On American VIX options

Jérôme Detemple* & Yerkin Kitapbayev†

4 August 2016

Abstract

In this paper, we extend the 3/2-model for VIX studied by Goard and Mazur (2013) and introduce the generalized 3/2 and 1/2 classes of volatility processes. Under these models, we study the pricing of European and American VIX options and, for the latter, we obtain an early exercise premium representation using a free-boundary approach and local time-space calculus. The optimal exercise boundary for the volatility is obtained as the unique solution to an integral equation of Volterra type.

We also consider a model mixing these two classes and formulate the corresponding optimal stopping problem in terms of the observed factor process. The price of an American VIX call is then represented by an early exercise premium formula. We show the existence of a pair of optimal exercise boundaries for the factor process and characterize them as the unique solution to a system of integral equations.

JEL Classification: C61, G13, G17.

Key Words: Stochastic volatility; generalized 3/2 and 1/2 models; generalized mixture models; American options; exercise premium; exercise boundaries; integral equations; local time.

1. Introduction

During recent decades, financial markets have experienced significant fluctuations in volatility. These events have spurred demands for volatility indicators and for derivative instruments to manage volatility risk. Nowadays, the most popular volatility measurement is the VIX, which is the implied volatility of 30-day S&P500 options. VIX futures contracts started to trade on March 26, 2004 on the CBOE. Options on the VIX, introduced on February 24, 2006, also by the CBOE, have proven increasingly popular with investors. Since their introduction, volume has grown from a daily average of 23,491 contracts in 2006 to 632,419 in 2014. This popularity stems in part from the recurrence of rapidly changing volatility episodes, especially during the recent crisis. VIX options provide an effective way to manage risks tied to volatility fluctuations.

The valuation of VIX options has been considered well before their actual introduction on the CBOE. The issue became of interest in the early 90s, around the time when the VIX index was introduced to measure volatility (see Whaley (1993)). Valuation formulas have developed

*Questrom School of Business, Boston University, Boston, USA. E-mail: detemple@bu.edu

†Questrom School of Business, Boston University, Boston, USA. E-mail: yerkin@bu.edu

around a set of well known models for the evolution of the underlying volatility. Formulas for European volatility options can be found in Whaley (1993) under the assumption of a geometric Brownian motion process (GBMP) and in Grünbichler and Longstaff (1996) for a mean-reverting square-root volatility process (MRSRP), also known as CIR process (see Cox, Ingersoll and Ross (1985)). For American-style volatility contracts, Detemple and Osakwe (2000) provide formulas for Geometric Brownian motion process (GBMP), mean-reverting Gaussian process (MRGP), mean-reverting square root process (MRSRP) and mean-reverting log process (MRLP). All these cases can be embedded in the volatility models,

$$dX_t = (\beta - \alpha X_t)\, dt - \kappa X_t^{\gamma} dB_t, \qquad \text{with } \gamma = 0, 1/2 \text{ or } 1 \tag{1.1}$$

$$d\ln X_t = (\beta - \alpha \ln X_t)\, dt - \kappa dB_t.$$

This last one is an exponential transform of a Gaussian process. Another transform of volatility that has been used to price variance contracts is the Heston (1993) model, where the local variance $v = X^2$ follows,

$$dv_t = (\beta - \alpha v_t)\, dt - \kappa v_t^{1/2} dB_t$$

which is a MRSRP for v. Contracts on realized variance or realized volatility, such as swaps and options, have been examined under this specification by several authors, including Broadie and Jain (2008). Realized variance/volatility contracts have also been priced under various generalizations of the Heston model, e.g., Elliott, Siu and Chan (2007) and Sepp (2008).

More recently, Goard and Mazur (2013) examine the valuation of VIX options under the $3/2$ specification,

$$dX_t = \left(\alpha X_t - \left(\beta - \kappa^2\right) X_t^2\right) dt + \kappa X_t^{3/2} dB_t. \tag{1.2}$$

This process was originally introduced by Heston (1997) and Platen (1997) to model the evolution of the local variance v of an asset return. An interesting feature of the process is that it allows for spikes, a property found in volatility data. Drimus (2001) provides empirical evidence in favor of the model for FX markets. Goard and Mazur, show that the $3/2$ specification, as a model for volatility, provides a better fit to the VIX data than various alternatives including GBMP, MRGP, MRSRP, MRLP and Heston's MRSRP for v. They also compute European VIX option prices under this model. Relying on this evidence, Liu (2015) formulates a free-boundary problem for the valuation of American VIX put option under the $3/2$ model and shows monotonicity properties of the option price function and optimal exercise boundary.

