := \frac{\mu_0}{\sigma} - \frac{1}{2}\sigma$$

The second term on the right-hand-side of (5.5) is the awkward history-dependent term which makes numerical solution of the partial information ESO problem difficult. We will develop a numerical approximation for the partial information problem in Section 5.6. For  $\lambda = 0$ , we see that  $\Phi$  becomes a deterministic function of the current stock price, and this feature was exploited by Décamps et al [15, 16], Klein [38] and Ekström and Lu [20]. See also Ekström and Lindberg [19]. This reduction results in simpler computations than we require in Section 5.6.

**5.2. The binomial tree setting.** Consider now a discrete time, discrete space setting. Recall that  $T > 0$  is the option maturity and divide the interval  $[0, T]$  into  $N$  steps. Each time step is of length  $h = T/N$ .<sup>1</sup> Define a switching, two-state Markov chain  $Y_k$ ,  $k \in \{0, \dots, N\}$  which is represented by the transition probability matrix

$$(5.6) \quad \begin{pmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{pmatrix} = \begin{pmatrix} \mathbb{P}(Y_{k+1} = 0 | Y_k = 0) & \mathbb{P}(Y_{k+1} = 1 | Y_k = 0) \\ \mathbb{P}(Y_{k+1} = 0 | Y_k = 1) & \mathbb{P}(Y_{k+1} = 1 | Y_k = 1) \end{pmatrix} \\ = \begin{pmatrix} e^{-\lambda_0 h} & 1 - e^{-\lambda_0 h} \\ 1 - e^{-\lambda_1 h} & e^{-\lambda_1 h} \end{pmatrix},$$

with an initial state  $Y_0 = i$ ,  $i \in \{0, 1\}$  and intensities  $\lambda_0, \lambda_1$ . In our setting, for the ESO problem with one switch,  $\lambda_1 = 0$  and  $Y_0 = 0$ . We set  $\lambda_0 \equiv \lambda$  for consistency for the remainder of the paper.

The stock returns  $R_k$ ,<sup>2</sup>  $k \in \{1, \dots, N-1\}$  are generated by a sequence of independent Bernoulli random variables. The stock price process  $X_k$  is then modelled as

$$(5.7) \quad X_{k+1} = R_{k+1} X_k, \quad k = 0, \dots, N-1,$$

where  $X_0 = x$  is the initial stock price. The stock return attains one of two possible values  $\mathfrak{u}$  and  $\mathfrak{d}$ , referred to as an up-return and down-return, respectively, and defined by

$$(5.8) \quad \mathfrak{u} = e^{\sigma\sqrt{h}}, \quad \mathfrak{d} = 1/\mathfrak{u}.$$

<sup>1</sup>Note there is no confusion with the process  $h_t$  defined earlier in (2.7).

<sup>2</sup>Note there is no confusion with the reward process defined earlier in (3.1).

This is the parameterisation of the standard Cox-Ross-Rubinstein (CRR) tree, see [13].

The background (full information) filtration  $\mathbb{F}$  is given by  $\mathcal{F}_k = \sigma(X_u, Y_u | u = 0, \dots, k)$ . We describe dynamics under the background filtration in the next section. The observation (partial information) filtration  $\widehat{\mathbb{F}}$  is given by  $\widehat{\mathcal{F}}_k \equiv \mathcal{F}_k^X = \sigma(X_u | u = 0, \dots, k)$ . When developing the probability filter in Section 5.4, we will use the fact that a stock return filtration given by  $\mathcal{F}_k^R = \sigma(R_u | u = 1, \dots, k)$  supplemented with  $X_0 = x$  carries the same information as the observation filtration  $\widehat{\mathbb{F}}$ .

**5.3. Dynamics under the background filtration.** The regime switching process is observable under the background filtration, and the probabilities of an up-return and down-return at step  $k \in \{0, \dots, N-1\}$  and in regime  $i \in \{0, 1\}$  are given by

$$(5.9a) \quad p_{ui} = \mathbb{P}(R_{k+1} = u | Y_{k+1} = i) = \frac{e^{\mu_i \sqrt{h}} - \mathfrak{d}}{u - \mathfrak{d}},$$

$$(5.9b) \quad p_{\mathfrak{d}i} = \mathbb{P}(R_{k+1} = \mathfrak{d} | Y_{k+1} = i) = 1 - p_{ui},$$

Observe that the stock drift  $\mu_i$ , which determines the expected return at step  $k+1$ , is aligned with the drift regime prevailing at step  $k+1$ . Alternatively, the stock drift can be aligned with the drift regime at step  $k$ . This alternative time discretisation is used, for example, in Bollen [6] or Liu [44]. We derived and implemented this alternative and obtained virtually indistinguishable results.

We now describe the evolution of the joint process  $(X_k, Y_k)$ . Since the stock price at step  $k+1$  is given by (5.7), where the return depends only on the Markov chain transition from step  $k$  to  $k+1$ , the process  $(X_k, Y_k)$  is Markov. Both the stock return and drift switching process take one of two possible values, and therefore from the known state  $(X_k = x, Y_k = i)$  at step  $k$ , the joint process is in one of four states at step  $k+1$ . The process evolution is given explicitly by

$$(5.10) \quad (X_k = x, Y_k = i) \longrightarrow \begin{cases} (X_{k+1} = xu, Y_{k+1} = 0) \text{ with probability } p_{u0}q_{i0}, \\ (X_{k+1} = x\mathfrak{d}, Y_{k+1} = 0) \text{ with probability } p_{\mathfrak{d}0}q_{i0}, \\ (X_{k+1} = xu, Y_{k+1} = 1) \text{ with probability } p_{u1}q_{i1}, \\ (X_{k+1} = x\mathfrak{d}, Y_{k+1} = 1) \text{ with probability } p_{\mathfrak{d}1}q_{i1}, \end{cases}$$

where each of the four transition probabilities  $p_{\cdot}q_{\cdot}$  represents the probability of arriving into the corresponding state at step  $k+1$ , conditioned on the state at step  $k$ .

**5.4. Dynamics under the observation filtration.** We now develop a binomial tree version of the filter of Section 2.2. In contrast to the full information case, knowledge of the drift state is not available, and so is estimated based on realised stock returns.

The filter estimate of  $Y_k$  will be denoted by  $\widehat{Y}_k$ , and it is an estimate based on the observation filtration of the probability that the drift switching process is in state 1 at step  $k$ . The filtered probability  $\widehat{Y}_k$  is defined by

$$(5.11) \quad \widehat{Y}_k := \mathbb{P}(Y_k = 1 | \widehat{\mathcal{F}}_k) = \mathbb{E}[Y_k | \widehat{\mathcal{F}}_k], \quad k \in \{1, \dots, N\}.$$

Since we consider two drift regimes only, we have  $\mathbb{P}(Y_k = 0 | \widehat{\mathcal{F}}_k) = 1 - \widehat{Y}_k$ . We divide the filtering operation at each stock price step into two steps. In the first step, the stock price return is predicted and in the second step the filter is updated.

