

# Asset liquidation under drift uncertainty and regime-switching volatility

Juozas Vaicenavicius\*

Friday 5<sup>th</sup> January, 2018

## Abstract

Optimal liquidation of an asset with unknown constant drift and stochastic regime-switching volatility is studied. The uncertainty about the drift is represented by an arbitrary probability distribution; the stochastic volatility is modelled by  $m$ -state Markov chain. Using filtering theory, an equivalent reformulation of the original problem as a four-dimensional optimal stopping problem is found and then analysed by constructing approximating sequences of three-dimensional optimal stopping problems. An optimal liquidation strategy and various structural properties of the problem are determined. Analysis of the two-point prior case is presented in detail, building on which, an outline of the extension to the general prior case is given.

*MSC 2010 subject classifications:* primary 60G40; secondary 91G80, 60J25.

*Keywords and phrases:* optimal liquidation, drift uncertainty, regime-switching volatility, sequential analysis, optimal stopping, stochastic filtering.

## 1 Introduction

Selling is a fundamental and ubiquitous economic operation. As the prices of goods fluctuate over time, ‘What is the best time to sell an asset to maximise revenue?’ qualifies as a basic question in Finance. Suppose that an asset needs to be sold before a known deterministic time  $T > 0$  and that the only source of information available to the seller is the price history. A natural mathematical reformulation of the aforementioned optimal selling question is to find a selling time  $\tau^* \in \mathcal{T}_T$  such that

$$\mathbb{E}[S_{\tau^*}] = \sup_{\tau \in \mathcal{T}_T} \mathbb{E}[S_{\tau}], \quad (1.1)$$

where  $\{S_t\}_{t \geq 0}$  denotes the price process and  $\mathcal{T}_T$  denotes the set of stopping times with respect to the price process  $S$ .

Many popular continuous models for the price process are of the form

$$dS_t = \alpha S_t dt + \sigma(t) S_t dW_t, \quad (1.2)$$

---

\*Department of Information Technology, Uppsala University, Box 337, 751 05 Uppsala, Sweden.  
Email: juozas.vaicenavicius@it.uu.se

where  $\alpha \in \mathbb{R}$  is called the *drift*, and  $\sigma \geq 0$  is known as the *volatility process*. Imposing simplifying assumptions that the volatility is independent of  $W$  as well as time-homogeneous, an  $m$ -state time-homogeneous Markov chain stands out as a basic though still rather flexible stochastic volatility model (proposed in [10]), which we choose to use in this article. The flexibility comes from the fact that we can choose the state space as well as the transition intensities between the states.

Though the problem (1.1) in which  $S$  follows (1.2) is well-posed mathematically, from a financial point of view, the known drift assumption is widely accepted to be unreasonable (e.g. see [29, Section 4.2 on p. 144]) and needs to be relaxed. Hence, using the Bayesian paradigm, we model the initial uncertainty about the drift by a probability distribution (known as the *prior* in Bayesian inference), which incorporates all the available information about the parameter and its uncertainty (see [13] for more on the interpretation of the prior). If the quantification of initial uncertainty is subjective, then the prior represents one's beliefs about how likely the drift is to take different values. To be able to incorporate arbitrary prior beliefs, we set out to solve the optimal selling problem (1.1) under an arbitrary prior for the drift.

In the present paper, we analyse and solve the asset liquidation problem (1.1) in the case when  $S$  follows (1.2) with  $m$ -state time-homogeneous Markov chain volatility and unknown drift, the uncertainty of which is modelled by an arbitrary probability distribution. The first time a particular four-dimensional process hits a specific boundary determining the stopping set is shown to be optimal. This stopping boundary has attractive monotonicity properties and can be found using the approximation procedure developed.

Let us elucidate our study of the optimal selling problem in more depth. Using the nonlinear filtering theory, the original selling problem with parameter uncertainty is rewritten as an equivalent optimal stopping problem of a standard form (i.e. without unknown parameters). In this new optimal stopping problem, the posterior mean serves as the underlying process and acts as a stochastic creation rate; the payoff function in the problem is constant. The posterior mean is shown to be the solution of an SDE depending on the prior and the whole volatility history. Embedding of the optimal stopping problem into a Markovian framework is non-trivial because the whole posterior distribution needs to be included as a variable. Fortunately, we show that having fixed the prior, the posterior is fully characterised by only two real-valued parameters: the posterior mean and, what we call, the *effective learning time*. As a result, we are able to define an associated Markovian value function with four underlying variables (time, posterior mean, effective learning time, and volatility) and study the optimal stopping problem as a four-dimensional Markovian optimal stopping problem (the volatility takes values in a finite set, but slightly abusing terminology, we still call it a dimension). Exploiting that the volatility is constant between the regime switches, we construct  $m$  sequences of simpler auxiliary three-dimensional Markovian optimal stopping problems whose values in the limit converge monotonically to the true value function. We show that the value function is decreasing in time and effective learning time as well as increasing and convex

in posterior mean. The first hitting time of a region specified by a stopping boundary that is a function of time, effective learning time, and volatility is shown to be optimal. The stopping boundary is increasing in time, effective learning time, and is the limit of a monotonically increasing sequence of boundaries from the auxiliary problems. Moreover, the approximation procedure using the auxiliary problems yields a method to calculate the value function as well as the optimal stopping boundary numerically.

In the two-point prior case, the posterior mean fully characterises the posterior distribution, making the problem more tractable and allowing us to obtain some additional results. In particular, we prove that, under a skip-free volatility assumption, the Markovian value function is decreasing in the volatility and that the stopping boundary is increasing in the volatility.

In a broader mathematical context, the selling problem investigated appears to be the first optimal stopping problem with parameter uncertainty and stochastic volatility to be studied in the literature. Thus it is plausible that ideas presented herein will find uses in other optimal stopping problems of the same type; for example, in classical problems of Bayesian sequential analysis (e.g. see [27, Chapter VI]) with stochastically evolving noise magnitude. It is clear to the author that with additional efforts a number of results of the article can be refined or generalised. However, the objective chosen is to provide an intuitive understanding of the problem and the solution while still maintaining readability and clarity. This also explains why, for the most part, we focus on the two-point prior case and outline an extension to the general prior case only at the end.

## 1.1 Related literature

There is a strand of research on asset liquidation problems in models with regime-switching volatility, alas, they either concern only a special class of suboptimal strategies or treat the drift as observable. In [33], a restrictive asset liquidation problem was proposed and studied; the drift as well as the volatility were treated as unobservable and the possibility to learn about the parameters from the observations was disregarded. The subsequent papers [31], [32], [15] explored various aspects of the same formulation. An optimal selling problem with the payoff  $e^{-r\tau}(S_\tau - K)$  was studied in [23] for the Black-Scholes model, in [19] for a two-state regime-switching model, and in [32] for an  $m$ -state model with finite horizon. In all three cases, the drift and the volatility are assumed to be fully observable.

In another strand of research, the optimal stopping problem (1.1) has been solved and analysed in the Black-Scholes model under arbitrary uncertainty about the drift. The two-point prior case was studied in [11], while the general prior case was solved in [13] using a different approach. This article can be viewed as a generalisation of [13] to include stochastic regime-switching volatility. Related option valuation problems under incomplete information were studied in [16], [30], both in the two-point prior case, and in [9] in the  $n$ -point prior case.

The approach we take to approximate a Markovian value function by a sequence of value functions of simpler constant volatility problems was used before in [21] to investi-

gate a finite-horizon American put problem (also, its slight generalisation) in a regime-switching model with full information. Regrettably, in the case of 3 or more volatility states, the recursive approximation step in [21, Section 5] contains a blunder; we rectify it in Section 3.2 of this article.

Though it is true that the current paper is a generalisation of [13] from constant volatility to the regime-switching stochastic volatility model, the extension is definitely not a straightforward one. Novel statistical learning intuitions were needed, and new proofs were developed to arrive at the results of the paper. One of the main insights of the optimal liquidation problem with constant volatility in [13] was that the current time and price were sufficient statistics for the optimal selling problem. However, changing the volatility from constant to stochastic makes the posterior distribution of the drift truly dependent on the price path. This raises questions whether an optimal liquidation problem can be treated using the mainstream finite-dimensional Markovian techniques at all, and also whether any of the developments from the constant volatility case can be taken advantage of. In the two-point prior case with regime-switching volatility, the following new insight was key. Despite the posterior being a path-dependent function of the stock price, we can show that the current time, posterior mean and instantaneous volatility (extracted from the price process) are sufficient statistics for the optimal liquidation problem. Alas, for any prior with more than two points in the support, the same triplet is no longer a sufficient statistic. Fortunately, if in addition to the time-price-volatility triplet we introduce an additional statistic, which we name the effective learning time, the resulting 4-tuple becomes a sufficient statistic for the selling problem under a general prior. Besides these insights, some new technicalities (in particular, Lemma (2.3)) stemming from stochastic volatility had to be resolved to reformulate the optimal selling problem into the standard Markovian form.

In relation to [21], though we employ the same general iterative approximation idea to construct an approximating sequence for the Markovian value function, the particulars, including proofs and results, are notably distinct. Firstly, we work in a more general setting, proving and formulating more abstract as well as, in multiple instances, new type of results. For example, we prove things in the  $m$ -state rather than the two-state regime-switching model. This allowed us to catch and correct an erroneous construction of the approximating sequence in [21] for models with more than two volatility states. Moreover, almost all the proofs follow different arguments either because of the structural differences in the selling problem or because we prefer another way, which seems to be more transparent and direct, to arrive at the results. Lastly, many of the results in the present paper are problem-specific and even not depend on the iterative approximation of the value function after all.

The idea to iteratively construct a sequence of auxiliary value functions that converge to the true value function in the limit is generic and has been many times successfully applied to optimal stopping problems with a countable number of discrete events (e.g. jumps, discrete observations). In the setting with partial observations, an iterative approximation scheme was employed in [3] to study the Poisson disorder detection

problem with unknown post-disorder intensity, then later, in [8], to analyse a combined Poisson-Wiener disorder detection problem, and, more recently, in [4], to investigate the Wiener disorder detection under discrete observations. In the fully observable setting, such iterative approximations go back to at least as early as [17], which deals with a Markovian optimal stopping problem with a piecewise deterministic underlying. In Financial Mathematics, iteratively constructed approximations were used in [2] and [5] to study the value functions of finite and perpetual American put options, respectively, for a jump diffusion. Besides optimal stopping, the iterative approximation technique was utilised for the singular control problem [14] of optimal dividend policy.

## 2 Problem set-up

We model a financial market on a filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$  satisfying the usual conditions. Here the measure  $\mathbb{P}$  denotes the physical probability measure. The price process is modelled by

$$dS_t = X S_t dt + \sigma(t) S_t dW_t, \quad (2.1)$$

where  $X$  is a random variable having probability distribution  $\mu$ ,  $W$  is a standard Brownian motion, and  $\sigma$  is a time-homogeneous right-continuous  $m$ -state Markov chain with a generator  $\Lambda = (\lambda_{ij})_{1 \leq i, j \leq m}$  and taking values  $\sigma_m \geq \dots \geq \sigma_1 > 0$ . Moreover, we assume that  $X$ ,  $W$ , and  $\sigma$  are independent. Since the volatility can be estimated from the observations of  $S$  in an arbitrary short period of time (at least in theory), it is reasonable to assume that the volatility process  $\{\sigma(t)\}_{t \geq 0}$  is observable. Hence the available information is modelled by the filtration  $\mathbb{F}^{S, \sigma} = \left\{ \mathcal{F}_t^{S, \sigma} \right\}_{t \geq 0}$  generated by the processes  $S$  and  $\sigma$  and augmented by the null sets of  $\mathcal{F}$ . Note that the drift  $X$  and the random driver  $W$  are not directly observable.