Although the evidence provided in Goard and Mazur (2013) shows that the $3/2$ model dominates the alternatives considered, the analysis performed tests for overidentifying restrictions relative to a specific benchmark. This benchmark has a more general structure that nests the various alternatives tested. It nevertheless imposes specific functional forms on the coefficients of the VIX process. Unconstrained GMM shows that the benchmark has an estimated γ of 1.48 and places large weights on the various nonlinear components in the drift. These results suggest that specifications deviating from the standard $3/2$ model are of interest for capturing complex aspects of the VIX behavior.

A recent step in that direction is taken by Grasselli (2015), who introduces the $4/2$ model for local variance process, which is the sum of a $1/2$ and a $3/2$ models. Instantaneous volatility, in the 4/2 model, is $a\sqrt{Y} + b/\sqrt{Y}$ where a, b are positive constants and Y follows a CIR process. This model has several interesting features. Most notably, variance is bounded away from zero, as suggested by the stylized facts reported in Gatheral (2008). The model also helps to explain observed shapes of the implied volatility surface. Grasselli (2015) studies the behavior of the 4/2 price process and derives the characteristic function of the log price. He also provides an exact simulation scheme based on the conditional distribution of the price.

This paper has several contributions. First, it introduces two new classes of volatility processes, the generalized 3/2 class $(A1)$ and the generalized 1/2 class $(A2)$. These two classes contain a variety of processes that are natural extensions of the 3/2 and 1/2 processes, yet remain tractable for valuation purposes. The computations of vanilla options and futures on VIX can be executed efficiently by standard numerical integration methods. Also, we note that models in the generalized 3/2 class produce a positive skew of implied volatilities which is the most relevant stylized fact of the VIX market, as documented by Mencia and Sentana (2013). Second, it provides explicit formulas (in the form of integrals) for European and American call and put options when the underlying volatility follows any process in $(A1)$ or $(A2)$. In the American case, an early exercise premium representation formula is derived using the free-boundary approach and local time-space calculus (see Peskir (2005a)). The optimal exercise boundary for the VIX process is characterized as the unique solution to a nonlinear integral equation of Volterra type. Third, we show that the value function of the optimal stopping problem satisfies a smooth-fit property along the optimal exercise boundary in the case where the dependence on the initial value of the underlying process is unknown. To the best of our knowledge, existing papers considered problems where the underlying processes have explicit initial dependence, e.g., Brownian motion, geometric Brownian motion, etc. Another aspect, outlined in the paper, is that numerical computations show a non-convexity of the American call price function with respect to the initial value of the VIX under the 3/2 model (see Figure 3), but not in the 1/2 model (see Figure 4) and in the classical models, e.g., GBMP and MRLP models (see Detemple and Osakwe (2000)).

The final section of the paper is devoted to the pricing of the American call when the VIX is modelled as the mixture of the two classes of models above, i.e., the sum of generalized 3/2- and 1/2-type processes. Equivalently, the VIX process is a function of a CIR process where this function is the sum of functions of $(A1)$ and $(A2)$ types. We show that, under certain assumptions, there exists a pair of optimal exercise boundaries for the underlying CIR process that can be obtained as the unique solution to a system of coupled integral equations. We then provide the early exercise premium representation formula for the American call price. This formula decomposes it into the sum of a European part and an early exercise premium which depends on the pair of exercise boundaries.

The paper is organized as follows. Section 2 describes the two classes of processes that are the focus of this study, formulates the pricing problem for an American VIX call as an optimal stopping problem and shows how to price a European VIX call. An associated free-boundary problem for the American call option is studied in Section 3. Section 4 derives the early exercise premium representation for the American call price and characterizes the optimal exercise boundary as the unique solution to a nonlinear integral equation. Section 5 provides corresponding results for European and American put options. Section 6 studies the VIX call price under the mixture model. The paper is completed by a technical appendix.

2. The generalized 3/2 and 1/2 models and VIX options

1. First let us consider the following two classes of functions

$(A1)$ Generalized 3/2-type: let $f(\cdot): I\!R_+ \to I\!R_+$ be a continuously differentiable, strictly decreasing and convex function ranging from $+\infty$ to 0. Let $g(\cdot)$ be the inverse of $f(\cdot)$, i.e. $f(g(x)) = x$ for $x > 0$.

$(A2)$ Generalized 1/2-type: let $f(\cdot): \mathbb{R}_+ \to \mathbb{R}_+$ be a continuously differentiable, strictly increasing and weakly concave function ranging from 0 to ∞. Let $g(\cdot)$ be the inverse of $f(\cdot)$, i.e. $f(g(x)) = x$.