*Step 1: Predicting the return.* In the first filtering step the return for the next period is predicted, which amounts to calculating the transition probabilities of stock price moves under the observation filtration. We will denote by  $p_{uy}$  the probability of an up-move, and by  $p_{\mathfrak{d}y}$  the probability of a down-move, where  $y$  in the subscript denotes the

dependency of the expected return on the filtered probability. The transition probability is given by

$$(5.12) \quad p_{\varrho y} = p_{\varrho 0} [q_{00}(1 - y) + q_{10}y] + p_{\varrho 1} [q_{01}(1 - y) + q_{11}y], \quad \varrho \in \{\mathfrak{u}, \mathfrak{d}\},$$

where the filtered probability  $y$  was calculated at step  $k$ . The above formula is obtained by using the law of total probability, the independence of returns, the Markov chain probabilities (5.6), and the filter definition (5.11). In particular,

$$\begin{aligned} p_{\varrho y} &= \mathbb{P}(R_{k+1} = \varrho | \widehat{\mathcal{F}}_k) \\ &= \sum_{i=0}^1 \mathbb{P}(R_{k+1} = \varrho | Y_{k+1} = i, \widehat{\mathcal{F}}_k) \mathbb{P}(Y_{k+1} = i | \widehat{\mathcal{F}}_k) \\ &= \sum_{i=0}^1 p_{\varrho i} \left[ q_{0i} \mathbb{P}(Y_k = 0 | \widehat{\mathcal{F}}_k) + q_{1i} \mathbb{P}(Y_k = 1 | \widehat{\mathcal{F}}_k) \right] \end{aligned}$$

and by substituting in the filtered probability definition (5.11), the filter prediction (5.12) is obtained. The filter is initialised by assuming

$$\widehat{Y}_0 = \mathbb{P}(Y_0 = 1) = \mathbb{E}[Y_0] = y_0 \in [0, 1),$$

and thus the first return prediction can be calculated according to (5.12).

*Step 2: Updating the filter.* In the second filtering step the filtered probability is updated by evaluating  $\widehat{Y}_{k+1} = \mathbb{P}(Y_{k+1} = 1 | \widehat{\mathcal{F}}_{k+1})$ , and we will denote by  $y_{\mathfrak{u}}$  the filtered probability when the stock price moves up from step  $k$  to  $k + 1$ , and by  $y_{\mathfrak{d}}$  the probability when the stock price moves down. The filtered probability update is given by

$$(5.13) \quad y_{\varrho} = \frac{p_{\varrho 1} [q_{01}(1 - y) + q_{11}y]}{p_{\varrho 0} [q_{00}(1 - y) + q_{10}y] + p_{\varrho 1} [q_{01}(1 - y) + q_{11}y]}, \quad \varrho \in \{\mathfrak{u}, \mathfrak{d}\}.$$

The filtered probability update is derived by a direct application of Bayes' formula. In particular,

$$\begin{aligned} (\widehat{Y}_{k+1} = y_{\varrho}) &= \mathbb{P}(Y_{k+1} = 1 | R_{k+1} = \varrho, \widehat{\mathcal{F}}_k) \\ &= \mathbb{P}(R_{k+1} = \varrho | Y_{k+1} = 1, \widehat{\mathcal{F}}_k) \times \mathbb{P}(Y_{k+1} = 1 | \widehat{\mathcal{F}}_k) / \mathbb{P}(R_{k+1} = \varrho | \widehat{\mathcal{F}}_k), \end{aligned}$$

and by using the arguments leading to (5.12) the expression (5.13) is obtained.

Observe from the return prediction (5.12) that the distribution of  $R_{k+1}$  depends on  $y$ , and therefore  $\mathbb{P}(R_{k+1} | \widehat{\mathcal{F}}_k) = \mathbb{P}(R_{k+1} | \widehat{Y}_k = y)$ . Further observe from the filter update (5.13) that the value of  $\widehat{Y}_{k+1}$  is calculated by evaluating  $\mathbb{P}(Y_{k+1} = 1 | R_{k+1} = \varrho, \widehat{Y}_k = y)$ . It follows that the joint process  $(R_k, \widehat{Y}_k)$  is Markov, and thus also  $(S_k, \widehat{Y}_k)$  is Markov. The evolution of the joint stock price and filtered probability process is given by

$$(5.14) \quad (X_k = x, \widehat{Y}_k = y) \longrightarrow \begin{cases} (X_{k+1} = x\mathfrak{u}, \widehat{Y}_{k+1} = y_{\mathfrak{u}}) & \text{with probability } p_{\mathfrak{u}y}, \\ (X_{k+1} = x\mathfrak{d}, \widehat{Y}_{k+1} = y_{\mathfrak{d}}) & \text{with probability } p_{\mathfrak{d}y}, \end{cases}$$

where the transition probabilities  $p_{\varrho y}$  are given by the predicted return probabilities (5.12). We can directly observe that the transition probabilities sum to unity. As in the full information case, we now verify the transition probabilities (5.12). We demonstrate the derivation for  $p_{\mathfrak{u}y}$  only, since  $p_{\mathfrak{d}y}$  follows by an obvious modification. The probability of

arriving into  $(xu, y_u)$  conditioned on  $(x, y)$  is given by

$$\begin{aligned}
& \mathbb{P}(X_{k+1} = xu, \hat{Y}_{k+1} = y_u | X_k = x, \hat{Y}_k = y) \\
&= \mathbb{P}(R_{k+1} = u, \hat{Y}_{k+1} = y_u | \hat{Y}_k = y) \\
&= \mathbb{P}(R_{k+1} = u, \hat{Y}_{k+1} = y_u, \hat{Y}_k = y) / \mathbb{P}(\hat{Y}_k = y) \\
&= \mathbb{P}(\hat{Y}_{k+1} = y_u | R_{k+1} = u, \hat{Y}_k = y) \times \mathbb{P}(R_{k+1} = u, \hat{Y}_k = y) / \mathbb{P}(\hat{Y}_k = y) \\
&= \mathbb{P}(R_{k+1} = u | \hat{Y}_k = y) \\
&= p_{uy}.
\end{aligned}$$

In the derivation above, we have used the independence of returns and the fact that  $\mathbb{P}(\hat{Y}_{k+1} = y_u | R_{k+1} = u, \hat{Y}_k = y) = 1$ .

**5.5. Optimal stopping by dynamic programming .** We tackle the optimal stopping problem under the full and partial information given by (2.8) and (2.9), respectively, by dynamic programming on the binomial trees developed in the previous section. We use standard results on optimal stopping by discrete-time dynamic programming, see, for example, Björk [5, Chapter 21] for a self-contained exposition of the general theory.

**5.5.1. The full information case.** The regime switching process is observable in the full information case. The discrete-time analogue to the value function (3.3) is

$$v(k, x, i) = \max_{k \leq \tau \leq N} \mathbb{E} \left[ e^{-r(\tau-k)h} (X_\tau - K)^+ \middle| X_k = x, Y_k = i \right], \quad i \in \{0, 1\}.$$

It follows from the general results on dynamic programming for optimal stopping that the value function for  $i \in \{0, 1\}$  is the solution to the recursive equation

$$\begin{aligned}
(5.15a) \quad v(k, x, i) &= \max \left\{ (x - K)^+, e^{-rh} \mathbb{E} \left[ v(k+1, X_{k+1}, Y_{k+1}) \middle| X_k = x, Y_k = i \right] \right\} \\
&= \max \left\{ (x - K)^+, e^{-rh} \sum_{j=0}^1 q_{ij} [p_{uj} v(k+1, xu, j) + p_{dj} v(k+1, x\mathfrak{d}, j)] \right\}
\end{aligned}$$

at step  $k = N-1, N-2, \dots, 0$ , and by the boundary condition

$$(5.15b) \quad v(N, x, i) = (x - K)^+.$$

at the final step  $N$ . We solve the value function for  $i \in \{0, 1\}$  by running the recursion (5.15) backwards on the binomial tree with transition probabilities according to (5.10). In order to run the recursion, two value function trees need to be implemented: the  $v(\cdot, \cdot, 0)$  tree for regime 0 and  $v(\cdot, \cdot, 1)$  tree for regime 1.