The optimal selling problem that we are interested in is

$$V = \sup_{\tau \in \mathcal{T}_T^{S, \sigma}} \mathbb{E}[S_\tau], \quad (2.2)$$

where  $\mathcal{T}_T^{S, \sigma}$  denotes the set of  $\mathbb{F}^{S, \sigma}$ -stopping times that are smaller or equal to a pre-specified time horizon  $T > 0$ .

**Remark 2.1.** It is straightforward to include a discount factor  $e^{-r\tau}$  in (2.2). In fact, it simply corresponds to a shift of the prior distribution  $\mu$  in the negative direction by  $r$ .

Let  $l := \inf \text{supp}(\mu)$  and  $h := \sup \text{supp}(\mu)$ . It is easy to see that if  $l \geq 0$ , then it is optimal to stop at the terminal time  $T$ . Likewise, if  $h \leq 0$ , then stopping immediately, i.e. at time zero, is optimal. The rest of the article focuses on the remaining and most interesting case.

**Assumption 2.2.**  $l < 0 < h$ .

## 2.1 Equivalent reformulation under a measure change

Let us write  $\hat{X}_t := \mathbb{E}[X | \mathcal{F}_t^{S,\sigma}]$ . Then the process

$$\hat{W}_t := \int_0^t \frac{1}{\sigma(s)} (X - \hat{X}_s) ds + W_t,$$

called the innovation process, is an  $\mathbb{F}^{S,\sigma}$ -Brownian motion (see [1, Proposition 2.30 on p. 33]).

**Lemma 2.3.** *The volatility process  $\sigma$  and the innovation process  $\hat{W}$  are independent.*

*Proof.* Since  $X$ ,  $W$ , and  $\sigma$  are independent, we can think of  $(\Omega, \mathcal{F}, \mathbb{P})$  as a product space  $(\Omega_{X,W} \times \Omega_\sigma, \mathcal{F}_{X,W} \otimes \mathcal{F}_\sigma, \mathbb{P}_{X,W} \times \mathbb{P}_\sigma)$ . Let  $A, A' \in \mathcal{B}(\mathbb{R}^{[0,T]})$ . Then

$$\begin{aligned} \mathbb{P}(\hat{W} \in A, \sigma \in A') &= \int_{\Omega_{X,W} \times \Omega_\sigma} \mathbb{1}_{\{\hat{W}(\omega_{X,W}, \omega_\sigma) \in A, \sigma(\omega_\sigma) \in A'\}} d(\mathbb{P}_{X,W} \times \mathbb{P}_\sigma)(\omega_{X,W}, \omega_\sigma) \\ &= \int_{\Omega_\sigma} \int_{\Omega_{X,W}} \mathbb{1}_{\{\hat{W}(\omega_{X,W}, \omega_\sigma) \in A\}} \mathbb{1}_{\{\sigma(\omega_\sigma) \in A'\}} d\mathbb{P}_{X,W}(\omega_{X,W}) d\mathbb{P}_\sigma(\omega_\sigma) \\ &= \int_{\Omega_\sigma} \mathbb{1}_{\{\sigma(\omega_\sigma) \in A'\}} \int_{\Omega_{X,W}} \mathbb{1}_{\{\hat{W}(\omega_{X,W}, \omega_\sigma) \in A\}} d\mathbb{P}_{X,W}(\omega_{X,W}) d\mathbb{P}_\sigma(\omega_\sigma) \\ &= \int_{\Omega_\sigma} \mathbb{1}_{\{\sigma(\omega_\sigma) \in A'\}} \mathbb{P}_{X,W}(\hat{W}(\cdot, \omega_\sigma) \in A) d\mathbb{P}_\sigma(\omega_\sigma) \\ &= \mathbb{P}(\hat{W} \in A) \mathbb{P}_\sigma(\sigma \in A') \\ &= \mathbb{P}(\hat{W} \in A) \mathbb{P}(\sigma \in A'), \end{aligned} \tag{2.3}$$

where the penultimate equality is justified by the fact that, for any fixed  $\omega_\sigma$ , the innovation process  $\hat{W}(\cdot, \omega_\sigma)$  is a Brownian motion under  $\mathbb{P}_{X,W}$ . Hence from (2.3), the processes  $\hat{W}$  and  $\sigma$  are independent.  $\square$

Defining a new equivalent measure  $\tilde{\mathbb{P}} \sim \mathbb{P}$  on  $(\Omega, \mathcal{F}_T)$  via the Radon-Nikodym derivative

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = e^{\int_0^T \sigma(t) d\hat{W}_t - \frac{1}{2} \int_0^T \sigma(t)^2 dt}$$

and writing

$$\begin{aligned} S_t &= S_0 e^{Xt + \int_0^t \sigma(s) dW_s - \frac{1}{2} \int_0^t \sigma(s)^2 ds} \\ &= S_0 e^{\int_0^t \hat{X}_s ds + \int_0^t \sigma(s) d\hat{W}_s - \frac{1}{2} \int_0^t \sigma(s)^2 ds}, \end{aligned}$$

we have that, for any  $\tau \in \mathcal{T}_T^{S,\sigma}$ ,

$$\mathbb{E}[S_\tau] = \tilde{\mathbb{E}} \left[ S_0 e^{\int_0^\tau \hat{X}_s ds} \right] = S_0 \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s ds} \right].$$

Moreover, by Girsanov's theorem, the process  $B_t := -\int_0^t \sigma(s) ds + \hat{W}_t$  is a  $\tilde{\mathbb{P}}$ -Brownian motion on  $[0, T]$ . In addition, Lemma 2.3 together with [1, Proposition 3.13] tells us that

the law of  $\sigma$  is the same under  $\tilde{\mathbb{P}}$  and  $\mathbb{P}$ , as well as that  $B$  and  $\sigma$  are independent under  $\tilde{\mathbb{P}}$ .

Without loss of generality, we set  $S_0 = 1$  throughout the article, so the optimal stopping problem (2.2) can be cast as

$$V = \sup_{\tau \in \mathcal{T}_T^{S,\sigma}} \tilde{\mathbb{E}}[e^{\int_0^\tau \hat{X}_s ds}]. \quad (2.4)$$

Between the volatility jumps, the stock price is a geometric Brownian motion with known constant volatility and unknown drift. Hence, by Corollary 3.4 in [13], we have that  $\mathbb{F}^{S,\sigma} = \mathbb{F}^{\hat{X},\sigma}$  and  $\mathcal{T}_T^{S,\sigma} = \mathcal{T}_T^{\hat{X},\sigma}$ , where  $\mathbb{F}^{\hat{X},\sigma}$  denotes the usual augmentation of the filtration generated by  $\hat{X}$  and  $\sigma$ , also,  $\mathcal{T}_T^{\hat{X},\sigma}$  denotes the set of  $\mathbb{F}^{\hat{X},\sigma}$ -stopping times not exceeding  $T$ . As a result, an equivalent reformulation of (2.4) is

$$V = \sup_{\tau \in \mathcal{T}_T^{\hat{X},\sigma}} \tilde{\mathbb{E}}[e^{\int_0^\tau \hat{X}_s ds}], \quad (2.5)$$

which we will study in the subsequent parts of the article.

## 2.2 Markovian embedding

In all except the last section of this article, we will focus on the special case when  $X$  has a two-point distribution  $\mu = \pi\delta_h + (1 - \pi)\delta_l$ , where  $h > l$ ,  $\pi \in (0, 1)$  are constants, and  $\delta_h, \delta_l$  are Dirac measures at  $h$  and  $l$ , respectively. In this special case, expressions are simpler and arguments are easier to follow than in the general prior case; still, most underlying ideas of the arguments are the same. Hence, we choose to understand the two-point prior case first, after which generalising the results to the general prior case will become a rather easy task.

Since the volatility is a known constant between the jump times, using the dynamics of  $\hat{X}$  in the constant volatility case (the equation (3.9) in [13]), the process  $\hat{X}$  is a unique strong solution of

$$d\hat{X}_t = \sigma(t)\phi(\hat{X}_t, \sigma(t)) dt + \phi(\hat{X}_t, \sigma(t)) dB_t, \quad (2.6)$$

where

$$\phi(x, \sigma) := \frac{1}{\sigma}(h - x)(x - l).$$

Now, we can embed the optimal stopping problem (2.4) into a Markovian framework by defining a Markovian value function

$$v(t, x, \sigma) := \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}}[e^{\int_0^\tau \hat{X}_s^{t,x,\sigma} ds}], \quad (t, x, \sigma) \in [0, T] \times (l, h) \times \{\sigma_1, \dots, \sigma_m\}. \quad (2.7)$$

Here  $\hat{X}^{t,x,\sigma}$  denotes the process  $\hat{X}$  in (2.6) started at time  $t$  with  $\hat{X}_t = x$ ,  $\sigma(t) = \sigma$ , and  $\mathcal{T}_{T-t}$  stands for the set of stopping times less or equal to  $T - t$  with respect to the

usual augmentation of the filtration generated by  $\{\hat{X}_{t+s}^{t,x,\sigma}\}_{s \geq 0}$  and  $\{\sigma(t+s)\}_{s \geq 0}$ . The formulation (2.7) has an interpretation of an optimal stopping problem with the constant payoff 1 and the discount rate  $-\hat{X}_s$ ; from now onwards, we will study this discounted problem. The notation  $v_i := v(\cdot, \cdot, \sigma_i)$  will often be used.

### 3 Approximation procedure

It is not clear how to compute  $v$  in (2.7) or analyse it directly. Hence, in this section, we develop a way to approximate the value function  $v$  by a sequence of value functions, corresponding to simpler constant volatility optimal stopping problems.

#### 3.1 Operator $J_i$

For the succinctness of notation, let  $\lambda_i := \sum_{j \neq i} \lambda_{ij}$  denote the total intensity with which the volatility jumps from state  $\sigma_i$ . Also, let us define

$$\eta_i^t := \inf\{s > 0 \mid \sigma(t+s) \neq \sigma(t) = \sigma_i\},$$

which is an  $\text{Exp}(\lambda_i)$ -distributed random variable representing the duration up to the first volatility change if started from the volatility state  $\sigma_i$  at time  $t$ .

Furthermore, let us define an operator  $J$  acting on a bounded  $f : [0, T] \times (l, h) \rightarrow \mathbb{R}$  by

$$(Jf)(t, x, \sigma_i) := \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_i^t\}} + e^{\int_0^{\eta_i^t} \hat{X}_{t+s}^{t,x,\sigma_i} ds} f(t + \eta_i^t, \hat{X}_{t+\eta_i^t}^{t,x,\sigma_i}) \mathbb{1}_{\{\tau \geq \eta_i^t\}} \right] \quad (3.1)$$

$$= \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t,x,\sigma_i} - \lambda_i ds} + \lambda_i \int_0^\tau e^{\int_0^u \hat{X}_{t+s}^{t,x,\sigma_i} - \lambda_i ds} f(t + u, \hat{X}_{t+u}^{t,x,\sigma_i}) du \right], \quad (3.2)$$

where  $\mathcal{T}_{T-t}$  denotes the set of stopping times less or equal to  $T - t$  with respect to the usual augmentation of the filtration generated by  $\{\hat{X}_{t+s}^{t,x,\sigma_i}\}_{s \geq 0}$  and  $\{\sigma(t+s)\}_{s \geq 0}$ . To simplify notation, we also define an operator  $J_i$  by

$$J_i f := (Jf)(\cdot, \cdot, \sigma_i).$$

Intuitively,  $(J_i f)$  represents a Markovian value function corresponding to optimal stopping before  $t + \eta_i^t$ , i.e. before the first volatility change after  $t$ , when, at time  $t + \eta_i^t < T$ , the payoff  $f(t + \eta_i^t, \hat{X}_{t+\eta_i^t}^{t,x,\sigma_i})$  is received provided stopping has not occurred yet.