In this paper we model the VIX, under the historical measure P, as follows

$$X_t = f(Y_t) \tag{2.1}$$

for $t \geq 0$ where f is either of type $(A1)$ or $(A2)$ above and a factor process $Y = (Y_t)_{t\geq 0}$ is given by

$$dY_t = \big(\beta - \alpha Y_t\big)dt - \kappa\sqrt{Y_t}dB_t, \quad Y_0 = y \tag{2.2}$$

where $\alpha, \beta, \kappa > 0$ are constant parameters and B is a standard Brownian motion. The process Y solving (2.2) follows a mean-reverting square-root process (MRSRP) and the random variable Y_t^y has non-central chi-squared density function $q(\widetilde{y}; t, y)$ (see, e.g., Cox, Ross and Ingersoll (1985)). Throughout this paper, we assume that $\beta \geq \kappa^2/2$ as Feller showed that under this condition Y is strictly positive. Hence X is well defined and is strictly positive for all $t > 0$. We note that the functions f are strictly monotone so that there is one-to-one relationship between the VIX process X and the factor process Y.

By using Ito's formula, we get the dynamics of X

$$\begin{aligned} dX_t = &\left(f'(g(X_t))\,(\beta - \alpha g(X_t)) + \frac{1}{2}\kappa^2 f''(g(X_t))\,g(X_t)\right)dt \\ &- \kappa f'(g(X_t))\sqrt{g(X_t)}dB_t \end{aligned} \tag{2.3}$$

for $t \geq 0$. As X is not the price of a traded asset, one should allow for the possibility of a nonzero market price of risk $\lambda(t, X)$ associated with the VIX. Following papers by Stein and Stein (1991) and Grünbichler and Longstaff (1996), we assume that the market price of risk is such that the risk-neutral process for X is of the same form as the real process (2.3). For this one chooses $\lambda(t, X_t)$ as $a/\sqrt{g(X_t)} + b\sqrt{g(X_t)} + cf''(g(X_t))\sqrt{g(X_t)}/f'(g(X_t))$. To avoid additional notations, we assume that the dynamics (2.3) is under some risk neutral measure Q and B is Q-SBM.

The specification (2.3) of type $(A1)$ includes several models of potential interest to describe the evolution of the VIX.

Example 2.1. *($3/2$-model) The 3/2-model is introduced by Goard and Mazur (2013). It is obtained by taking $f(y) = 1/y$. Then, $g(x) = 1/x$, $f'(y) = -1/y^2$, $f''(y) = 2/y^3$ and*

$$dX_t = \big(\alpha X_t - \big(\beta - \kappa^2\big)X_t^2\big)dt + \kappa X_t^{3/2}dB_t. \tag{2.4}$$

The 3/2 model has elasticity of variance equal to $\varepsilon = 3$. It also displays mean reversion if $\beta > \kappa^2$. The speed of mean reversion $(\beta - \kappa^2)X_t$ is linear in the VIX. The constant attractor is $\alpha/(\beta - \kappa^2)$.

Example 2.2. *($\nu+1/2$ -model) Let* $f(y) = 1/y^\nu$ *where* $\nu > 0$ *and* $\beta > \frac{1}{2}\kappa^2(\nu+1)$ *. Then,* $g(x) = (1/x)^{1/\nu}$ *,* $f'(y) = -\nu/y^{\nu+1}$ *,* $f''(y) = \nu(\nu+1)/y^{\nu+2}$ *and*

$$dX_t = \nu\left(\alpha X_t - \left(\beta - \frac{1}{2}\kappa^2(\nu+1)\right)X_t^{1+1/\nu}\right)dt + \nu\kappa X_t^{\nu+1/2}dB_t. \tag{2.5}$$

For this specification the elasticity of variance is $\varepsilon = 2\nu+1$ *. The process has linear speed of mean reversion* $\nu\left(\beta - \frac{1}{2}\kappa^2(\nu+1)\right)X_t$ *and constant attractor* $\left(\alpha/\left(\beta-\frac{1}{2}\kappa^2(\nu+1)\right)\right)^\nu$ *. The* $3/2$ *model is obtained when* $\nu = 1$ *.*

Example 2.3. *(mixture* $\nu_j+1/2$ *,* $j = 1,...,n$ *model) Let* $f(y) = \sum_j \omega_j/y^{\nu_j}$ *where* $\nu_j \geq 1$ *,* $\omega_j > 0, j = 1,...,n$ *so that*

$$\sum_j \omega_j \frac{1}{g(x)^{\nu_j}} = x, \qquad f'(y) = -\sum_j \omega_j \frac{\nu_j}{y^{\nu_j+1}}, \qquad f''(y) = \sum_j \omega_j \frac{\nu_j(\nu_j+1)}{y^{\nu_j+2}}$$

and

$$dX_t = \left(\alpha \sum_j \omega_j \frac{\nu_j}{g(X_t)^{\nu_j}} - \left(\beta - \tfrac{1}{2}\kappa^2\right)\sum_j \omega_j \frac{\nu_j(\nu_j+1)}{g(X_t)^{\nu_j+1}}\right)dt - \kappa\sum_j \omega_j \frac{\nu_j}{g(X_t)^{\nu_j+1/2}}dB_t. \tag{2.6}$$

The elasticity of variance is a non-linear function of the VIX. The process has non-linear speed of mean reversion and a constant attractor.