The optimal stopping time for a fixed step  $k$  and for  $i \in \{0, 1\}$  is given by

$$\bar{\tau} = \inf \{k \leq m \leq N : (m, X_m, Y_m) \notin \mathcal{C}_i\},$$

where  $\mathcal{C}_i$  is the continuation region defined by

$$\mathcal{C}_i := \{(m, x, i) : v(m, x, i) > (x - K)^+\}, \quad i = 0, 1.$$

Note that since we have two drift regimes  $i$  is a binary variable, and there exist a pair of two-dimensional continuation regions.

5.5.2. *The partial information case.* The regime switching process is not observable in the partial information case. The discrete-time analogue of the value function (4.1) is

$$u(k, x, y) = \sup_{k \leq \tau \leq N} \mathbb{E} \left[ e^{-r(\tau-k)h} (X_\tau - K)^+ \middle| X_k = x, \hat{Y}_k = y \right].$$

It follows from the general results that the optimal value function is the solution to the recursive equation given by

$$\begin{aligned} u(k, x, y) &= \max \left\{ (x - K)^+, e^{-rh} \mathbb{E} \left[ u(k+1, X_{k+1}, \hat{Y}_{k+1}) \middle| X_k = x, \hat{Y}_k = y \right] \right\}, \\ (5.16a) \quad &= \max \left\{ (x - K)^+, e^{-rh} \left[ p_{uy} u(k+1, xu, y_u) + p_{\mathfrak{d}y} u(k+1, x\mathfrak{d}, y_{\mathfrak{d}}) \right] \right\}, \end{aligned}$$

at a step  $k = N-1, N-2, \dots, 0$ , and by the boundary condition

$$(5.16b) \quad u(N, x, y) = (x - K)^+.$$

at the final step  $N$ . We solve the value function by running the recursion (5.16) backward on the binomial tree with transition probabilities according to (5.14). However, running the partial information backward recursion is not as straightforward as running the full information recursion as the filtered probability process  $\hat{Y}_k$  is path-dependent. We suggest an approximate solution method to tackle the path-dependency in the following subsection.

The optimal stopping time for a fixed step  $k$  is given by

$$\tau^* = \inf \{ k \leq m \leq N : (m, X_m, \hat{Y}_m) \notin \hat{\mathcal{C}} \},$$

where  $\hat{\mathcal{C}}$  is the continuation region defined by

$$\hat{\mathcal{C}} = \{ (m, x, y) : u(m, x, y) > (x - K)^+ \}.$$

**5.6. An approximation algorithm for the partial information recursion.** Assume that  $\hat{Y}_k = y$ ,  $X_k = x$  and consider the following two stock price paths. First, the stock price moves up and down, giving  $X_{k+2} = xu\mathfrak{d}$ . Second, the stock price moves down and up, giving  $X_{k+2} = x\mathfrak{d}u$ . Since the binomial tree recombines,  $xu\mathfrak{d} = x\mathfrak{d}u = x$ . However, the value of  $\hat{Y}_{k+2}$  implied from  $y$  and  $u\mathfrak{d}$  is in general different from the value of  $\hat{Y}_{k+2}$  implied from  $y$  and  $\mathfrak{d}u$ , as can be seen from (5.13). This means that a binomial tree carrying the filtered probability values does not recombine. The stock price tree at step  $k$  consists of  $k+1$  lattice nodes and paths, whereas the filtered probability tree would have  $2^k$  nodes and paths. The exponential growth in the lattice substantially restricts the number of computationally feasible steps. To overcome this computational burden, we develop an approximation in the spirit of Hull and White [32], Barraquand and Pudet [4], or Klassen [37]. These papers focus on Asian type options, where the path-dependency follows from the need to consider the time-average stock price in the value function. The convergence of the approximations suggested in the aforementioned papers is studied in Forsyth et al [23].

Let  $X_{k,j}$  denote the stock price at step  $k$  and a state space position  $j$ . Analogously, we denote by  $\hat{Y}_{k,j}$  and  $U_{k,j}$  the filtered probability and the partial information value function, respectively, at a corresponding tree node  $(k, j)$ . As we have argued above  $\hat{Y}_{k,j}$  is not unique, but depends on the path leading to the node  $(k, j)$ . Specifically, the number of  $\hat{Y}_{k,j}$  values is given by the binomial coefficient  $\binom{k}{j}$ , with the same applying to  $U_{k,j}$ . Our approximation algorithm is implemented as follows.

- (1) At each node  $(k, j)$  we assume  $L$  different values of  $\widehat{Y}_{k,j}$ , and create the following equidistantly spaced  $\bar{y}$ -grid:

$$0 = \bar{y}^1 < \bar{y}^2 < \dots < \bar{y}^{L-1} < \bar{y}^L = 1.$$

For each  $\bar{y}^l$ ,  $l = 1, \dots, L$ , we compute according to (5.13) the filtered probability in the case of an up-return and down-return, and denote the probability by  $y_u^l$  and  $y_d^l$ , respectively. Furthermore, we find two values  $\bar{y}^{u_1}$  and  $\bar{y}^{u_2}$  in the  $\bar{y}$ -grid such that they are the nearest values to  $y_u^l$  and satisfy

$$\bar{y}^{u_1} \leq y_u^l \leq \bar{y}^{u_2}.$$

The superscripts  $u_1, u_2 \in \{1, \dots, L\}$  thus denote a position in the  $\bar{y}$ -grid. Analogously, we find two values  $\bar{y}^{d_1}, \bar{y}^{d_2}$ , and thus two positions  $d_1, d_2$  corresponding to  $y_d^l$ . Note that this calculation is carried out only once, because the  $\bar{y}$ -grid is the same for each node of the binomial tree.

- (2) The partial information value function is a function of the filtered probability, and thus we need to evaluate  $L$  value functions at each node of the binomial tree. We denote by  $U_{k,j}^l$  the value function at a node  $(k, j)$  and corresponding to the probability grid point  $\bar{y}^l$ . We thus work with an  $U_{k,j}$ -grid consisting of  $U_{k,j}^1, \dots, U_{k,j}^L$  values.
- (3) At terminal tree nodes we use the boundary condition (5.16b) to compute  $U_{N,j}^l$ ,  $j = 0, \dots, N$ ,  $l = 1, \dots, L$ . Note that at a given terminal node  $(N, j)$  the value function  $U_{N,j}^l$  is the same for all  $l$ , but  $U_{k,j}^l$  starts to vary with respect to  $l$  for  $k < N$ . At each node  $(N-1, j)$  we use the recursive equation (5.16a) to calculate

$$U_{N-1,j}^l = \max \left\{ (x - K)^+, e^{-rh} \left( p_{uy} U_{N,j+1}^l + p_{dy} U_{N,j}^l \right) \right\},$$

where  $p_y$  is given by (5.12) with  $y = \bar{y}^l$ .