**Proposition 3.1.** *Let  $f : [0, T] \times (l, h) \rightarrow \mathbb{R}$  be bounded. Then*

- (i)  $Jf$  is bounded;
- (ii)  $f$  increasing in the second variable  $x$  implies that  $Jf$  is increasing in the second variable  $x$ ;

- (iii)  $f$  decreasing in the first variable  $t$  implies that  $Jf$  is decreasing in the first variable  $t$ ;
- (iv)  $f$  increasing and convex in the second variable  $x$  implies that  $Jf$  is increasing and convex in the second variable  $x$ ;
- (v)  $J$  preserves order, i.e.  $f_1 \leq f_2$  implies  $Jf_1 \leq Jf_2$ ;
- (vi)  $Jf \geq 1$ .

*Proof.* All except claim (iv) are straightforward consequences of the representation (3.2). To prove (iv), we will approximate the optimal stopping problem (3.2) by Bermudan options.

Let  $i$  and  $n$  be fixed. We will approximate the value function  $J_i f$  by a value function  $w_{i,n}^{(f)}$  of a corresponding Bermudan problem with stopping allowed only at times  $\{\frac{kT}{2^n} : k \in \{0, 1, \dots, 2^n\}\}$ . We define  $w_{i,n}^{(f)}$  recursively as follows. First,

$$w_{i,n}^{(f)}(T, x) := 1.$$

Then, starting with  $k = 2^n$  and continuing recursively down to  $k = 1$ , we define

$$w_{i,n}^{(f)}(t, x) = \begin{cases} g(t, x, \frac{kT}{2^n}), & t \in (\frac{(k-1)T}{2^n}, \frac{kT}{2^n}), \\ g(\frac{(k-1)T}{2^n}, x, \frac{kT}{2^n}) \vee 1, & t = \frac{(k-1)T}{2^n}, \end{cases} \quad (3.3)$$

where the function  $g$  is given by

$$g(t, x, \frac{kT}{2^n}) := \tilde{\mathbb{E}} \left[ e^{\int_t^{\frac{kT}{2^n}} \hat{X}_s^{t,x,\sigma_i} - \lambda_i ds} w_{i,n}^{(f)} \left( \frac{kT}{2^n}, \hat{X}_{\frac{kT}{2^n}}^{t,x,\sigma_i} \right) + \int_t^{\frac{kT}{2^n}} e^{\int_t^u \hat{X}_s^{t,x,\sigma_i} - \lambda_i ds} f(u, \hat{X}_u^{t,x,\sigma_i}) du \right]. \quad (3.4)$$

Next, we show by backward induction on  $k$  that  $w_{i,n}^{(f)}$  is increasing and convex in the second variable  $x$ . Suppose that for some  $k \in \{1, 2, \dots, 2^n\}$ , the function  $w_{i,n}^{(f)}(\frac{kT}{2^n}, \cdot)$  is increasing and convex (the assumption clearly holds for the base step  $k = 2^n$ ). Let  $t \in [\frac{(k-1)T}{2^n}, \frac{kT}{2^n})$ . Then, since  $f$  is also increasing and convex in the second variable  $x$ , we have that the function  $g(t, \cdot, \frac{kT}{2^n})$ , and so  $w_{i,n}^{(f)}(t, \cdot)$ , is convex by [12, Theorem 5.1]. Moreover, from (3.4) and [28, Theorem IX.3.7], it is clear that  $w_{i,n}^{(f)}(t, \cdot)$  is increasing. Consequently, by backward induction, we obtain that the Bermudan value function  $w_{i,n}^{(f)}$  is increasing and convex in the second variable.

Letting  $n \nearrow \infty$ , the Bermudan value  $w_{i,n}^{(f)} \nearrow J_i f$  pointwise. As a result,  $J_i f$  is increasing and convex in the second argument, since convexity and monotonicity are preserved when taking pointwise limits.  $\square$

The sets

$$\begin{aligned}\mathcal{C}_i^f &:= \{(t, x) \in [0, T] \times (l, h) : (J_i f)(t, x) > 1\}, \\ \mathcal{D}_i^f &:= \{(t, x) \in [0, T] \times (l, h) : (J_i f)(t, x) = 1\} = [0, T] \times (l, h) \setminus \mathcal{C}_i^f,\end{aligned}\tag{3.5}$$

correspond to continuation and stopping sets for the stopping problem  $J_i f$  as the next proposition shows.

**Proposition 3.2** (Optimal stopping time). *The stopping time*

$$\tau_{\sigma_i}^f(t, x) = \inf\{u \in [0, T - t] : (t + u, \hat{X}_{t+u}^{t, x, \sigma_i}) \in \mathcal{D}_i^f\}\tag{3.6}$$

is optimal for the problem (3.2).

*Proof.* A standard application of Theorem D.12 in [20].  $\square$

**Proposition 3.3.** *If a bounded  $f : [0, T] \times (l, h) \rightarrow \mathbb{R}$  is decreasing in the first variable as well as increasing and convex in the second, then  $J_i f$  is continuous.*

*Proof.* The argument is a trouble-free extension of the proof of the third part of Theorem 3.10 in [13]; still, we include it for completeness. Before we begin, in order to simplify notation, we will write  $u := J_i f$ .

Firstly, we let  $r \in (l, h)$  and will prove that there exists  $K > 0$  such that, for every  $t \in [0, T]$ , the map  $x \mapsto J_i f(t, x)$  is  $K$ -Lipschitz continuous on  $(l, r]$ . To obtain a contradiction, assume that there is no such  $K$ . Then, by convexity of  $u$  in the second variable, there is a sequence  $\{t_n\}_{n \geq 0} \subset [0, T]$  such that the left-derivatives  $\partial_2^- u(t_n, r) \nearrow \infty$ . Hence, for  $r' \in (r, h)$ , the sequence  $u(t_n, r') \rightarrow \infty$ , which contradicts that  $u(t_n, r') \leq u(0, r') < \infty$  for all  $n \in \mathbb{N}$ .

Now, it remains to show that  $u$  is continuous in time. Assume for a contradiction that the map  $t \mapsto u(t, x_0)$  is not continuous at  $t = t_0$  for some  $x_0$ . Since  $u$  is decreasing in time,  $u(\cdot, x_0)$  has a negative jump at  $t_0$ . Next, we will investigate the cases  $u(t_0-, x_0) > u(t_0, x_0)$  and  $u(t_0, x_0) > u(t_0+, x_0)$  separately.

Suppose  $u(t_0-, x_0) > u(t_0, x_0)$ . By Lipschitz continuity in the second variable, there exists  $\delta > 0$  such that, writing  $\mathcal{R} = (t_0 - \delta, t_0) \times (x_0 - \delta, x_0 + \delta)$ ,

$$\inf_{(t, x) \in \mathcal{R}} u(t, x) > u(t_0, x_0 + \delta).\tag{3.7}$$

Thus  $\mathcal{R} \subseteq \mathcal{C}_i^f$ . Let  $t \in (t_0 - \delta, t_0)$  and  $\tau_{\mathcal{R}} := \inf\{s \geq 0 : (t + s, \hat{X}_{t+s}^{t, x, \sigma_i}) \notin \mathcal{R}\}$ . Then,

by the martingality in the continuation region,

$$\begin{aligned}
u(t, x_0) &= \tilde{\mathbb{E}} \left[ e^{\int_0^{\tau_{\mathcal{R}}} \hat{X}_{t+u}^{t, x_0, \sigma_i} - \lambda_i \, du} u(t + \tau_{\mathcal{R}}, \hat{X}_{t+\tau_{\mathcal{R}}}^{t, x_0, \sigma_i}) \right. \\
&\quad \left. + \int_0^{\tau_{\mathcal{R}}} e^{\int_0^u \hat{X}_{t+s}^{t, x_0, \sigma_i} - \lambda_i \, ds} f(t + u, \hat{X}_{t+u}^{t, x_0, \sigma_i}) \, du \right] \\
&\leq \tilde{\mathbb{E}} \left[ e^{(t_0-t)(x_0+\delta)^+} u(t, x_0 + \delta) \mathbb{1}_{\{t+\tau_{\mathcal{R}} < t_0\}} \right. \\
&\quad \left. + e^{(t_0-t)(x_0+\delta)^+} u(t_0, x_0 + \delta) \mathbb{1}_{\{t+\tau_{\mathcal{R}} = t_0\}} \right. \\
&\quad \left. + \int_0^{t_0-t} e^{\int_0^u \hat{X}_{t+s}^{t, x_0, \sigma_i} - \lambda_i \, ds} |f(t + u, \hat{X}_{t+u}^{t, x_0, \sigma_i})| \, du \right] \\
&\leq e^{(t_0-t)(x_0+\delta)^+} u(t, x_0 + \delta) \tilde{\mathbb{P}}(t + \tau_{\mathcal{R}} < t_0) + e^{(t_0-t)(x_0+\delta)^+} u(t_0, x_0 + \delta) \\
&\quad + \int_0^{t_0-t} \tilde{\mathbb{E}} \left[ e^{\int_0^u \hat{X}_{t+s}^{t, x_0, \sigma_i} - \lambda_i \, ds} |f(t + u, \hat{X}_{t+u}^{t, x_0, \sigma_i})| \right] \, du \\
&\rightarrow u(t_0, x_0 + \delta)
\end{aligned}$$

as  $t \rightarrow t_0$ , contradicting (3.7).

The other case to consider is  $u(t_0, x_0) > u(t_0+, x_0)$ ; we look into the situation  $u(t_0, x_0) > u(t_0+, x_0) > 1$  first. The local Lipschitz continuity in the second variable and the decay in the first variable imply that there exist  $\epsilon > 0$  and  $\delta > 0$  such that, writing  $\mathcal{R} = (t_0, t_0 + \epsilon] \times [x_0 - \delta, x_0 + \delta]$ ,

$$u(t_0, x_0) > \sup_{(t,x) \in \mathcal{R}} u(t, x) \geq \inf_{(t,x) \in \mathcal{R}} u(t, x) > 1. \quad (3.8)$$

Hence,  $\mathcal{R} \subseteq \mathcal{C}_i^f$  and writing  $\tau_{\mathcal{R}} := \inf\{s \geq 0 : (t_0 + s, \hat{X}_{t_0+s}^{t_0, x_0, \sigma_i}) \notin \mathcal{R}\}$  we have

$$\begin{aligned}
u(t_0, x_0) &= \tilde{\mathbb{E}} \left[ e^{\int_0^{\tau_{\mathcal{R}}} \hat{X}_{t_0+u}^{t_0, x_0, \sigma_i} - \lambda_i \, du} u(t_0 + \tau_{\mathcal{R}}, \hat{X}_{t_0+\tau_{\mathcal{R}}}^{t_0, x_0, \sigma_i}) \right. \\
&\quad \left. + \int_0^{\tau_{\mathcal{R}}} e^{\int_0^u \hat{X}_{t_0+s}^{t_0, x_0, \sigma_i} - \lambda_i \, ds} f(t_0 + u, \hat{X}_{t_0+u}^{t_0, x_0, \sigma_i}) \, du \right] \\
&\leq \tilde{\mathbb{E}} \left[ e^{\epsilon(x_0+\delta)^+} u(t_0, x_0 + \delta) \mathbb{1}_{\{\tau_{\mathcal{R}} < \epsilon\}} \right] \\
&\quad + \tilde{\mathbb{E}} \left[ e^{\epsilon(x_0+\delta)^+} u(t_0 + \epsilon, x_0 + \delta) \mathbb{1}_{\{\tau_{\mathcal{R}} = \epsilon\}} \right. \\
&\quad \left. + \int_0^{\epsilon} e^{\int_0^u \hat{X}_{t_0+s}^{t_0, x_0, \sigma_i} - \lambda_i \, ds} |f(t_0 + u, \hat{X}_{t_0+u}^{t_0, x_0, \sigma_i})| \, du \right] \\
&\leq e^{\epsilon(x_0+\delta)^+} u(t_0, x_0 + \delta) \tilde{\mathbb{P}}(\tau_{\mathcal{R}} < \epsilon) + e^{\epsilon(x_0+\delta)^+} u(t_0 + \epsilon, x_0 + \delta) \\
&\quad + \int_0^{\epsilon} \tilde{\mathbb{E}} \left[ e^{\int_0^u \hat{X}_{t_0+s}^{t_0, x_0, \sigma_i} - \lambda_i \, ds} |f(t_0 + u, \hat{X}_{t_0+u}^{t_0, x_0, \sigma_i})| \right] \, du \\
&\rightarrow u(t_0+, x_0 + \delta)
\end{aligned}$$

as  $\epsilon \searrow 0$ , which contradicts (3.8).