The specification (2.3) of type $(A2)$ contains another set of relevant models for the VIX.

Example 2.4. *($1/2$ -model) See e.g. Grunblicher and Longstaff (1996). It is obtained by taking* $f(y) = y$ *, a weakly concave function. Then* $g(x) = x$ *,* $f'(y) = 1$ *,* $f''(y) = 0$ *and*

$$dX_t = (\beta - \alpha X_t)\,dt - \kappa X_t^{1/2}dB_t. \tag{2.7}$$

The $1/2$ *-model has elasticity of variance equal to* $\varepsilon = 1$ *. It also displays mean reversion. The speed of mean reversion* α *is constant. The long run mean is* β/α *.*

Example 2.5. *($1\text{–}1/(2\nu)$ model) Let* $f(y) = y^\nu$ *where* $\nu \in (0,1]$ *and* $\beta + \frac{1}{2}\kappa^2(\nu-1) > 0$ *. Then,* $g(x) = x^{1/\nu}$ *,* $f'(y) = \nu y^{\nu-1}$ *,* $f''(y) = \nu(\nu-1)y^{\nu-2}$ *and*

$$dX_t = \nu X_t^{1-1/\nu}\left(\beta + \frac{1}{2}\kappa^2(\nu-1) - \alpha X_t^{1/\nu}\right)dt - \kappa\nu X_t^{1-1/(2\nu)}dB_t. \tag{2.8}$$

The $1\text{–}1/(2\nu)$ *model has non-linear elasticity of variance* $\varepsilon(x) = (2-1/\nu)/x$ *. It also displays non-linear mean reversion with speed of mean reversion* $\alpha\nu X_t^{1-1/\nu}$ *. The attracting value is* $\left(\left(\beta + \frac{1}{2}\kappa^2(\nu-1)\right)/\alpha\right)^\nu$ *. The* $1/2$ *- model is obtained for* $\nu = 1$ *.*

Example 2.6. *(mixture* $1\text{–}1/(2\nu_j)$ *,* $j = 1,...,n$ *model) Let* $f(y) = \sum_j \omega_j y^{\nu_j}$ *where* $\nu_j \geq 1$ *,* $\omega_j > 0, j = 1,...,n$ *and* $\beta > \frac{1}{2}\kappa^2(1-\nu_j)$ *,* $j = 1,...,n$ *. Then,*

$$\sum_j \omega_j g(x)^{\nu_j} = x, \qquad f'(y) = \sum_j \omega_j \nu_j y^{\nu_j-1}, \qquad f''(y) = \sum_j \omega_j \nu_j(\nu_j-1)\,y^{\nu_j-2}$$

and

$$(2.9)\qquad dX_t = \left(\sum_j \omega_j \nu_j g\left(X_t\right)^{\nu_j-1}\left(\beta-\frac{1}{2}\kappa^2\left(1-\nu_j\right)\right)-\alpha\sum_j \omega_j\nu_j g\left(X_t\right)^{\nu_j}\right)dt - \kappa\sum_j \omega_j\nu_j g\left(X_t\right)^{\nu_j-1/2}dB_t.$$

The elasticity of variance is a non-linear function of the VIX. The process has non-linear speed of mean reversion and a non-linear attractor. This mixture model behaves as the $\max_j \nu_j+1/2$ *model as* $x\to 0$ *. It behaves as the* $\min_j \nu_j+1/2$ *model as* $x\to\infty$ *.*

2. Here, we justify the relevance and choice of models based on the classes of functions $(A1)$ and $(A2)$ and the process Y. As shown by Goard and Mazur (2013), the 3/2-model (Example 2.1) provides a better fit to the VIX data than various alternatives including GBMP, MRGP, MRSRP and MRLP. Notable features of this model are (i) a high power law of 3/2 which can reduce the heteroskedasticity of volatility and (ii) a nonlinear drift that generates substantial nonlinear mean-reverting behaviour when the volatility exceeds its long-run mean. Another important feature of this framework is that it reproduces the positive skew of implied volatilities which is the most relevant stylized fact of the VIX market, see, e.g., Mencia and Sentana (2013). We note that all the models in Examples 2.2-2.3 exhibit this important property.