- (4) We iterate backwards through steps  $k = N-2, \dots, 0$ . At each node  $(k, j)$  we compute according to the recursive equation (5.16a) the value function

$$U_{k,j}^l = \max \left\{ (x - K)^+, e^{-rh} \left( p_{uy} \tilde{U}_{k+1,j+1}^{u,l} + p_{dy} \tilde{U}_{k+1,j}^{d,l} \right) \right\},$$

where  $p_y$  is given by (5.12) with  $y = \bar{y}^l$ , and where  $\tilde{U}_{k+1,j}^{d,l}$  is an interpolated value function from the previous recursion step. The interpolated value function  $\tilde{U}_{k+1,j+1}^{u,l}$  corresponds to  $y_u^l$  and is given by

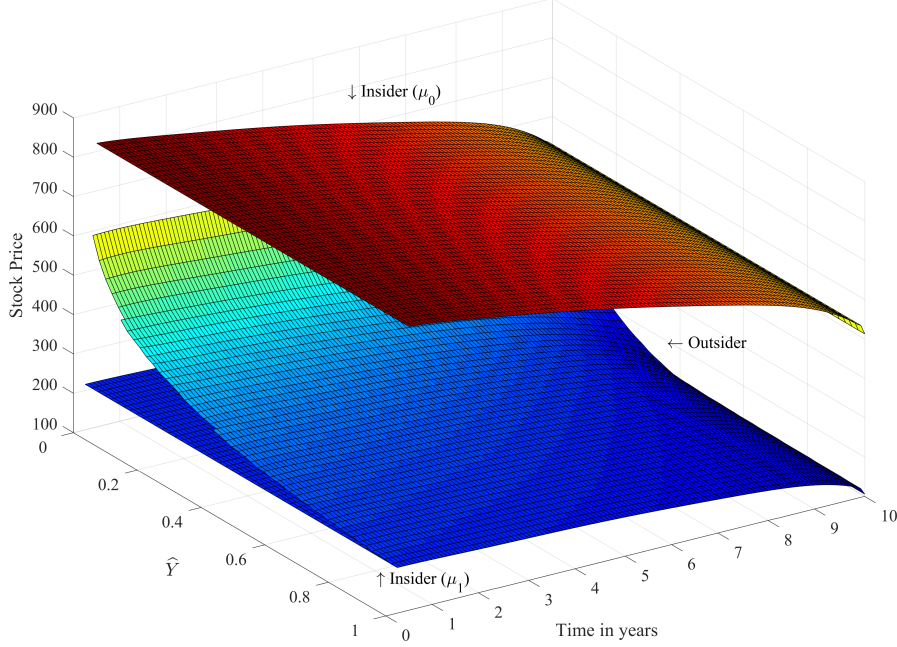
$$\tilde{U}_{k+1,j+1}^{u,l} = U_{k+1,j+1}^{u_1} + \frac{U_{k+1,j+1}^{u_2} - U_{k+1,j+1}^{u_1}}{y^{u_2} - y^{u_1}} (y_u^l - y^{u_1}),$$

where  $U_{k+1,j+1}^{u_1}$  and  $U_{k+1,j+1}^{u_2}$  is located at the  $u_1$  and  $u_2$  position of the  $U_{k+1,j+1}$ -grid, and thus corresponds to  $\bar{y}^{u_1}$  and  $\bar{y}^{u_2}$  in the  $\bar{y}$ -grid. Note that  $\tilde{U}_{k+1,j+1}^{u,l}$  does not exactly equal  $U_{k+1,j+1}^{u,l}$  due to a nonlinearity of  $y \rightarrow U$ . Analogously, we enumerate  $\tilde{U}_{k+1,j}^{d,l}$ . In other words, for a given  $\bar{y}^l$  at node  $(k, j)$  we have carried out a value function approximation based on maps  $y_u^l \rightarrow \tilde{U}_{k+1,j+1}^{u,l}$ ,  $y_d^l \rightarrow \tilde{U}_{k+1,j}^{d,l}$ .

- (5) At the initial node  $(0, 0)$ , we assume a unique  $y$ , and therefore we obtain a unique value function  $U_{0,0}$ .

We have also tested a more sophisticated algorithm with a probability grid specific to each tree node. In particular, for each stock price node it is possible to calculate the minimal and maximal filtered probability. These probability extremes are then used as

the probability grid bounds. This sophistication of the probability grid leads to ESO values which differ from those we present at most in the first decimal place.

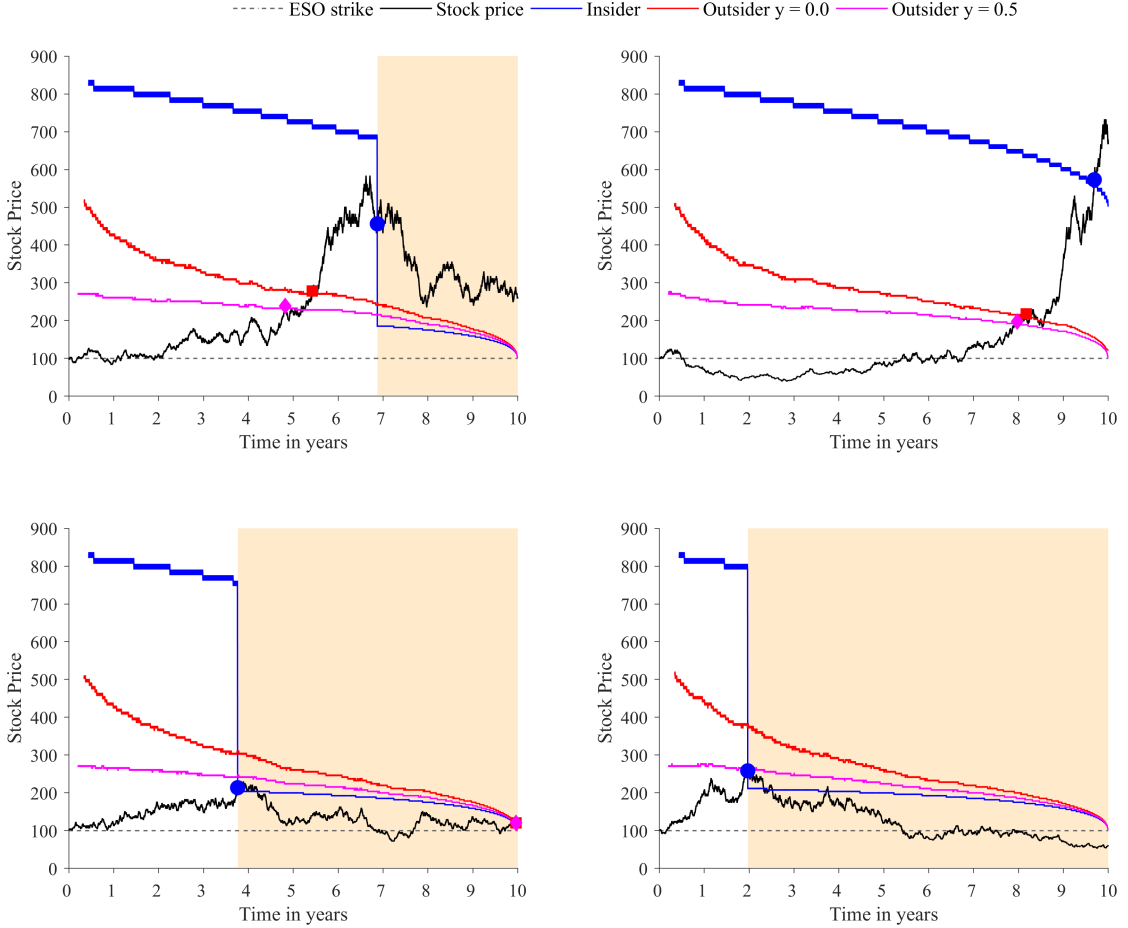


**FIGURE 1. Exercise surfaces under full (insider) and partial information (outsider) against time and the filtered process  $\hat{Y}$ .** The uppermost and lowermost surfaces are those of the insider: the uppermost surface  $x_0(t)$  in regime 0 with  $\mu_0$ , and the lowermost surface  $x_1(t)$  in regime 1 with  $\mu_1$ . These do not depend upon  $\hat{Y}$  so each surface for the insider is constant in the  $\hat{Y}$  direction, and has been plotted as a surface for comparison with the outsider's surface, which does depend upon  $\hat{Y}$ . The outsider's exercise surface  $x^*(t, y)$  depends upon the filtered process  $\hat{Y}_t$  and lies between the insider's two surfaces. (The exercise surface for the outsider should be viewed as conditional on the joint attainability of the filtered probability and stock price. Some of the filtered probabilities may not be attainable for depicted stock price exercise levels.) 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  $\lambda = 10\%$ , volatility  $\sigma = 30\%$ , and the riskfree rate is  $r = 2.5\%$ . The binomial trees use  $N = 2500$  steps and the filtered probability  $\hat{Y}$  grid has 250 points ( $L = 250$ ). Note also that exercise boundaries have been smoothed using a polynomial regression.