Lastly, suppose that  $u(t_0, x_0) > u(t_0+, x_0) = 1$ . By Lipschitz continuity in the second

variable, there exists  $\delta > 0$  such that

$$\inf_{x \in (x_0 - \delta, x_0)} u(t_0, x) > u(t_0, x_0) = 1. \quad (3.9)$$

Consequently,  $(t_0, T] \times (x_0 - \delta, x_0) \subseteq \mathcal{D}_i^f$ . Hence the process  $\hat{X}^{t_0, x_0 - \delta/2, \sigma_i}$  hits the stopping region immediately and so  $(t_0, x_0 - \delta/2) \in \mathcal{D}_i^f$ , which contradicts (3.9).  $\square$

**Proposition 3.4** (Optimal stopping boundary).

Let  $f : [0, T] \times (l, h) \rightarrow \mathbb{R}$  be bounded, decreasing in the first variable as well as increasing and convex in the second variable. Then the following hold.

(i) There exists a function  $b_{\sigma_i}^f : [0, T] \rightarrow [l, h]$  that is both increasing, right-continuous with left limits, and satisfies

$$\mathcal{C}_i^f = \{(t, x) \in [0, T] \times (l, h) : x > b_{\sigma_i}^f(t)\}. \quad (3.10)$$

(ii) The pair  $(J_i f, b_{\sigma_i}^f)$  satisfies the free-boundary problem

$$\begin{cases} \partial_t u(t, x) + \sigma_i \phi(x, \sigma_i) \partial_x u(t, x) + \frac{1}{2} \phi(x, \sigma_i)^2 \partial_{xx} u(t, x) \\ \quad + (x - \lambda_i) u(t, x) + \lambda_i f(t, x) = 0, & \text{if } x > b_{\sigma_i}^f(t), \\ u(t, x) = 1, & \text{if } x \leq b_{\sigma_i}^f(t) \text{ or } t = T. \end{cases} \quad (3.11)$$

*Proof.* (i) By Proposition 3.1 (iv), there exists a unique function  $b_{\sigma_i}^f$  satisfying (3.10). Moreover, by Proposition 3.1 (iii), this boundary  $b_{\sigma_i}^f$  is increasing. Hence, using Proposition 3.3, we also obtain that  $b_{\sigma_i}^f$  is right-continuous with left limits.

(ii) The proof follows a well-known standard argument (e.g. see [20, Theorem 7.7 in Chapter 2]), thus we omit it.  $\square$

## 3.2 A sequence of approximating problems

Let us define a sequence of stopping times  $\{\xi_n^t\}_{n \geq 0}$  recursively by

$$\begin{aligned} \xi_0^t &:= 0, \\ \xi_n^t &:= \inf\{s > \xi_{n-1}^t : \sigma(t+s) \neq \sigma(t+\xi_{n-1}^t)\}, \quad n > 0. \end{aligned}$$

Here  $\xi_n^t$  represents the duration until the  $n$ -th volatility jump since time  $t$ . Furthermore, let us define a sequence of operators  $\{J^{(n)}\}_{n \geq 0}$  by

$$(J^{(n)} f)(t, x, \sigma_i) := \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t, x, \sigma_i} ds} \mathbb{1}_{\{\tau < \xi_n^t\}} + e^{\int_0^{\xi_n^t} \hat{X}_{t+s}^{t, x, \sigma_i} ds} f(t + \xi_n^t, \hat{X}_{t+\xi_n^t}^{t, x, \sigma_i}) \mathbb{1}_{\{\tau \geq \xi_n^t\}} \right], \quad (3.12)$$

where  $f : [0, T] \times (l, r) \rightarrow \mathbb{R}$  is bounded. In particular, note that  $J^{(0)}f = f$  and  $J^{(1)}f = Jf$ . Similarly as for the operator  $J$ , we define  $J_i^{(n)}$  by

$$J_i^{(n)}f := (J^{(n)}f)(\cdot, \cdot, \sigma_i).$$

**Proposition 3.5.** *Let  $n \geq 0$  and  $i \in \{0, \dots, m\}$ . Then*

$$J_i^{(n+1)} = J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j^{(n)} \right). \quad (3.13)$$

*Proof.* The proof is by induction. In order to present the argument of the proof while keeping intricate notation at bay, we will only prove that, for a bounded  $f : [0, T] \times (l, h) \rightarrow \mathbb{R}$  and  $x \in (l, h)$ , the identity  $(J_i^{(2)}f)(t, x) = (J_i(\sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j f))(t, x)$  holds. The induction step  $J_i^{(n+1)} = J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j^{(n)} \right)$  follows a similar argument, though with more abstract notation. Note that without loss of generality, we can assume  $t = 0$ , which we do.

Firstly, we will show  $(J_i^{(2)}f)(0, x) \leq J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} (J_j f) \right)(0, x)$  and then the opposite inequality. For  $j \in \mathbb{N}$ , we will write  $\xi_j$  instead of  $\xi_j^0$  as well as will use the notation  $\eta_j := \xi_j - \xi_{j-1}$ . Let  $\tau \in \mathcal{T}_T$  and consider

$$\begin{aligned} A(\tau) &:= \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_1\}} + e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\eta_1 \leq \tau < \xi_2\}} + e^{\int_0^{\xi_2} \hat{X}_s^{0,x,\sigma_i} ds} f(\xi_2, \hat{X}_{\xi_2}^{0,x,\sigma_i}) \mathbb{1}_{\{\tau \geq \xi_2\}} \right] \\ &= \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_1\}} + \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\eta_1 \leq \tau < \xi_2\}} \right. \right. \\ &\quad \left. \left. + e^{\int_0^{\xi_2} \hat{X}_s^{0,x,\sigma_i} ds} f(\xi_2, \hat{X}_{\xi_2}^{0,x,\sigma_i}) \mathbb{1}_{\{\tau \geq \xi_2\}} \mid \mathcal{F}_{\eta_1}^{\hat{X}^{0,x,\sigma_i}, N} \right] \right], \end{aligned} \quad (3.14)$$

where  $\{N_t\}_{t \geq 0}$  denotes the process counting the volatility jumps. The inner conditional expectation in (3.14) satisfies

$$\begin{aligned} &\tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\eta_1 \leq \tau < \xi_2\}} + e^{\int_0^{\xi_2} \hat{X}_s^{0,x,\sigma_i} ds} f(\xi_2, \hat{X}_{\xi_2}^{0,x,\sigma_i}) \mathbb{1}_{\{\tau \geq \xi_2\}} \mid \mathcal{F}_{\eta_1}^{\hat{X}^{0,x,\sigma_i}, N} \right] \\ &= e^{\int_0^{\eta_1} \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\eta_1 \leq \tau\}} \tilde{\mathbb{E}} \left[ e^{\int_{\eta_1}^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \xi_2\}} \right. \\ &\quad \left. + e^{\int_{\eta_1}^{\xi_2} \hat{X}_s^{0,x,\sigma_i} ds} f(\xi_2, \hat{X}_{\xi_2}^{0,x,\sigma_i}) \mathbb{1}_{\{\tau \geq \xi_2\}} \mid \mathcal{F}_{\eta_1}^{\hat{X}^{0,x,\sigma_i}, N} \right] \\ &= e^{\int_0^{\eta_1} \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\eta_1 \leq \tau\}} \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} \tilde{\mathbb{E}}_{\eta_1, \hat{X}_{\eta_1}^{0,x,\sigma_i}, \sigma_j} \left[ e^{\int_0^{\tilde{\tau}} \hat{X}_{\eta_1+s} ds} \mathbb{1}_{\{\tilde{\tau} < \eta_2\}} \right. \\ &\quad \left. + e^{\int_0^{\eta_2} \hat{X}_{\eta_1+s} ds} f(\eta_1 + \eta_2, \hat{X}_{\eta_1+\eta_2}) \mathbb{1}_{\{\tilde{\tau} \geq \eta_2\}} \right], \end{aligned} \quad (3.15)$$

where  $\tilde{\tau} = \tau - \eta_1$  in the case  $\eta_1 \leq \tau \leq T$ . Therefore, substituting (3.15) into (3.14) and

then taking a supremum over  $\tilde{\tau}$ , we get

$$\begin{aligned}
A(\tau) &\leq \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_1\}} \right. \\
&\quad + e^{\int_0^{\eta_1} \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau \geq \eta_1\}} \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} \sup_{\tilde{\tau} \in \mathcal{T}_{T-T \wedge \eta_1}} \tilde{\mathbb{E}}^{\eta_1, \hat{X}_{\eta_1}^{0,x,\sigma_i}, \sigma_j} \left[ e^{\int_0^{\tilde{\tau}} \hat{X}_{\eta_1+s} ds} \mathbb{1}_{\{\tilde{\tau} < \eta_2\}} \right. \\
&\quad \left. \left. + e^{\int_0^{\eta_2} \hat{X}_{\eta_1+s} ds} f(\eta_1 + \eta_2, \hat{X}_{\eta_1+\eta_2}) \mathbb{1}_{\{\tilde{\tau} \geq \eta_2\}} \right] \right] \\
&= \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_1\}} + e^{\int_0^{\eta_1} \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau \geq \eta_1\}} \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} (J_j f)(\eta_1, X_{\eta_1}^{0,x,\sigma_i}) \right] \quad (3.16)
\end{aligned}$$

Taking a supremum over  $\tau$  in (3.16), we obtain

$$(J_i^{(2)} f)(0, x) = \sup_{\tau \in \mathcal{T}_T} A(\tau) \leq J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} (J_j f) \right) (0, x). \quad (3.17)$$