However, there are at least two reasons to consider generalizations of the 3/2 model. Firstly, Goard and Mazur (2013) performed tests for overidentifying restrictions relative to a specific benchmark which has a more general structure that nests the various alternatives tested. Unconstrained GMM shows that the benchmark has an estimated γ of 1.48 and places large weights on the various nonlinear components in the drift. These findings motivate us to vary the power of the diffusion coefficient (Example 2.2) and combine different powers (Example 2.3) in order to obtain a better fit to the VIX data. Secondly, if one chooses parameters under a risk neutral measure to exactly match at-the-money vanilla options then the 3/2 model generally undervalues both in- and out-of-the money option prices, see, e.g., Section 7 in Goard and Mazur (2013). By taking models in Examples 2.2. and 2.3 one can adjust the tail behaviour of VIX either at 0 or at high levels and therefore improve model prices for in-the-money or out-of-the money vanillas compared to Example 2.1.

A thorough empirical analysis of the models introduced here is clearly needed. This is left for future research as the main aim of this paper is to provide a rigorous analysis of American options under these new specifications for VIX and numerical illustrations of the theoretical results. Nevertheless, based on the discussion above, it is reasonable to introduce the generalized 3/2 and 1/2 models. Note also that building on Mencia and Sentana (2013), it might be useful to specify the parameter β of Y in (2.2) as a stochastic process instead of a constant in order to improve the fit to the VIX futures term structure. In this case, the American option pricing problem becomes a three-dimensional optimal stopping problem. This extension is left for future research as well.

3. In this paper, we study the American call and put VIX options under the model (2.3) with f of types $(A1)$ and $(A2)$. By definition, the payoff of the American VIX call at exercise time $\tau\in[0,T]$ is $(X_\tau-K)^+ := \max(X_\tau-K,0)$ where $K>0$ is the strike and $T>0$ is the expiry date. The rational price C^A of the American VIX call at time $t=0$ is the value

function of the following optimal stopping problem

$$C^A = \sup_{0 \le \tau \le T} \mathsf{E} e^{-r\tau}(X_\tau - K)^+ \tag{2.10}$$

where the supremum is taken over all stopping times τ of process X, the expectation E is taken under a risk neutral measure Q and $r > 0$ is the constant interest rate.

As the process X is time-homogeneous Markov and (2.10) is a finite horizon problem, we will study the problem (2.10) in the Markovian setting and hence, we introduce dependence on time t and the initial value of X:

$$C^A(t,x) = \sup_{0 \le \tau \le T-t} \mathsf{E} e^{-r\tau} G(X_\tau^x) \tag{2.11}$$

for $t \in [0,T)$ and $x > 0$ where X^x means that the process X starts from $X_0^x = x$ and the payoff function G is given by

$$G(x) := (x - K)^+ \tag{2.12}$$

for $x > 0$. We tackle the problem (2.11) in Sections 3 and 4. The discussion of the American put option follows in Section 5.

4. Now we introduce the rational price function of European VIX call option

$$C^E(t,x) = e^{-r(T-t)} \mathsf{E}(X_{T-t}^x - K)^+ \tag{2.13}$$

for $t \in [0,T)$ and $x > 0$. A formula for (2.13) in the $3/2$ model was derived by Goard and Mazur (2013) using the fact that the process $(1/X_t)_{t \ge 0}$ is a mean-reverting square-root process. We exploit a similar idea and recall that $X_t = f(Y_t)$ so that using the known probability density function $q(\widetilde{y}; t, y)$ of Y_t, one can compute (2.13) by numerical integration in an efficient way for f of $3/2$ type as follows

$$C^E(t,x) = e^{-r(T-t)} \int_0^{g(K)} (f(\widetilde{y}) - K)\, q(\widetilde{y}; u, g(x))\, d\widetilde{y} \tag{2.14}$$

for $t \in [0,T)$ and $x > 0$ as g is decreasing in this case. When f is of $1/2$ type so that g is increasing, the European price is

$$C^E(t,x) = e^{-r(T-t)} \int_{g(K)}^{\infty} (f(\widetilde{y}) - K)\, q(\widetilde{y}; u, g(x))\, d\widetilde{y}. \tag{2.15}$$

3. The free-boundary problem for the American VIX call option

In this section we will reduce the problem (2.11) to a free-boundary problem and the latter will be tackled in the next section using the local time-space calculus (see Peskir (2005a)). First, using that the payoff function $G(x)$ is continuous and standard arguments (see e.g.