**5.7. Numerical example: ESO valuation & exercise.** Given 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 first illustrate the exercise surfaces generated by the model for the insider and outsider. Figure 1 plots the exercise surfaces under full and partial information against time and the filtered process  $\hat{Y}$ . We set the switch intensity to be  $\lambda = 10\%$  which implies a probability of 63 % of  $\mu_0$  switching to  $\mu_1$  during the option's life. Recall the two thresholds of the insider,  $x_0(t)$  in regime 0 and  $x_1(t)$  in regime 1, do not depend upon  $\hat{Y}$ . Hence on the graph, each surface for the insider is constant in the  $\hat{Y}$  direction, and has been plotted as a surface for comparison with the outsider's surface, which does depend upon  $\hat{Y}$ . As expected from Corollary 3.3, we have



the ordering  $x_0(t) > x_1(t) \geq K$ ;  $t \in [0, T]$  and the two terminal values of the insider's boundaries are given by  $x_0(10) = 500$  and  $x_1(10) = 100$ . We see the exercise surface for the outsider,  $x^*(t, y)$  is indeed decreasing in time and  $y$ , as proven earlier in Lemma 4.2. Finally, in the limit as  $y \rightarrow 1$ ,  $x^*(t, y)$  approaches  $x_1(t)$ . Thus the numerical results are in accordance with the theoretical results of Sections 3 and 4, giving a check on the numerical algorithm.



**FIGURE 2. Monte Carlo simulations of the stock price, thresholds and exercise decisions of the insider and outsider.** In each panel, the shaded background indicates the switch in drift regime to  $\mu_1 < \mu_0$ . In each panel we display the stock price, the insider's exercise boundary, and the outsider's exercise boundary under two alternative assumptions on the outsider's initial probability of being in regime 1:  $\hat{Y}_0 = \mathbb{E}[Y_0] = y_0 = 0$  and  $y_0 = 0.5$ . Exercise decisions of the (insider, outsider with  $y_0 = 0$ , outsider with  $y_0 = 0.5$ ) are marked with (circles, squares, diamonds). 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  $\lambda = 10\%$ , volatility  $\sigma = 30\%$ , and the riskfree rate is  $r = 2.5\%$ . The binomial trees use  $N = 2500$  steps and the filtered probability grid for  $\hat{Y}$  has 250 points ( $L = 250$ ).

We are of course interested in the ESO exercise policies. To illustrate insider and outsider exercise patterns, we ran simulations as follows. Stock price paths and drift switching processes are simulated on the binomial tree. A set of four simulations are plotted in Figure 2. In each panel we display the stock price, the insider's exercise

boundary, and the outsider's exercise boundary under two alternative assumptions on the outsider's initial probability of being in regime 1:  $\hat{Y}_0 = \mathbb{E}[Y_0] = y_0 = 0$  and  $y_0 = 0.5$ . The shaded area in each panel denotes the time after the switch has occurred, ie. the drift has switched from  $\mu_0$  to  $\mu_1$ . Exercise decisions are recorded on each plot for both the outsider with  $y_0 = 0$  (with a square) and  $y_0 = 0.5$  (with a diamond) as well as the insider (with a circle).

The top left panel demonstrates a scenario where the insider exercises in direct response to the switch and benefits from the additional information. In this panel, the outsider has already exercised (in both alternatives for the outsider's starting probability of regime 1) as the stock price crosses the outsider's boundaries. The insider continues to wait as he knows the switch has not occurred. He then benefits with a larger exercise payoff by exercising exactly at the changepoint.

In the top right panel, there is no switch during the option's life and the stock price performs very well. The stock price first reaches the boundary of the outsider with  $y_0 = 0.5$ , then the boundary of the outsider with  $y_0 = 0$  and finally, the much higher boundary of the insider. Under this scenario, the insider has benefited from the additional information (the knowledge that the switch has not occurred) and has secured a much higher payoff than the outsider.

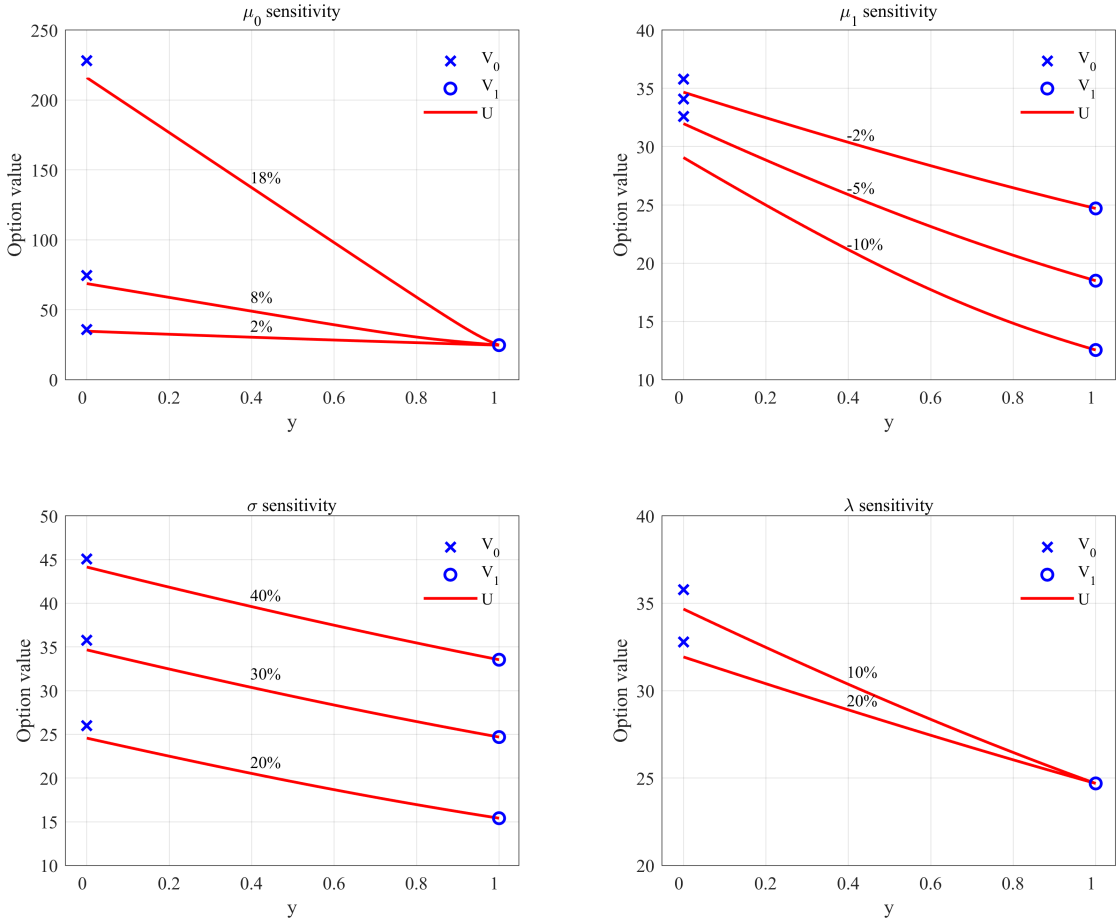
In both lower panels, the stock price performs less well and prior to the changepoint, none of the agent's exercise boundaries have been reached. In both panels, the insider exercises on the changepoint, significantly in-the-money. In the lower-left hand panel, the outsider exercises just prior to maturity when the option is just in-the-money, but has received much less than the insider. In the lower-right hand panel, the outsider never exercises, as the stock price puts the option out-of-the-money for much of it's life beyond about 5.5 years. Again, the insider has benefitted from the knowledge of the switch and obtained a significant payoff.