It remains to establish the opposite inequality. Let  $\tau \in \mathcal{T}_T$  and define

$$\check{\tau} := \tau \mathbb{1}_{\{\tau \leq \eta_1\}} + (\eta_1 \wedge T + \tau_{\sigma(\eta_1)}) \mathbb{1}_{\{\tau > \eta_1\}}, \quad (3.18)$$

where  $\tau_{\sigma(\eta_1)} := \tau_{\sigma(\eta_1)}^f(\eta_1 \wedge T, \hat{X}_{\eta_1 \wedge T}^{0,x,\sigma_i})$ . Clearly,  $\check{\tau} \in \mathcal{T}_T$ . Then

$$\begin{aligned}
&(J_1^{(2)} f)(0, x) \\
&\geq A(\check{\tau}) \\
&= \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_1\}} + e^{\int_0^{\eta_1} \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau \geq \eta_1\}} \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} \tilde{\mathbb{E}}^{\eta_1, \hat{X}_{\eta_1}^{0,x,\sigma_i}, \sigma_j} \left[ e^{\int_0^{\tau_{\sigma_j}} \hat{X}_{\eta_1+s} ds} \mathbb{1}_{\{\tau_{\sigma_j} < \eta_2\}} \right. \right. \\
&\quad \left. \left. + e^{\int_0^{\eta_2} \hat{X}_{\eta_1+s} ds} f(\eta_1 + \eta_2, \hat{X}_{\eta_1+\eta_2}) \mathbb{1}_{\{\tau_{\sigma_j} \geq \eta_2\}} \right] \right] \\
&= \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_1\}} + e^{\int_0^{\eta_1} \hat{X}_s^{0,x,\sigma_i} ds} \mathbb{1}_{\{\tau \geq \eta_1\}} \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} (J_j f)(\eta_1, \hat{X}_{\eta_1}^{0,x,\sigma_i}) \right],
\end{aligned}$$

where Proposition 3.2 was used to obtain the last equality. Hence, by taking supremum over stopping times  $\tau \in \mathcal{T}_T$ , we get

$$(J_i^{(2)} f)(0, x) \geq J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} (J_j f) \right) (0, x). \quad (3.19)$$

Finally, (3.17) and (3.19) taken together imply

$$(J_i^{(2)} f)(0, x) = J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} (J_j f) \right) (0, x).$$

□

**Remark 3.6.** In [21], the authors use the same approximation procedure for an optimal

stopping problem with regime switching volatility as in this article. Unfortunately, a mistake is made in equation (18) of [21], which wrecks the subsequent approximation procedure when the number of volatility states is greater than 2. The identity (18) therein should be replaced by (3.13).

### 3.3 Convergence to the value function

**Proposition 3.7** (Properties of the approximating sequence).

- (i) The sequence of functions  $\{J^{(n)}1\}_{n \geq 0}$  is increasing, bounded from below by 1 and from above by  $e^{hT}$ .
- (ii) Every  $J^{(n)}1$  is decreasing in the first variable  $t$  as well as increasing and convex in the second variable  $x$ .
- (iii) The sequence of functions

$$J^{(n)}1 \nearrow v \quad \text{pointwise as } n \nearrow \infty.$$

Moreover, the approximation error

$$\|v - J^{(n)}1\|_\infty \leq e^{hT} \lambda T \frac{(\lambda T)^{n-1}}{(n-1)!} \text{ as } n \rightarrow \infty, \quad (3.20)$$

where  $\lambda := \max\{\lambda_i : 1 \leq i \leq m\}$ .

- (iv) For every  $n \in \mathbb{N} \cup \{0\}$ ,

$$J_m^n 1 \leq J^{(n)}1 \leq J_1^n 1. \quad (3.21)$$

*Proof.* (i) The statement that  $\{J_i^{(n)}1\}_{n \geq 0}$  is increasing, bounded from below by 1 and from above by  $e^{hT}$  is a direct consequence of the definition (3.12).

- (ii) The claim that every  $J_i^{(n)}1$  is decreasing in the first variable  $t$  as well as increasing and convex in the second variable  $x$  follows by a straightforward induction on  $n$ , using Proposition 3.1 (iii),(iv) and Proposition 3.5 at the induction step.

- (iii) First, let  $i \in \{1, \dots, m\}$  and note that, for any  $n \in \mathbb{N}$ ,

$$J_i^{(n)}1 \leq v_i.$$

Here the inequality holds by suboptimality, since  $J_i^{(n)}1$  corresponds to an expected payoff of a particular stopping time in the problem (2.4). Next, define

$$U_n^{(i)}(t, x) := \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t,x,\sigma_i} ds} \mathbb{1}_{\{\tau < \xi_n^t\}} \right].$$

Then

$$U_n^{(i)}(t, x) \leq (J_i^{(n)}1)(t, x) \leq v_i(t, x) \leq U_n^{(i)}(t, x) + e^{h(T-t)}\mathbb{P}(\xi_n^t \leq T-t). \quad (3.22)$$

Since it is a standard fact that the  $n^{\text{th}}$  jump time, call it  $\zeta_n$ , of a Poisson process with jump intensity  $\lambda := \max\{\lambda_i : 1 \leq i \leq m\}$  follows the Erlang distribution, we have

$$\begin{aligned} \mathbb{P}(\xi_n^t \leq T-t) &\leq \mathbb{P}(\zeta_n \leq T-t) \\ &= \frac{1}{(n-1)!} \int_0^{\lambda(T-t)} u^{n-1} e^{-u} du \\ &\leq \lambda T \frac{(\lambda T)^{n-1}}{(n-1)!}. \end{aligned}$$

Therefore, by (3.22),

$$\|v - J^{(n)}1\|_\infty \leq e^{hT} \lambda T \frac{(\lambda T)^{n-1}}{(n-1)!} \text{ as } n \rightarrow \infty.$$

- (iv) The string of inequalities (3.21) will be proved by induction. First, the base step is obvious. Now, suppose (3.21) holds for some  $n \geq 0$ . Hence, for any  $i \in \{1, \dots, m\}$ ,

$$J_m^n 1 \leq \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j^{(n)} 1 \leq J_1^n 1. \quad (3.23)$$

Let us fix  $i \in \{1, \dots, m\}$ . By Proposition 3.1 (iv), every function in (3.23) is convex in the spatial variable  $x$ , thus [12, Theorem 6.1] yields

$$J_m^{n+1} 1 \leq J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j^{(n)} 1 \right) \leq J_1^{n+1} 1.$$

As  $i$  was arbitrary, we also have

$$J_{\sigma_m}^{n+1} 1 \leq J^{(n+1)} 1 \leq J_{\sigma_1}^{n+1} 1. \quad (3.24)$$

□

**Remark 3.8.** If instead of 1 we choose the constant function  $e^{hT}$  to apply the operators  $J_i^{(n)}$  to, then, following the same strategy as above,  $\{J_i^{(n)} e^{hT}\}_{n \geq 0}$  is a decreasing sequence of functions with the limit  $J_i^{(n)} e^{hT} \searrow v_i$  pointwise as  $n \nearrow \infty$ .

Let  $\mathcal{B}_b([0, T] \times (l, h); \mathbb{R})$  denote the set of bounded functions from  $[0, T] \times (l, h)$  to  $\mathbb{R}$

and define an operator  $\tilde{J} : \mathcal{B}_b([0, T] \times (l, h); \mathbb{R})^m \rightarrow \mathcal{B}_b([0, T] \times (l, h); \mathbb{R})^m$  by

$$\tilde{J} \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} := \begin{pmatrix} J_1(\sum_{j \neq 1} \frac{\lambda_{1j}}{\lambda_1} f_j) \\ \vdots \\ J_m(\sum_{j \neq m} \frac{\lambda_{mj}}{\lambda_m} f_j) \end{pmatrix}.$$

**Proposition 3.9.** (i) Let  $f \in \mathcal{B}_b([0, T] \times (l, h); \mathbb{R})^m$ . Then

$$\lim_{n \rightarrow \infty} \tilde{J}^n f = \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}.$$

(ii) The vector  $(v_1, \dots, v_m)^{tr}$  of value functions is a fixed point of the operator  $\tilde{J}$ , i.e.

$$\tilde{J} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}. \quad (3.25)$$

*Proof.* (i) Observe that the argument in the proof of part (iii) of Proposition 3.7 also gives that  $J_i^{(n)} g \rightarrow v_i$  as  $n \rightarrow \infty$  for any bounded  $g$ . Hence to finish the proof it is enough to recall the relation (3.13) in Proposition 3.5.

(ii) Let  $i \in \{1, \dots, m\}$ . By Proposition 3.5,

$$J_i^{(n+1)} 1 = J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j^{(n)} 1 \right). \quad (3.26)$$

By Proposition 3.7 (iii), for every  $j \in \{1, \dots, m\}$ , the sequence  $J_j^{(n)} 1 \nearrow v_j$  as  $n \nearrow \infty$ , so, letting  $n \nearrow \infty$  in (3.26), the monotone convergence theorem tells us that

$$v_i = J_i \left( \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} v_j \right). \quad (3.27)$$

□

## 4 The value function and the stopping strategy

In this section, we show that the value function  $v$  has attractive structural properties and identify an optimal strategy for the liquidation problem (2.7). The first passage time below a boundary, which is an increasing function of time and volatility, is proved to be optimal. Moreover, we provide a method to approximate the optimal stopping boundary by demonstrating that it is a limit of an increasing sequence of stopping boundaries

coming from easier auxiliary problems of Section 3.

**Theorem 4.1** (Properties of the value function).

- (i)  $v$  is decreasing in the first variable  $t$  as well as increasing and convex in the second variable  $x$ .
- (ii)  $v_i$  is continuous for every  $i \in \{1, \dots, m\}$ .
- (iii)

$$\check{v}_{\sigma_m} \leq v \leq \check{v}_{\sigma_1}, \quad (4.1)$$

where  $\check{v}_{\sigma_i} : [0, T] \times (l, h) \rightarrow \mathbb{R}$  denotes the Markovian value function as in (2.7), but for a price process (2.1) with constant volatility  $\sigma_i$ .

- Proof.* (i) Since, by Proposition 3.7 (ii), every  $J^{(n)}1$  is decreasing in the first variable  $t$ , increasing and convex in the second variable  $x$ , these properties are also preserved in the pointwise limit  $\lim_{n \rightarrow \infty} J^{(n)}1$ , which is  $v$  by Proposition 3.7 (iii).
- (ii) Using part (i) above, the claim follows from Proposition 3.9 (ii), i.e. from the fact that  $(v_1, \dots, v_m)^{tr}$  is a fixed point of a regularising operator  $\tilde{J}$  in the sense of Proposition 3.3.
- (iii) Letting  $n \rightarrow \infty$  in (3.21), Proposition 3.7 (iii) gives us (4.1). □

For the optimal liquidation problem (2.4) with constant volatility  $\sigma$ , i.e. in the case  $\sigma_1 = \dots = \sigma_m = \sigma$ , it has been shown in [13] that an optimal liquidation strategy is characterised by a increasing continuous stopping boundary  $\check{b}_\sigma : [0, T) \rightarrow [l, 0]$  with  $\check{b}_\sigma(T-) = 0$  such that the stopping time  $\check{\tau}_\sigma = \inf\{t \geq 0 : \hat{X}_t \leq \check{b}_\sigma(t)\} \wedge T$  is optimal. It turns out that the optimal liquidation strategy within our regime-switching volatility model shares some similarities with the constant volatility case as the next theorem shows.

**Theorem 4.2** (Optimal liquidation strategy).

- (i) For every  $i \in \{1, \dots, m\}$ , there exists  $b_{\sigma_i} : [0, T) \rightarrow [l, 0]$  that is increasing, right-continuous with left limits, satisfies the equality  $b_{\sigma_i}(T-) = 0$  and the identity

$$\mathcal{C}_i^{u_i} = \{(t, x) \in [0, T) \times (l, h) : x > b_{\sigma_i}(t)\}, \quad (4.2)$$

where  $u_i := \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} v_j$ . Moreover,

$$\check{b}_{\sigma_1} \leq b_{\sigma_i} \leq \check{b}_{\sigma_m}.$$

for any  $i \in \{1, \dots, m\}$ .

- (ii) The stopping strategy

$$\tau^* := \inf\{s \in [0, T-t) : \hat{X}_{t+s}^{t,x,\sigma} \leq b_{\sigma(t+s)}(t+s)\} \wedge (T-t).$$

is optimal for the optimal selling problem (2.7).