Corollary 2.9 (Finite horizon) with Remark 2.10 in Peskir and Shiryaev (2006)), we have that the continuation and exercise regions read, respectively

$$\mathcal{C} = \{\, (t,x) \in [0,T)\times[0,\infty) : C^A(t,x) > G(x) \,\} \tag{3.1}$$

$$\mathcal{E} = \{\, (t,x) \in [0,T)\times[0,\infty) : C^A(t,x) = G(x) \,\} \tag{3.2}$$

and the optimal stopping time in (2.11) is given by

$$\tau = \inf \{\ 0 \le s \le T : (t+s, X_s^x) \in \mathcal{E}\ \}. \tag{3.3}$$

Before starting our analysis, we recall an important result for our purposes on flows of stochastic differential equations. The underlying model satisfies the conditions of Theorem 37 of Chapter V, Section 7 in Protter (1990), i.e. which simply requires only Lipschitz coefficients of SDE (2.3), so that we have the following inequality

$$\Big[\mathsf{E} \sup_{0\le u\le T-t} (X_u^x - X_u^y)^2\Big]^{1/2} \le C_L\, |x-y| \tag{3.4}$$

for $x,y>0$ and some constant $C_L>0$. We will use this estimate for the proof of the smooth-fit property.

1. We show that the price function C^A is continuous on $[0,T)\times(0,\infty)$. It follows that

$$\begin{aligned}
0 \le C^A(t,x) - C^A(t,y) &\le \sup_{0\le\tau\le T-t} \mathsf{E} e^{-r\tau}\,(X_\tau^x - X_\tau^y) \le \mathsf{E} \sup_{0\le u\le T-t} (X_u^x - X_u^y) \\
&\le \Big(\mathsf{E} \sup_{0\le u\le T-t} (X_u^x - X_u^y)^2\Big)^{1/2} \le C_L(x-y)
\end{aligned} \tag{3.5}$$

for $x\ge y$ and $t\in[0,T)$ where we used that $\sup(f)-\sup(g)\le \sup(f-g)$ and $(x-K)^+ - (y-K)^+ \le (x-y)^+$ for $x,y,K\in I\!R$, the comparison theorem for solutions of SDEs (i.e. $Q(X_s^x \ge X_s^y,\ s\ge 0)=1$), Holder inequality and the inequality (3.4) . From (3.5) we see that $x\mapsto C^A(t,x)$ is continuous uniformly over $t\in[0,T]$. Thus to prove that C^A is continuous on $[0,T)\times(0,\infty)$ it is enough to show that $t\mapsto C^A(t,x)$ is continuous on $[0,T]$ for each $x>0$ given and fixed. For this, take any $t_1<t_2$ in $[0,T]$ and let τ_1 be an optimal stopping time for $C^A(t_1,x)$. Setting $\tau_2=\tau_1\wedge(T-t_2)$ and using that $t\mapsto C^A(t,x)$ is decreasing on $[0,T]$, we have

$$0 \le C^A(t_1,x) - C^A(t_2,x) \le \mathsf{E} e^{-r\tau_1} G(X_{\tau_1}^x) - \mathsf{E} e^{-r\tau_2} G(X_{\tau_2}^x) \le \mathsf{E}\big(X_{\tau_1}^x - X_{\tau_2}^x\big)^+. \tag{3.6}$$

Letting first $t_2-t_1\to 0$ and using $\tau_1-\tau_2\to 0$ we see that $C^A(t_1,x)-C^A(t_2,x)\to 0$ by dominated convergence. This shows that $t\mapsto C^A(t,x)$ is continuous on $[0,T]$, and the proof of the initial claim is complete.

2. Now we get some initial insights into the structure of exercise region $\mathcal{E}$.

(i) We first calculate the function $H(x) := (\text{Ł}_X G - rG)(x)$ for $x\in(0,\infty)$ (which is the instantaneous benefit of waiting to exercise) where

$$\text{Ł}_X = \Big(f'(g(x))\big(\beta-\alpha g(x)\big) + \frac{1}{2}\kappa^2 f''(g(x))g(x)\Big)\frac{d}{dx} + \frac{1}{2}\kappa^2 (f'(g(x)))^2 g(x)\,\frac{d^2}{dx^2} \tag{3.7}$$

is the infinitesimal generator of X . As $G(x)=(x-K)^+$, we have that

$$H(x) = h(x)I(x \geq K) \tag{3.8}$$

for $x\in(0,\infty)$ where

$$h(x) = f'(g(x))\big(\beta - \alpha g(x)\big) + \frac{1}{2}\kappa^2 f''(g(x))g(x) - r(x-K) \tag{3.9}$$

for $x>0$. Throughout the paper, the following condition is imposed on the model and we note that all models in Examples 2.1-2.6 satisfy this assumption (the verification is provided in the Appendix):

Assumption R: There exists $x^*>0$ such that $h(x)\geq 0$ if and only if $x\leq x^*$.

We could assume a weaker condition, that there exists $x^*>0$ such that $H(x)\geq 0$ if and only if $x\leq \max(K,x^*)$. We use **Assumption R** in order to have a unified condition for both call and put options, and it is enough for models of interest such as Examples 2.1-2.6.