Figure 3 displays comparative statics for full and partial information option values. Each panel plots option value against  $\hat{Y}_0 = \mathbb{E}[Y_0] = y_0$ , the outsider's initial probability of being in state 1. In each panel, one input parameter ( $\mu_0, \mu_1, \sigma$ , and  $\lambda$ ) is varied at a time. Some overall observations are that we always have the relation  $V_0 \geq U \geq V_1$  so the outsider's option value for any value of  $y_0$  lies between the two insider values. We also observe that the outsider's option value,  $U$ , is monotonically decreasing as  $y_0$  increases between zero and one. That is, the outsider reduces her valuation as the initial probability of being in the low drift regime increases. As  $y_0 \rightarrow 1$ , we see  $U \rightarrow V_1$ . That is, once the outsider knows with probability 1 that the stock is in regime 1, she places the same value on the option as the insider does in that state, as there are no further switching possibilities. However, when  $y_0 = 0$  and the outsider knows the price begins in regime 0, she still places less value on the option than the insider who knows for certain that the price is in regime 0. This is because the insider will benefit from knowing the timing of the potential switch in regime, whilst the outsider must filter the probability of the switching process.

We now consider how option values change with market parameters. In the top left panel, we see a higher  $\mu_0$  results in higher option values for both the insider in state 0 and the outsider (except in the limit as  $y_0 \rightarrow 1$ ). The insider's value in state 1 does not vary with  $\mu_0$ , as once the switch has occurred, this drift is irrelevant to the insider. We vary  $\mu_1$  in the top right panel. As we would expect, higher (less negative) values of the drift in state 1 increases option values for the insider in both states, and the outsider. The lower left panel displays the behaviour of option values as volatility  $\sigma$  changes. Option values are increasing in volatility for the insider in both states and the outsider as might be expected due to the convexity of the call payoff and the assumption of

constant volatility. Proving monotonicity in volatility would follow almost immediately from the convexity in  $x$  in Lemmas 3.1 and 4.2 along the lines of Hobson [30]. See also El Karoui, Jeanblanc and Shreve [21]. The lower right panel displays option values for alternative values of the transition intensity,  $\lambda$ . We see a higher probability of a downward jump in drift reduces option values for the insider in state 0 and the outsider (except in the limit as  $y_0 \rightarrow 1$ ). The insider's option value in state 1 does not vary with  $\lambda$  as the single switch has already occurred.

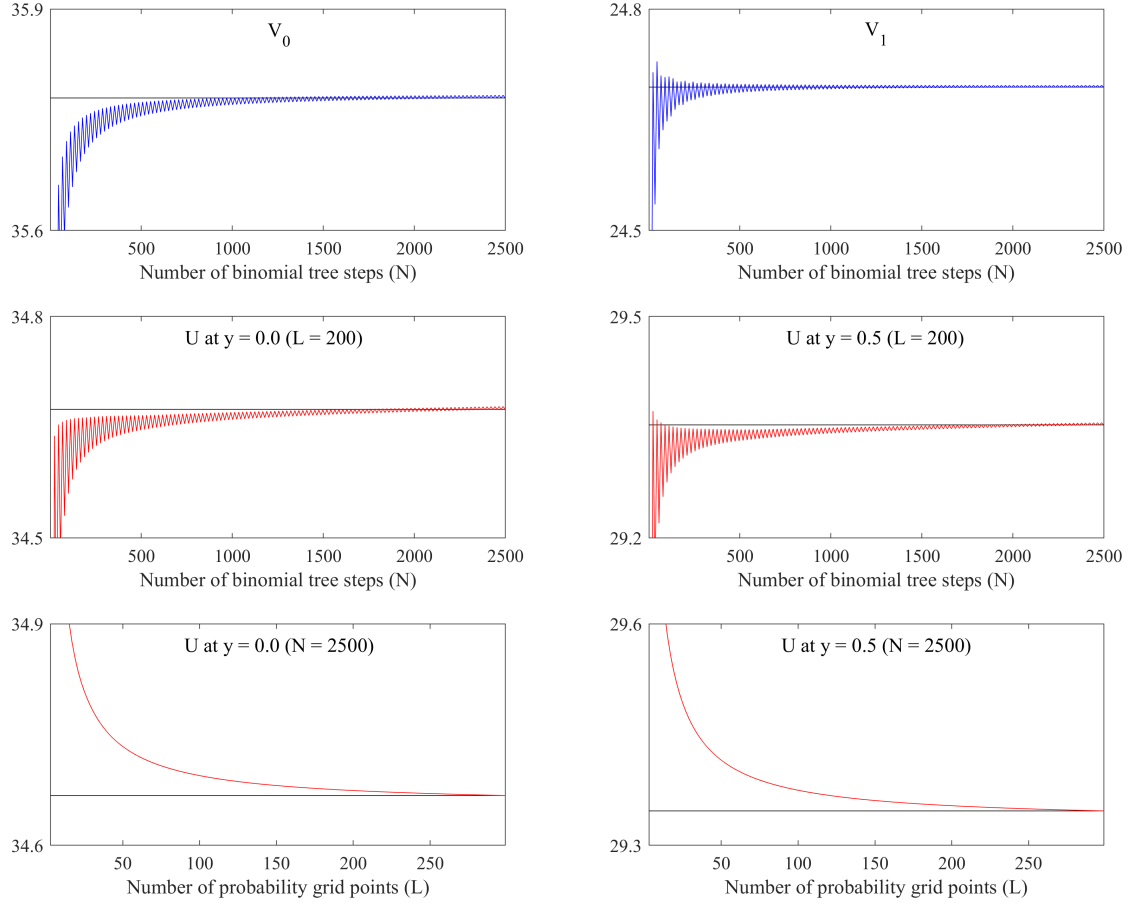
Table 1 reports more detailed comparative statics for option values of the insider in both states and the outsider, but only for two values  $y_0 = 0, y_0 = 0.5$  of the initial probability of being in state 1. The comparative statics re-enforce the above conclusions concerning the impact of market parameters on values.