(iii) For  $i \in \{1, \dots, m\}$ , the boundaries

$$b_{\sigma_i}^{g_i^{(n)}} \searrow b_{\sigma_i} \quad \text{pointwise as } n \nearrow \infty,$$

$$\text{where } g_i^{(n)} := \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j^{(n)} 1.$$

(iv) The pairs  $(v_1, b_{\sigma_1}), (v_2, b_{\sigma_2}), \dots, (v_m, b_{\sigma_m})$  satisfy a coupled system of  $m$  free-boundary problems with each being

$$\begin{cases} \partial_t v_i(t, x) + \sigma_i \phi(x, \sigma_i) \partial_x v_i(t, x) + \frac{1}{2} \phi(x, \sigma_i)^2 \partial_{xx} v_i(t, x) \\ \quad + (x - \lambda_i) v_i(t, x) + \sum_{j \neq i} \lambda_{ij} v_j(t, x) = 0, & \text{if } x > b_i(t), \\ v_i(t, x) = 1, & \text{if } x \leq b_i(t) \text{ or } t = T, \end{cases} \quad (4.3)$$

where  $i \in \{1, \dots, m\}$ .

*Proof.* (i) The existence of  $b_{\sigma_i} : [0, T) \rightarrow [l, h]$  that is increasing, right-continuous with left limits, and satisfies (4.2) follows from the fixed-point property (3.25), and Theorem 4.1 (i),(ii). Since the range of  $\check{b}_{\sigma_1}, \check{b}_{\sigma_m}$  is  $[l, 0]$  and  $\check{b}_{\sigma_1}(T-) = \check{b}_{\sigma_m}(T-) = 0$ , using Theorem 4.1 (iii), we also conclude that  $\check{b}_{\sigma_1} \leq b_{\sigma_i} \leq \check{b}_{\sigma_m}$  and that  $b_{\sigma_i}(T-) = 0$  for every  $i$ .

(ii) Let us define  $\mathcal{D} := \{(t, x, \sigma) \in [0, T] \times (l, h) \times \{\sigma_1, \dots, \sigma_m\} : v(t, x, \sigma) = 1\}$ . Then  $\tau_{\mathcal{D}} := \inf\{s \geq 0 : (t + s, \hat{X}_{t+s}^{t, x, \sigma(t)}, \sigma(t + s)) \in \mathcal{D}\}$  is optimal for the problem (2.7) by [27, Corollary 2.9]. Lastly, from the fixed-point property (3.25) and Proposition 3.2, we conclude that  $\tau^* = \tau_{\mathcal{D}}$ , which finishes the proof.

(iii) Since  $J_i^{(n)} 1 \nearrow v_i$  as  $n \nearrow \infty$  and  $J_i^{(n)} 1 \geq 1$  for all  $n$ , we have that  $\lim_{n \nearrow \infty} b_{\sigma_i}^{g_i^{(n)}} \geq b_{\sigma_i}$ . Also, if  $x < \lim_{n \nearrow \infty} b_{\sigma_i}^{g_i^{(n)}}(t)$ , then  $J_i^{(n)} 1(t, x) = 1$  for all  $n \in \mathbb{N}$  and so  $v_i(t, x) = \lim_{n \nearrow \infty} J_i^{(n)} 1(t, x) = 1$ . Hence,  $\lim_{n \nearrow \infty} b_{\sigma_i}^{g_i^{(n)}} \leq b_{\sigma_i}$ . As a result,  $\lim_{n \nearrow \infty} b_{\sigma_i}^{g_i^{(n)}} = b_{\sigma_i}$ .

(iv) The free-boundary problem is a consequence of Proposition 3.4 (ii) and the fixed-point property (3.25). □

**Remark 4.3.** Establishing uniqueness of a classical solution to a time non-homogeneous free-boundary problem is typically a technical task (see [24] for an example). Not being central to the mission of the paper, the uniqueness of solution to the free-boundary problems (4.3) and (3.11) has not been pursued.

**Remark 4.4** (A possible alternative approach). It is worth pointing out that a potential alternative approach for the study of the value function and the optimal strategy is to study directly the variational inequality formulation (e.g., see [26, Section 5.2]) arising from the optimal stopping problem (2.7). The variational inequality would need to be studied using PDE techniques (likely, similar to the techniques of [6]) to obtain desired

regularity and structural properties of the value function and the stopping region. Though the author is unaware of any work studying exactly this type of free-boundary problem directly in detail, there are related results on existence and uniqueness of solutions to free-boundary problems of American option pricing in stochastic volatility models; see [24]. Still, with this alternative in mind (if possible at all), our approach is likely to be simpler since the analysis of the full variational inequality in three dimensions is instead replaced by an investigation of a sequence of standard tractable two-dimensional optimal stopping problems. Not least, additional analytical technicalities are avoided.

For further study of the problem in this section, we will make a structural assumption about the Markov chain modelling the volatility.

**Assumption 4.5.** The Markov chain  $\sigma$  is *skip-free*, i.e. for all  $i \in \{1, \dots, m\}$ ,

$$\lambda_{ij} = 0 \text{ if } j \notin \{i-1, i, i+1\}.$$

As many popular financial stochastic volatility models have continuous trajectories, and a skip-free Markov chain is a natural discrete state-space approximation of a continuous process, Assumption 4.5 does not appear to be a severe restriction.

**Lemma 4.6.** *Let  $\delta > 0$ ,  $g : (l, h) \times [0, \infty) \rightarrow [0, \infty)$  be increasing and convex in the first variable as well as decreasing in the second. Then  $u : (l, h) \times \{\sigma_1, \dots, \sigma_m\} \rightarrow \mathbb{R}$  defined by*

$$u(x, \sigma_i) := \mathbb{E} \left[ e^{\int_0^\delta \hat{X}_u^{x, \sigma_i} du} g(\hat{X}_\delta^{x, \sigma}, \sigma(\delta)) \right] \quad (4.4)$$

*is increasing and convex in the first variable as well as decreasing in the second.*

*Proof.* We will prove the claim using a coupling argument. Let  $(\Omega', \mathcal{F}', \tilde{\mathbb{P}}')$  be a probability triplet supporting a Brownian motion  $B$ , and two volatility processes  $\sigma^1, \sigma^2$  with the state space and transition densities as in (2.1). In addition, we assume that  $B$  is independent of  $(\sigma^1, \sigma^2)$ , that the starting values satisfy  $\sigma^1(0) = \sigma_i \leq \sigma_j = \sigma^2(0)$ , and that  $\sigma^1(t) \leq \sigma^2(t)$  for all  $t \geq 0$ . Also, let  $\hat{X}^1$  and  $\hat{X}^2$  denote the solutions to (2.6) when  $\sigma$  is replaced by  $\sigma^1$  and  $\sigma^2$ , respectively.

Let us fix an arbitrary  $\omega_0 \in \Omega'$ . Since  $\hat{W}$  is independent of  $\sigma^1$ ,

$$\tilde{E}' \left[ e^{\int_0^\delta (\hat{X}^1)_u^x du} g((\hat{X}^1)_\delta^x, \sigma^1(\delta)) \mid \mathcal{F}_\delta^{\sigma^1} \right] (\omega_0) = \tilde{\mathbb{E}}' \left[ e^{\int_0^\delta (\tilde{X}^1)_u^x du} g((\tilde{X}^1)_\delta^x, \sigma^1(\delta, \omega_0)) \right], \quad (4.5)$$

where  $\tilde{X}^1$  denotes the process  $\hat{X}^1$  with the volatility process  $\sigma^1$  replaced by a deterministic function  $\sigma^1(\cdot, \omega_0)$ . Furthermore, the right-hand (and so the left-hand side) in (4.5) as a function of  $x$  is increasing by [28, Theorem IX.3.7] as well as convex by [12, Theorem 5.1]. Hence

$$u(\cdot, \sigma_i) : x \mapsto \tilde{\mathbb{E}}' \left[ \tilde{\mathbb{E}}' \left[ e^{\int_0^\delta (\hat{X}^1)_u^x du} g((\hat{X}^1)_\delta^x, \sigma^1(\delta)) \mid \mathcal{F}_\delta^{\sigma^1} \right] \right]$$

is increasing and convex. Next, we observe that

$$\begin{aligned}
& \tilde{\mathbb{E}}' \left[ e^{\int_0^\delta (\hat{X}^1)_u^x du} g((\hat{X}^1)_\delta^x, \sigma^1(\delta)) \mid \mathcal{F}_\delta^{\sigma^1, \sigma^2} \right] (\omega_0) \\
& \geq \tilde{\mathbb{E}}' \left[ e^{\int_0^\delta (\hat{X}^2)_u^\delta du} g((\hat{X}^2)_\delta^x, \sigma^1(\delta)) \mid \mathcal{F}_\delta^{\sigma^1, \sigma^2} \right] (\omega_0) \\
& \geq \tilde{\mathbb{E}}' \left[ e^{\int_0^\delta (\hat{X}^2)_u^x du} g((\hat{X}^2)_\delta^x, \sigma^2(\delta)) \mid \mathcal{F}_\delta^{\sigma^1, \sigma^2} \right] (\omega_0).
\end{aligned} \tag{4.6}$$

In the above, having in mind that the conditional expectations can be rewritten as ordinary expectations similarly as in (4.5), the first inequality followed by [12, Theorem 6.1], the second by the decay of  $g$  in the second variable. Integrating both sides of (4.6) over all possible  $\omega_0 \in \Omega'$  with respect to  $d\mathbb{P}'$ , we get that

$$u(x, \sigma_1) \geq u(x, \sigma_2).$$

Thus we can conclude that  $u$  is increasing and convex in the first variable as well as decreasing in the second.  $\square$

**Theorem 4.7** (Ordering in volatility).

(i)  $v$  is decreasing in the volatility variable, i.e.

$$v_{\sigma_1} \geq v_{\sigma_2} \geq \dots \geq v_{\sigma_m}.$$

(ii) The boundaries are ordered in volatility as

$$b_{\sigma_1} \leq b_{\sigma_2} \leq \dots \leq b_{\sigma_m}.$$

*Proof.* (i) We will prove the claim by approximating the value function  $v$  by a sequence of value functions  $\{v_n\}_{n \geq 0}$  of corresponding Bermudan optimal stopping problems. Let  $v_n$  denote the value function as in (2.7), but when stopping is allowed only at times  $\{\frac{kT}{2^n} : k \in \{0, 1, \dots, 2^n\}\}$ .

Let us fix  $n \in \mathbb{N}$ . We will show that, for any given  $k \in \{0, \dots, 2^n\}$  and any  $t \in [\frac{k}{2^n}T, T]$ , the value function  $v_n(t, x, \sigma)$  is increasing and convex in  $x$  as well as decreasing in  $\sigma$  (note that here  $\sigma$  denotes the initial value of the process  $t \mapsto \sigma(t)$ ). The proof is by backwards induction from  $k = 2^n$  down to  $k = 0$ . Since  $v_n(T, \cdot, \cdot) = 1$ , the base step  $k = 2^n$  holds trivially. Now, suppose that, for some given  $k \in \{0, \dots, 2^n\}$ , the value  $v_n(t, x, \sigma)$  is increasing and convex in  $x$  as well as decreasing in  $\sigma$  for any  $t \in [\frac{k}{2^n}T, T]$ . Then, Lemma 4.6 tells us that for any fixed  $t \in [\frac{(k-1)T}{2^n}, \frac{kT}{2^n})$ ,

$$f(t, x, \sigma) := \tilde{\mathbb{E}} \left[ e^{\int_t^{\frac{kT}{2^n}} \hat{X}_u^{t,x,\sigma} du} v_n \left( \frac{kT}{2^n}, \hat{X}_{\frac{kT}{2^n}}^{t,x,\sigma}, \sigma \left( \frac{kT}{2^n} \right) \right) \right],$$

is increasing and convex in  $x$  as well as decreasing in  $\sigma$ . Consequently, since

$$v_n(t, x, \sigma) = \begin{cases} f(t, x, \sigma), & t \in (\frac{(k-1)T}{2^n}, \frac{kT}{2^n}), \\ f(t, x, \sigma) \vee 1, & t = \frac{(k-1)T}{2^n}, \end{cases} \quad (4.7)$$

the value  $v_n(t, x, \sigma)$  is increasing and convex in  $x$  as well as decreasing in  $\sigma$  for any fixed  $t \in [\frac{k-1}{2^n}T, T]$ . Hence, by backwards induction,  $v_n$  is increasing and convex in the second argument  $x$  as well as decreasing in the third argument  $\sigma$ .