(ii) We now use the Ito-Tanaka's formula and the definition of H to obtain

$$\mathsf{E}e^{-r\tau}G(X^x_\tau) = \; G(x) + \mathsf{E}\int_0^\tau e^{-rs}H(X^x_s)ds + \frac{1}{2}\mathsf{E}\int_0^\tau e^{-rs}d\ell^K_s(X^x) \tag{3.10}$$

for $x\in(0,\infty)$ and any stopping time τ of the process X where $(\ell^K_s(X))_{s\geq 0}$ is the local time process of X at level K

$$\ell^K_s(X^x) := Q - \lim_{\varepsilon\downarrow 0}\frac{1}{2\varepsilon}\int_0^s I(K-\varepsilon < X^x_u < K+\varepsilon)d\langle X,X\rangle_u \tag{3.11}$$

and $d\ell^K_s(X^x)$ refers to the integration with respect to the continuous increasing function $s\mapsto \ell^K_s(X^x)$. The equation (3.10) and **Assumption R** show that it is not optimal to exercise the call option when $X_t\leq \max(K,x^*)$ as $H(X_t)\geq 0$ in this region and thus both integral terms on the right-hand side of (3.10) are non-negative. This fact can be also explained in the particular case where $X_t<K$ as follows: by exercising below K , the option holder receives a null payoff, however by waiting would have a positive probability of collecting a strictly positive payoff in the future.

Another implication of (3.10) is that the exercise region is non-empty for all $t\in[0,T)$, as for large $x\uparrow\infty$ the integrand H is negative and the local time term is zero, and thus due to a lack of time to compensate the negative H , it is optimal to stop at once.

3. Next we prove further properties of the exercise region $\mathcal{E}$ and define the optimal exercise boundary.

(i) As the payoff function in (2.11) is time-independent, it follows that the map $t\mapsto C^A(t,x)$ is non-increasing on $[0,T]$ for each $x>0$ so that $C^A(t_1,x)-G(x)\geq C^A(t_2,x)-G(x)\geq 0$ for $0\leq t_1<t_2<T$ and $x\in(0,\infty)$. Now, if we take a point $(t_1,x)\in\mathcal{E}$, i.e. $C^A(t_1,x)=G(x)$, then $(t_2,x)\in\mathcal{E}$ as well, which shows that the exercise region is increasing in t . In other words, $\mathcal{E}$ is right-connected.

(ii) Now let us take $t>0$ and $x>y>\max(K,x^*)$ such that $(t,y)\in\mathcal{E}$. Then, by right-connectedness of the exercise region, we have that $(s,y)\in\mathcal{E}$ as well for any $s>t$.

If we now run the process (t, X_t) from (t,x), we cannot hit the level $\max(K, x^*)$ before exercise (as $x > y$), thus the local time term in (3.10) is 0 and integrand H is negative (by **Assumption R**). Therefore it is optimal to exercise at (t,x) and we get up-connectedness of the exercise region $\mathcal{E}$.

(iii) From *(i)* - *(ii)* and paragraph 2 *(ii)* above we can conclude that there exists an optimal exercise boundary $b : [0,T] \to (0,\infty)$ such that

$$\tau_b = \inf\ \{\ 0 \le s \le T-t : X_s^x \ge b(t+s)\ \} \tag{3.12}$$

is optimal in (2.11) and $\max(K, x^*) < b(t) < \infty$ for $t \in [0,T)$. Moreover, b is decreasing on $[0,T)$.

Remark 3.1. *If* ***Assumption R*** *does not hold and the function $h(x)$ changes sign more than once for $x > K$, then there are the same number of exercise boundaries. Therefore the exercise region $\mathcal{E}$ is not up-connected.*

4. Now we prove that the smooth-fit condition along the boundary b holds

$$C_x^A(t, b(t)-) = C_x^A(t, b(t)+) = G'(b(t)) = 1 \tag{3.13}$$

for all $t \in [0,T)$. To the best of our knowledge, in the literature on optimal stopping problems, the smooth-fit property has been proven in models where the dependence of X^x on x is given explicitly (e.g. Brownian motion or geometric Brownian motion), however in our model such dependence is unknown. For this reason, we provide another proof based on the inequality (3.4).

(i) First let us fix a point $(t,x) \in [0,T) \times (0,\infty)$ lying on the boundary b so that $x = b(t)$. Then we have

$$\frac{C^A(t,x) - C^A(t,x-\varepsilon)}{\varepsilon} \le \frac{G(x) - G(x-\varepsilon)}{\varepsilon} \tag{3.14}$$

and taking the limit as $\varepsilon \downarrow 0$, we get

$$C_x^A(t, x-) \le G'(x) = 1 \tag{3.15}$$

where the left-hand derivative exists by monotonicity of $x \mapsto C^A(t,x)$ on $(0,\infty)$ for any fixed $t \in [0,T)$.