**FIGURE 3. Comparative Statics for the full and partial information option values.** Each panel plots option value against  $\hat{Y}_0 = \mathbb{E}[Y_0] = y_0$ , the outsider's initial probability of being in state 1. Solid lines denote the outsider's option value  $U$ , which varies with  $y_0$ . The two styles of markers (cross, open circle) denote the insider's option values ( $V_0$  in state 0,  $V_1$  in state 1). Each panel varies a single parameter ( $\mu_0, \mu_1, \sigma$ , and  $\lambda$ ) keeping the others fixed. Values for the varying parameter are given in each panel. The option maturity is ten years and granted at-the-money with  $X_0 = K = 100$ . Parameters fixed in each panel (apart from the one varying in each panel) are: expected returns in the two regimes are  $\mu_0 = 2\%$ ,  $\mu_1 = -2\%$ , transition intensity  $\lambda = 10\%$ , volatility  $\sigma = 30\%$ , and the riskfree rate is  $r = 2.5\%$ . The binomial trees use  $N = 2500$  steps and the filtered probability  $\hat{Y}$  grid has 250 points ( $L = 250$ ).

| read:   |           | $\lambda = 10\%$ |       |         |       |       |       | $\lambda = 20\%$ |      |         |      |       |      |
|---------|-----------|------------------|-------|---------|-------|-------|-------|------------------|------|---------|------|-------|------|
|         |           | $\sigma = 20\%$  |       |         |       |       |       | $\sigma = 20\%$  |      |         |      |       |      |
| $v_0$   | $v_1$     | $\mu_1$          |       |         |       |       |       | $\mu_1$          |      |         |      |       |      |
| $u_0$   | $u_{0.5}$ | -2%              |       | -5%     |       | -10%  |       | -2%              |      | -5%     |      | -10%  |      |
| $\mu_0$ | 2%        | 26.0             | 15.4  | 24.7    | 10.5  | 23.7  | 6.4   | 23.0             | 15.4 | 20.9    | 10.5 | 19.4  | 6.4  |
|         |           | 24.6             | 19.6  | 22.1    | 15.4  | 19.8  | 11.7  | 21.9             | 18.5 | 18.8    | 14.2 | 16.0  | 10.4 |
|         | 8%        | 65.4             | 15.4  | 64.2    | 10.5  | 63.4  | 6.4   | 47.9             | 15.4 | 45.9    | 10.5 | 44.5  | 6.4  |
|         |           | 60.0             | 35.7  | 54.8    | 29.9  | 50.0  | 25.5  | 41.3             | 27.0 | 35.5    | 21.1 | 30.5  | 16.2 |
|         | 18%       | 223.3            | 15.4  | 222.4   | 10.5  | 221.9 | 6.4   | 138.9            | 15.4 | 137.4   | 10.5 | 136.4 | 6.4  |
|         |           | 211.0            | 111.4 | 202.3   | 103.8 | 193.4 | 97.3  | 124.6            | 68.5 | 114.8   | 60.5 | 105.4 | 53.5 |
|         |           | $\sigma = 30\%$  |       |         |       |       |       | $\sigma = 30\%$  |      |         |      |       |      |
|         |           |                  |       | $\mu_1$ |       |       |       |                  |      | $\mu_1$ |      |       |      |
|         |           | -2%              |       | -5%     |       | -10%  |       | -2%              |      | -5%     |      | -10%  |      |
| $\mu_0$ | 2%        | 35.8             | 24.7  | 34.1    | 18.5  | 32.6  | 12.5  | 32.8             | 24.7 | 30.1    | 18.5 | 27.7  | 12.5 |
|         |           | 34.7             | 29.3  | 32.0    | 24.5  | 29.1  | 19.4  | 31.9             | 28.2 | 28.3    | 23.0 | 24.6  | 17.8 |
|         | 8%        | 74.4             | 24.7  | 72.7    | 18.5  | 71.3  | 12.5  | 57.3             | 24.7 | 54.5    | 18.5 | 52.2  | 12.5 |
|         |           | 68.7             | 44.1  | 62.9    | 37.0  | 56.8  | 30.4  | 50.5             | 36.2 | 44.2    | 29.5 | 37.9  | 22.7 |
|         | 18%       | 228.2            | 24.7  | 226.7   | 18.5  | 225.5 | 12.5  | 145.5            | 24.7 | 143.0   | 18.5 | 141.0 | 12.5 |
|         |           | 216.0            | 117.7 | 206.0   | 108.3 | 194.5 | 99.0  | 130.4            | 75.2 | 118.5   | 65.1 | 106.0 | 55.3 |
|         |           | $\sigma = 40\%$  |       |         |       |       |       | $\sigma = 40\%$  |      |         |      |       |      |
|         |           |                  |       | $\mu_1$ |       |       |       |                  |      | $\mu_1$ |      |       |      |
|         |           | -2%              |       | -5%     |       | -10%  |       | -2%              |      | -5%     |      | -10%  |      |
| $\mu_0$ | 2%        | 45.1             | 33.5  | 43.2    | 26.5  | 41.3  | 19.2  | 42.1             | 33.5 | 39.0    | 26.5 | 36.0  | 19.2 |
|         |           | 44.1             | 38.5  | 41.3    | 33.2  | 38.0  | 27.2  | 41.3             | 37.3 | 37.4    | 31.6 | 33.1  | 25.4 |
|         | 8%        | 84.2             | 33.5  | 82.2    | 26.5  | 80.3  | 19.2  | 67.0             | 33.5 | 63.8    | 26.5 | 60.7  | 19.2 |
|         |           | 77.9             | 52.5  | 71.6    | 44.8  | 64.8  | 36.8  | 59.8             | 45.4 | 53.4    | 38.2 | 46.5  | 30.5 |
|         | 18%       | 235.8            | 33.5  | 233.8   | 26.5  | 232.0 | 19.2  | 154.1            | 33.5 | 150.9   | 26.5 | 148.0 | 19.2 |
|         |           | 223.5            | 125.1 | 213.1   | 114.7 | 200.3 | 103.5 | 138.4            | 82.8 | 125.4   | 71.5 | 111.0 | 59.7 |

TABLE 1. **Comparative Statics for the full and partial information option values.** Each subpanel of four numbers contains the option values for the insider in state 0, the insider in state 1, the outsider for  $y_0 = 0$  and the outsider for  $y_0 = 0.5$ . 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  $\lambda = 10\%, 20\%$ , volatility  $\sigma = 20\%, 30\%, 40\%$ , and the riskfree rate is fixed at  $r = 2.5\%$ . The binomial trees use  $N = 2500$  steps and the filtered probability  $\hat{Y}$  grid has 250 points ( $L = 250$ ).



**FIGURE 4. Convergence of the full and partial information algorithms.** The top two panels display option values for the insider in regimes 0 and 1 for varying numbers of binomial steps up to  $N = 2500$ . The middle two panels plot option values for the outsider for  $y_0 = 0$  (left) and  $y_0 = 0.5$  (right) for varying numbers of binomial steps up to  $N = 2500$ . The lower two panels show option values for the outsider for  $y_0 = 0$  (left) and  $y_0 = 0.5$  (right) for varying the number of grid points  $L$  in the filtered probability  $\hat{Y}$ , up to  $L = 300$ . The number of binomial steps is fixed at  $N = 2500$ . 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  $\lambda = 10\%$ , volatility  $\sigma = 30\%$ , and the riskfree rate is  $r = 2.5\%$ .

We conclude our numerical study with a discussion of the convergence of the full and partial information algorithms, see Figure 4. The top pair of panels provide option values for the insider in regime 0 (left panel) and regime 1 (right panel) for varying numbers of binomial steps up to  $N = 2500$ . The middle pair of panels gives outsider option values for  $y_0 = 0$  (left panel) and  $y_0 = 0.5$  (right panel) against varying numbers of binomial steps up to  $N = 2500$  and for a fixed value of  $L = 200$ . Each of these shows both algorithms converge rapidly and taking  $N = 2500$  in our earlier analysis is sufficient. The lower pair of panels provide outsider option values for  $y_0 = 0$  (left panel) and  $y_0 = 0.5$  (right panel) against varying number of grid points  $L$  in the filtered probability  $\hat{Y}$  up to  $L = 300$  whilst fixing  $N = 2500$ . We can see a grid with  $L = 250$  with  $N = 2500$  is adequate for generating prices.