Finally, since  $v_n \rightarrow v$  pointwise as  $n \rightarrow \infty$ , we can conclude that the value function  $v$  is decreasing in  $\sigma$ .

- (ii) From the proof of Theorem 4.2 (ii), the claim is a direct consequence of part (i) above. □

**Remark 4.8.**

1. The value function is decreasing in the initial volatility (Theorem 4.7 (i)) also when the volatility is any continuous time-homogeneous positive Markov process independent of the driving Brownian motion  $W$ . The assertion is justified by inspection of the proof of Lemma 4.6 in which no crossing of the volatility trajectories was important, not the Markov chain structure.
2. Though there are no grounds to believe that any of the boundaries  $b_{\sigma_1}, \dots, b_{\sigma_m}$  is discontinuous, proving their continuity, except for the lowest one, is beyond the power of customary techniques. Continuity of the lowest boundary can be proved similarly as in the proof of part 4 of [13, Theorem 3.10], exploiting the ordering of the boundaries. The stumbling block for proving continuity of the upper boundaries is that, at a downward volatility jump time, the value function has a positive jump whose magnitude is difficult to quantify.

## 5 Generalisation to an arbitrary prior

In this section, we generalise most results of the earlier parts to the general prior case. In what follows, the prior  $\mu$  of the drift is no longer a two-point but an arbitrary probability distribution.

### 5.1 Two-dimensional characterisation of the posterior distribution

Let us first think a bit more abstractly to develop intuition for the arbitrary prior case. According to the Kushner-Stratonovich stochastic partial differential equation (SPDE) for the posterior distribution (see Section 3.2 of [7]), if we take the innovation process driving the SPDE and the volatility as the available information sources, then the posterior distribution is a measure-valued Markov process. Unfortunately, there does not exist any applicable general methods to solve optimal stopping problems for measure-valued

stochastic processes. If only we were able to characterise the posterior distribution process by an  $\mathbb{R}^n$ -valued Markovian process (with respect to the filtration generated by the innovation and the volatility processes), then we should manage to reduce our optimal stopping problem with a stochastic measure-valued underlying to an optimal stopping problem with a  $\mathbb{R}^n$ -valued Markovian underlying. Mercifully, this wishful thinking turns out to be possible in reality as we shall soon see.

Unlike in the problem with constant volatility studied in [13], when the volatility is varying, the pair consisting of the elapsed time  $t$  and the posterior mean  $\hat{X}_t$  is not sufficient (with an exception of the two-point prior case studied before) to characterise the posterior distribution  $\mu_t$  of  $X$  given  $\mathcal{F}_t^{S,\sigma}$ . Hence we need some additional information to describe the posterior distribution. Quite surprisingly, all this needed additional information can be captured in a single additional observable statistic which we will name the ‘effective learning time’. We start the development by first introducing some useful notation.

Define  $Y_t^{(i)} := Xt + \sigma_i W_t$  and let  $\mu_{t,y}^{(i)}$  denote the posterior distribution of  $X$  at time  $t$  given  $Y_t^{(i)} = y$ . It needs to be mentioned that, for any given prior  $\mu$ , the distributions of  $X$  given  $\mathcal{F}_t^{Y^{(i)}}$  and  $X$  given  $Y_t^{(i)}$  are equal (see Proposition 3.1 in [13]), which justifies our conditioning only on the last value  $Y_t^{(i)}$ . Also, recall that  $l = \inf \text{supp}(\mu)$ ,  $h = \sup \text{supp}(\mu)$ .

The next lemma provides the key insight allowing to characterise the posterior distribution by only two parameters.

**Lemma 5.1.** *Let  $\sigma_2 \geq \sigma_1 > 0$ . Then*

$$\{\mu_{t,y}^{(1)} : t > 0, y \in \mathbb{R}\} = \{\mu_{t,y}^{(2)} : t > 0, y \in \mathbb{R}\},$$

*i.e. the sets of possible conditional distributions of  $X$  in both cases are the same.*

*Proof.* Let  $t > 0, y \in \mathbb{R}$ . By the standard filtering theory (a generalised Bayes’ rule),

$$\mu_{t,y}^{(i)}(du) := \frac{e^{\frac{2uy - u^2 t}{2\sigma_i^2}} \mu(du)}{\int_{\mathbb{R}} e^{\frac{2uy - u^2 t}{2\sigma_i^2}} \mu(du)}. \quad (5.1)$$

Then taking  $r = \left(\frac{\sigma_1}{\sigma_2}\right)^2 t$  and  $y_1 = \left(\frac{\sigma_1}{\sigma_2}\right)^2 y$ , we have that

$$\mu_{t,y}^{(2)}(du) = \mu_{r,y_1}^{(1)}(du).$$

□

From Lemma 5.1 and [13, Lemma 3.3] we obtain the following important corollary, telling us that, having fixed a prior, any possible posterior distribution can be fully characterised by only two parameters.

**Corollary 5.2.** *Let  $t > 0$ . Then, for any posterior distribution  $\mu_t(\cdot) = \mathbb{P}(X \in \cdot | \mathcal{F}_t^{S,\sigma})(\omega)$ , there exists  $(r, x) \in (0, T] \times (l, h)$  such that  $\mu_t = \mu_{r,y_1}^{(1)}$ , where  $y_1(r, x)$  is defined as*

the unique value satisfying  $\mathbb{E}[X | Y_r^{(1)} = y_1(r, x)] = x$ . In particular, we can take  $r = \int_0^t \left( \frac{\sigma_1}{\sigma(u)(\omega)} \right)^2 du$  and  $y_1(r, x) = \int_0^t \left( \frac{\sigma_1}{\sigma(u)(\omega)} \right)^2 dY_u(\omega)$ , where  $Y_u = \log(S_u) + \frac{1}{2} \int_0^u \sigma(b)^2 db$ .

When the volatility varies, so does the speed of learning about the drift. The corollary tells us that we can interpret  $r$  as the *effective learning time* measured under the constant volatility  $\sigma_1$ . The intuition for the name is that even though the volatility is varying over time, the same posterior distribution  $\mu_t$  can be also be obtained in a constant volatility model with the constant volatility  $\sigma_1$ , just at a different time  $r$  and at a different value of the price  $S$ .

**Remark 5.3.** It is worth remarking that Corollary 5.2 also holds for any reasonable positive volatility process. Indeed, using the Kallianpur-Striebel formula with time-dependent volatility (see Theorem 2.9 on page 39 of [7]), the proof of Lemma 5.1 equally applies for an arbitrary positive time-dependent volatility and immediately yields the result of the corollary.

Next, we make a convenient technical assumption about the prior distribution  $\mu$ .

**Assumption 5.4.** The prior distribution  $\mu$  is such that

1.  $\int_{\mathbb{R}} e^{au^2} \mu(du) < \infty$  for some  $a > 0$ ,
2.  $\psi(\cdot, \cdot) : [0, T] \times (l, h) \rightarrow \mathbb{R}$  defined by

$$\psi(t, x) := \frac{1}{\sigma_1} (\mathbb{E}[X^2 | Y_t^1 = y_1(t, x)] - x^2) = \frac{1}{\sigma_1} \text{Var}(X | Y_t^1 = y_1(t, x))$$

is a bounded function that is Lipschitz continuous in the second variable.

In particular, all compactly supported distributions as well as the normal distribution are known to satisfy Assumption 5.4 (see [13]), so it is an inconsequential restriction for practical applications.

## 5.2 Markovian embedding

Similarly as in the two-point prior case, we will study the optimal stopping problem (2.5) by embedding it into a Markovian framework. With Corollary 5.2 telling us that the effective learning time  $r$  and the posterior mean  $x$  fully characterise the posterior distribution, now, we can embed the optimal stopping problem (2.5) into the standard Markovian framework by defining the Markovian value function

$$v(t, x, r, \sigma) := \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t, x, r, \sigma} ds} \right], (t, x, r, \sigma) \in [0, T] \times (l, h) \times [0, T] \times \{\sigma_1, \dots, \sigma_m\}. \quad (5.2)$$

Here the process  $\hat{X} = \hat{X}^{t,x,r,\sigma_i}$  evolves according to

$$\begin{cases} d\hat{X}_{t+s} = \sigma_1 \psi(r_{t+s}, \hat{X}_{t+s}) ds + \frac{\sigma_1}{\sigma(t+s)} \psi(r_{t+s}, \hat{X}_{t+s}) dB_{t+s}, & s \geq 0, \\ dr_{t+s} = \left( \frac{\sigma_1}{\sigma(t+s)} \right)^2 ds, & s \geq 0, \\ \hat{X}_t = x, \\ r_t = r, \\ \sigma(t) = \sigma_i; \end{cases} \quad (5.3)$$

the given dynamics of  $\hat{X}$  is a consequence of Corollary 5.2 and the evolution equation of  $\hat{X}$  in the constant volatility case (see the equation (3.9) in [13]). Also, in (5.3), the process  $B_t = \int_0^t \sigma(u) du + \hat{W}_t$  is a  $\tilde{\mathbb{P}}$ -Brownian motion. Lastly, in (5.2),  $\mathcal{T}_{T-t}$  denotes the set of stopping times less or equal to  $T-t$  with respect to the usual augmentation of the filtration generated by  $\{\hat{X}_{t+s}^{t,x,r,\sigma_i}\}_{s \geq 0}$  and  $\{\sigma(t+s)\}_{s \geq 0}$ .

**Remark 5.5.** Let us note that in light of the observations of Section 5.1, if the regime-switching volatility were replaced by a different stochastic volatility process, the same Markovian embedding (5.2) could still be useful for the to study an altered problem.

### 5.3 Outline of the approximation procedure and main results

Under an arbitrary prior, the approximation procedure of Section 3 can also be applied, however, the operators  $J$  and  $J^{(n)}$  need to be redefined in a suitable way. We redefine the operator  $J$  to act on a function  $f : [0, T] \times (l, h) \times [0, T] \rightarrow \mathbb{R}$  as

$$\begin{aligned} (Jf)(t, x, r, \sigma_i) &:= \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t,x,r,\sigma_i} ds} \mathbb{1}_{\{\tau < \eta_i^t\}} + e^{\int_0^{\eta_i^t} \hat{X}_{t+s}^{t,x,r,\sigma_i} ds} f \left( t + \eta_i^t, \hat{X}_{t+\eta_i^t}^{t,x,r,\sigma_i}, r_{t+\eta_i^t}^{t,r} \right) \mathbb{1}_{\{\tau \geq \eta_i^t\}} \right] \\ &= \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t,x,r,\sigma_i} - \lambda_i ds} + \int_0^\tau e^{\int_0^u \hat{X}_{t+s}^{t,x,r,\sigma_i} - \lambda_i ds} f \left( t + u, \hat{X}_{t+u}^{t,x,r,\sigma_i}, r_{t+\eta_i^t}^{t,r} \right) du \right] \end{aligned} \quad (5.4)$$

and then the operator  $J_i$  as  $J_i f := (Jf)(\cdot, \cdot, \sigma_i)$ . Intuitively,  $(J_i f)$  represents a Markovian value function corresponding to optimal stopping before  $t + \eta_i^t$ , i.e. before the first volatility change after  $t$ , when, at time  $t + \eta_i^t < T$ , the payoff  $f \left( t + \eta_i^t, \hat{X}_{t+\eta_i^t}^{t,x,r,\sigma_i}, r_{t+\eta_i^t}^{t,r} \right)$  is received, provided stopping has not occurred yet. The underlying process in the optimal stopping problem  $J_i f$  is the diffusion  $(t, \hat{X}_t, r_t)$ .