## 6. CONCLUSIONS

We have analysed the ESO exercise decisions of agents who have full and partial information on a negative drift change point, from both a theoretical and numerical perspective. The results illustrate that, to a high degree, greater knowledge of the change point leads to more advantageous exercise patterns, demonstrating some empirically observed features of ESO exercise by insiders.

## REFERENCES

- [1] D. ABOODY, J. HUGHES, J. LIU, AND W. SU, *Are executive stock option exercises driven by private information?*, Review of Accounting Studies, 14 (2008), pp. 551–571.
- [2] A. BAIN AND D. CRISAN, *Fundamentals of stochastic filtering*, vol. 60 of Stochastic Modelling and Applied Probability, Springer, New York, 2009.
- [3] G. BARONE-ADESI AND R. J. ELLIOTT, *Approximations for the values of American options*, Stochastic Anal. Appl., 9 (1991), pp. 115–131.
- [4] J. BARRAQUAND AND T. PUDET, *Pricing of american path-dependent contingent claims*, Mathematical Finance, 6 (1996), pp. 17–51.
- [5] T. BJÖRK, *Arbitrage theory in continuous time*, Oxford University Press, 2009.
- [6] N. P. BOLLEN, *Valuing options in regime-switching models*, The Journal of Derivatives, 6 (1998), pp. 38–49.
- [7] R. BROOKS, D. CHANCE, AND B. CLINE, *Private information and the exercise of executive stock options*, Financial Management, Fall (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.
- [9] J. CARPENTER, *The exercise and valuation of executive stock options*, Journal of Financial Economics, 48 (1998), pp. 127–158.
- [10] J. CARPENTER AND B. REMMERS, *Executive stock option exercises and inside information*, Journal of Business, 74 (2001), pp. 513–534.
- [11] J. CARPENTER, R. STANTON, AND N. WALLACE, *The importance of behavioral factors in the exercise and valuation of employee stock options*, Working paper, (2015).
- [12] D. CICERO, *The manipulation of executive stock option exercise strategies: Information timing and backdating*, Journal of Finance, 64 (2009), pp. 2627–2663.
- [13] J. C. COX, S. A. ROSS, AND M. RUBINSTEIN, *Option pricing: A simplified approach*, Journal of Financial Economics, 7 (1979), pp. 229–263.
- [14] M. DAI, Q. ZHANG, AND Q. ZHU, *Trend following trading under a regime switching model*, SIAM J. Financial Math., 1 (2010), pp. 780–810.
- [15] J.-P. DÉCAMPS, T. MARIOTTI, AND S. VILLENEUVE, *Investment timing under incomplete information*, Math. Oper. Res., 30 (2005), pp. 472–500.
- [16] ———, *Investment timing under incomplete information: erratum*, Math. Oper. Res., 34 (2009), pp. 255–256.
- [17] J. DETEMPLE AND S. SUNDARESAN, *Nontraded asset valuation with portfolio constraints: A binomial approach*, Review of Financial Studies, 12 (1999), pp. 835–872.
- [18] J. DETEMPLE AND W. TIAN, *The valuation of American options for a class of diffusion processes*, Management Science, 48 (2002), pp. 917–937.
- [19] E. EKSTRÖM AND C. LINDBERG, *Optimal closing of a momentum trade*, Journal of Applied Probability, 2 (2013), pp. 374–387.
- [20] E. EKSTRÖM AND B. LU, *Optimal selling of an asset under incomplete information*, Int. J. Stoch. Anal., (2011), pp. Art. ID 543590, 17.
- [21] N. EL KAROUI, M. JEANBLANC, AND S. SHREVE, *Robustness of the black and scholes formula*, Mathematical Finance, 2 (1998), pp. 93–126.
- [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. FORSYTH, K. VETZAL, AND R. ZVAN, *Convergence of numerical methods for valuing path-dependent options using interpolation*, Review of Derivatives Research, 5 (2002), pp. 273–314.
- [24] A. FRIEDMAN, *Partial differential equations of parabolic type*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 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.

- [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.
- [27] M. GRASSELLI AND V. HENDERSON, *Risk aversion and block exercise of executive stock options*, J. Econom. Dynam. Control, 33 (2009), pp. 109–127.
- [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.
- [29] B. HALL AND K. MURPHY, *Stock options for undiversified executives*, Journal of Accounting and Economics, 33 (2002), pp. 3–42.
- [30] D. HOBSON, *Volatility mis-specification, option pricing and super-replication via coupling*, Annals of Applied Probability, 8 (1998), pp. 193–205.
- [31] S. HUDDART AND M. LANG, *Information distribution within firms: evidence from stock option exercises*, Journal of Accounting and Economics, 34 (2003), pp. 3–31.
- [32] J. HULL AND A. WHITE, *Efficient procedures for valuing european and american path-dependent options*, The Journal of Derivatives, 1 (1993), pp. 21–31.
- [33] A. JOBERT AND L. C. G. ROGERS, *Option pricing with Markov-modulated dynamics*, SIAM J. Control Optim., 44 (2006), pp. 2063–2078.
- [34] I. KARATZAS AND S. E. SHREVE, *Brownian motion and stochastic calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.
- [35] ———, *Methods of mathematical finance*, vol. 39 of Applications of Mathematics (New York), Springer-Verlag, New York, 1998.
- [36] A. Q. M. KHALIQ AND R. H. LIU, *New numerical scheme for pricing American option with regime-switching*, Int. J. Theor. Appl. Finance, 12 (2009), pp. 319–340.
- [37] T. R. KLASSEN, *Simple, fast, and flexible pricing of asian options*, Journal of Computational Finance, 4 (2001), pp. 89–124.
- [38] M. KLEIN, *Comment on “Investment timing under incomplete information”*, Math. Oper. Res., 34 (2009), pp. 249–254.
- [39] N. KULATILAKA AND A. MARCUS, *Valuing employee stock options*, Financial Analysts Journal, 50 (1994), pp. 46–56.
- [40] H. KUNITA, *Stochastic flows and stochastic differential equations*, vol. 24 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1997. Reprint of the 1990 original.
- [41] H. LE AND C. WANG, *A finite time horizon optimal stopping problem with regime switching*, SIAM J. Control Optim., 48 (2010), pp. 5193–5213.
- [42] 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.
- [43] ———, *Exponential hedging with optimal stopping and application to employee stock option valuation*, SIAM J. Control Optim., 48 (2009), pp. 1422–1451.
- [44] R. LIU, *Regime-switching recombining tree for option pricing*, International Journal of Theoretical and Applied Finance, 13 (2010), pp. 479–499.
- [45] 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.
- [46] G. PESKIR AND A. SHIRYAEV, *Optimal stopping and free-boundary problems*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2006.
- [47] L. C. G. ROGERS AND D. WILLIAMS, *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.
- [48] N. TOUZI, *American options exercise boundary when the volatility changes randomly*, Appl. Math. Optim., 39 (1999), pp. 411–422.

VICKY HENDERSON, DEPARTMENT OF STATISTICS, ZEEMAN BUILDING, UNIVERSITY OF WARWICK, COVENTRY CV4 7AL, UK

*E-mail address:* Vicky.Henderson@warwick.ac.uk

KAMIL KLADÍVKO, SCHOOL OF BUSINESS, ÖREBRO UNIVERSITY, 701 82 ÖREBRO, SWEDEN

*E-mail address:* kladivko@gmail.com

MICHAEL MONOYIOS, MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, UK

*E-mail address:* monoyios@maths.ox.ac.uk