The majority of the results in Sections 3 and 4 generalise nicely to an arbitrary prior case. Proposition 3.1 extends word by word; the proofs are analogous, just the second property of  $\psi$  from [13, Proposition 3.6] needs to be used for Proposition 3.1 (iv). In addition, we have that  $f$  decreasing in  $r$  implies that  $J_i f$  is decreasing in  $r$ , which is proved by a Bermudan approximation argument as in Proposition 3.1 (iv) using the time decay of  $\psi$  from [13, Proposition 3.6]. As a result, for  $f : [0, T] \times (l, h) \times [0, T] \rightarrow \mathbb{R}$  that is decreasing in the first and third variables as well as increasing (though not too fast as  $x \nearrow \infty$ ) and convex in the second, there exists a function (a stopping boundary)  $b_{\sigma_i}^f : [0, T] \times [0, T] \rightarrow [l, 0]$  that is increasing in both variables and such that the continuation

region  $\mathcal{C}_i^f := \{(t, x, r) \in [0, T] \times (l, h) \times [0, T] : (J_i f)(t, x, r) > 1\}$  (optimality shown as in Proposition 3.2) satisfies

$$\mathcal{C}_i^f = \{(t, x, r) \in [0, T] \times (l, h) \times [0, T] : x > b_{\sigma_i}^f(t, r)\}.$$

In addition, each pair  $(J_i f, b_{\sigma_i}^f)$  solves the free-boundary problem

$$\begin{cases} \partial_t u(t, x, r) + \left(\frac{\sigma_1}{\sigma_i}\right)^2 \partial_r u(t, x, r) + \sigma_1 \psi(r, x) \partial_x u(t, x, r) \\ + \frac{1}{2} \left(\frac{\sigma_1}{\sigma_i}\right)^2 \psi(r, x)^2 \partial_{xx} u(t, x, r) + (x - \lambda_i) u(t, x, r) + \lambda_i f(t, x, r) = 0, \text{ if } x > b_{\sigma_i}^f(t, r), \\ u(t, x, r) = 1, \text{ if } x \leq b_{\sigma_i}^f(t, r) \text{ or } t = T. \end{cases}$$

With the operator  $J^{(n)}$  redefined as

$$(J^{(n)} f)(t, x, r, \sigma_i) := \sup_{\tau \in \mathcal{T}_{T-t}} \tilde{\mathbb{E}} \left[ e^{\int_0^\tau \hat{X}_{t+s}^{t,x,r,\sigma_i} ds} \mathbb{1}_{\{\tau < \xi_n^t\}} + e^{\int_0^{\xi_n^t} \hat{X}_{t+s}^{t,x,r,\sigma_i} ds} f(t + \xi_n^t, \hat{X}_{t+\xi_n^t}^{t,x,r,\sigma_i}, r_{t+\xi_n^t}^{t,r}) \mathbb{1}_{\{\tau \geq \xi_n^t\}} \right],$$

the crucial Proposition 3.5 holds word by word. Furthermore, the sequence of functions  $\{J^{(n)} 1\}_{n \geq 0}$  is increasing, bounded from below by 1 with each  $J^{(n)} 1$  being decreasing in the first and third variables as well as increasing and convex in the second variable  $x$ . As desired,

$$J^{(n)} 1 \nearrow v \quad \text{pointwise as } n \nearrow \infty,$$

so the value function  $v$  is decreasing in the first and third variables as well as increasing and convex in the second variable; again,  $v$  is a fixed point of  $\tilde{J}$ . Moreover, the uniform approximation error result (3.20) also holds for compactly supported priors (with an obvious reinterpretation  $h = \sup(\text{supp } \mu)$ ). We can also show (by a similar argument as in Theorem 4.2 (iii)) that

$$b_{\sigma_i}^{g_i^{(n)}} \searrow b_{\sigma_i} \quad \text{pointwise as } n \nearrow \infty,$$

where  $g_i^{(n)} := \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} J_j^{(n)} 1$  and the limit  $b_{\sigma_i}$  is a function increasing in both variables. Lastly, by similar arguments as before, the stopping time

$$\tau^* = \inf\{s \in [0, T-t] : \hat{X}_{t+s}^{t,x,r,\sigma} \leq b_{\sigma(t+s)}(t+s, r_{t+s})\} \wedge (T-t)$$

is optimal for the liquidation problem (2.5).

**Remark 5.6.** The higher volatility, the slower learning about the drift, so under Assumption 4.5 it is tempting to expect that the value function  $v$  is decreasing in the volatility variable and so the stopping boundaries  $b_{\sigma_1} \leq b_{\sigma_2} \leq \dots \leq b_{\sigma_m}$  also in the case of an arbitrary prior distribution  $\mu$ . Regrettably, proving (or disproving) such monotonicity in volatility has not been achieved by the author.

## References

- [1] Bain, A., Crisan, D. *Fundamentals of stochastic filtering*. Stochastic Modelling and Applied Probability, 60. Springer, New York, 2009.
- [2] Bayraktar, E. A Proof of the Smoothness of the Finite Time Horizon American Put Option for Jump Diffusions. *SIAM Journal on Control and Optimization*, vol. 48, no. 2, 2009, 551-572.
- [3] Bayraktar, E., Dayanik, S., Karatzas, I. Adaptive Poisson Disorder Problem. *Annals of Applied Probability*, vol. 16, no. 3, 2006, 1190-1261.
- [4] Bayraktar, E., Kravitz, R. Quickest Detection with Discretely Controlled Observations. *Sequential Analysis*, vol. 34, no. 1, 2015, 77-133.
- [5] Bayraktar, E. On the perpetual American put options for level dependent volatility models with jumps. *Quantitative Finance*, vol. 11, no. 3, 2011, 335-341.
- [6] Bensoussan, A., Lions, J.-L. *Applications of Variational Inequalities in Stochastic Control*. Studies in Mathematics and its Applications, vol. 12. North-Holland, Amsterdam, 1982.
- [7] Crisan, D., Rozovskii, B. *The Oxford handbook of nonlinear filtering*. Oxford University Press, 2011.
- [8] Dayanik, S., Poor, H. V., Sezer, S. O. Multisource Bayesian sequential change detection. *Annals of Applied Probability*, vol. 18, no. 2, 2008, 552-590.
- [9] Décamps, J.-P., Mariotti, T., Villeneuve, S. Investment timing under incomplete information. *Math. Oper. Res.*, vol. 30, no. 2, 2005, 472-500.
- [10] Di Masi, G. B., Kabanov, Y. M., Runggaldier, W. J. Mean-variance hedging of options on stocks with Markov volatilities. *Theory of Probability & Its Applications*, vol. 39, no. 1, 172-182.
- [11] Ekström, E., Lu, B. Optimal selling of an asset under incomplete information. *International Journal of Stochastic Analysis*, vol. 2011, 2011, ID 543590.
- [12] Ekström, E., Tysk, J. Convexity theory for the term structure equation. *Finance and Stochastics*, vol. 12, no. 1, 2008, 117-147.
- [13] Ekström, E., Vaicenavicius, J. Optimal liquidation of an asset under drift uncertainty. *SIAM Journal on Financial Mathematics*, vol. 7, no. 1, 2016, 357-381
- [14] Ekström, E., Lu, B. The Optimal Dividend Problem in the Dual Model. *Advances in Applied Probability*, vol. 46, no. 3, 2014, 746-765.

- [15] Eloe, P., Liu R.H., Yatsuki M., Yin G., Zhang, Q. Optimal selling rules in a regime-switching exponential Gaussian diffusion model. *SIAM Journal on Applied Mathematics*. vol. 69, no. 3, 2008, 810-829.
- [16] Gapeev, P. Pricing of perpetual American options in a model with partial information. *International Journal of Theoretical and Applied Finance*, vol. 15, no. 1, 2012, ID 1250010.
- [17] Gugerui, U. S. Optimal stopping of a piecewise-deterministic Markov process. *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 19, no. 4, 1986, 221-236.
- [18] Guo, X. Zhang, Q. Closed-form solutions for perpetual American put options with regime switching. *SIAM Journal on Applied Mathematics*, vol. 64, no. 6, 2004, 2034-2049.
- [19] Guo, X. Zhang, Q. Optimal selling rules in a regime switching model. *IEEE Transactions on Automatic Control*, vol. 50, 2005, 1450-1455.
- [20] Karatzas, I., Shreve, S. *Methods of Mathematical Finance*. Applications of Mathematics, 39. Springer-Verlag, New York, 1998.
- [21] Le, H., Wang, C. A Finite Time Horizon Optimal Stopping Problem with Regime Switching. *SIAM Journal on Control and Optimization*, vol. 48, no. 8, 5193-5213.
- [22] Lu, B. Optimal selling of an asset with jumps under incomplete information. *Applied Mathematical Finance*, vol. 20, no. 6, 2013, 599610.
- [23] Øksendal, B. *Stochastic Differential Equations: an introduction with applications*. Sixth edition. Springer, 2007.
- [24] Pascucci, A. Free boundary and optimal stopping problems for American Asian options. *Finance and Stochastics*, vol. 12, no. 1, 2008, 21-41.
- [25] Pemy, M., Zhang, Q. Optimal stock liquidation in a regime switching model with finite time horizon. *Journal of Mathematical Analysis and Applications*. vol. 321, no. 2, 2006, 537-552.
- [26] Pham, H. Continuous-time stochastic control and optimization with financial applications, 61. Springer, 2009.
- [27] Peskir, G., Shiryaev, A. *Optimal stopping and free-boundary problems*. Lectures in Mathematics, ETH Zürich. Birkhäuser Verlag, Basel, 2006.
- [28] Revuz, D., Yor, M. *Continuous martingales and Brownian motion*. Third edition. Grundlehren der Mathematischen Wissenschaften, 293. Springer-Verlag, Berlin, 1999.

- [29] Rogers, L.C.G. *Optimal Investment*. Springer Briefs in Quantitative Finance. Springer, 2013.
- [30] Vannestål, M. Exercising American options under uncertainty. Working paper, 2017.
- [31] Yin, G., Liu, R.H., Zhang, Q. Recursive algorithms for stock liquidation: A stochastic optimization approach, *SIAM Journal on Optimization*, vol. 13, no. 1, 2002, 240-263.
- [32] Yin, G., Zhang, G., Liu, F., Liu, R. H., Cheng, Y. Stock liquidation via stochastic approximation using Nasdaq daily and intra-day data. *Mathematical Finance*, vol. 16, no. 1, 2006, 217-236.
- [33] Zhang, Q. Stock trading: An optimal selling rule. *SIAM Journal on Control and Optimization*, vol. 40, no. 1, 2001, 64-87.
- [34] Zhang, Q., Yin, G., Liu, R. H. A near-optimal selling rule for a two-time-scale market model. *Multiscale Modeling & Simulation*, vol. 4, no. 1, 2005, 172-193.