(ii) To prove the reverse inequality, we set $\tau_\varepsilon = \tau_\varepsilon(t, x-\varepsilon)$ as an optimal stopping time for $C^A(t, x-\varepsilon)$. Using that X is a regular diffusion and $t \mapsto b(t)$ is decreasing we have that $\tau_\varepsilon \to 0$ as $\varepsilon \to 0$ Q-a.s. By the comparison theorem for solutions of SDEs and noting that

$$\begin{aligned} &G(X_{\tau_\varepsilon}^x) - G(X_{\tau_\varepsilon}^{x-\varepsilon}) \\ &\quad = \big(X_{\tau_\varepsilon}^x - X_{\tau_\varepsilon}^{x-\varepsilon}\big)\, I(X_{\tau_\varepsilon}^{x-\varepsilon} \ge K) + (X_{\tau_\varepsilon}^x - K) I(X_{\tau_\varepsilon}^x \ge K \ge X_{\tau_\varepsilon}^{x-\varepsilon}) \\ &\quad \ge \big(X_{\tau_\varepsilon}^x - X_{\tau_\varepsilon}^{x-\varepsilon}\big)\, I(X_{\tau_\varepsilon}^{x-\varepsilon} \ge K) \end{aligned} \tag{3.16}$$

we obtain

$$\frac{1}{\varepsilon}\Big(C^A(t,x) - C^A(t,x-\varepsilon)\Big) \tag{3.17}$$

$$
\begin{aligned}
&\geq \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(X^x_{\tau_\varepsilon} - X^{x-\varepsilon}_{\tau_\varepsilon}\right) I(X^{x-\varepsilon}_{\tau_\varepsilon} \geq K)\right] \\
&= \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(X^x_{\tau_\varepsilon} - X^{x-\varepsilon}_{\tau_\varepsilon}\right)\right] - \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(X^x_{\tau_\varepsilon} - X^{x-\varepsilon}_{\tau_\varepsilon}\right) I(X^{x-\varepsilon}_{\tau_\varepsilon} \leq K)\right].
\end{aligned}
$$

Then the second term on the right-hand side of (3.17) goes to 0 as $\varepsilon \to 0$ as

$$
\begin{aligned}
0 &\leq \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(X^x_{\tau_\varepsilon} - X^{x-\varepsilon}_{\tau_\varepsilon}\right) I(X^{x-\varepsilon}_{\tau_\varepsilon} \leq K)\right] \\
&\leq \frac{1}{\varepsilon}\left(\mathsf{E}\left(X^x_{\tau_\varepsilon} - X^{x-\varepsilon}_{\tau_\varepsilon}\right)^2\right)^{1/2}\left(Q(X^{x-\varepsilon}_{\tau_\varepsilon} \leq K)\right)^{1/2} \\
&\leq \frac{1}{\varepsilon}\left(\mathsf{E}\sup_{0\leq u\leq T-t}\left(X^x_u - X^{x-\varepsilon}_u\right)^2\right)^{1/2}\left(Q(X^{x-\varepsilon}_{\tau_\varepsilon} \leq K)\right)^{1/2} \\
&\leq C_L\left(Q(X^{x-\varepsilon}_{\tau_\varepsilon} \leq K)\right)^{1/2} \to 0
\end{aligned} \tag{3.18}
$$

where we used the Holder inequality, the inequality (3.4) and that the latter probability goes to zero because $x > K$. Now we turn to the first term on the right-hand side of (3.17). Using Ito's formula, we have:

$$
\frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(X^x_{\tau_\varepsilon} - X^{x-\varepsilon}_{\tau_\varepsilon}\right)\right] = 1 + \frac{1}{\varepsilon}\mathsf{E}\left[\int_0^{\tau_\varepsilon} e^{-rs}\left(\omega(X^x_s) - \omega(X^{x-\varepsilon}_s)\right) ds\right] \tag{3.19}
$$

where $\omega(x) := f'(g(x))\big(\beta - \alpha g(x)\big) + \frac{1}{2}\kappa^2 f''(g(x))g(x) - rx$ for $x > 0$. We show that the second term of (3.19) goes to 0 as $\varepsilon \to 0$

$$
\begin{aligned}
0 &\leq \frac{1}{\varepsilon}\mathsf{E}\left|\int_0^{\tau_\varepsilon} e^{-rs}(\omega(X^x_s) - \omega(X^{x-\varepsilon}_s))ds\right| \leq \frac{1}{\varepsilon}\left[\mathsf{E}\int_0^{\tau_\varepsilon} e^{-rs}|\omega'(\xi_s)|(X^x_s - X^{x-\varepsilon}_s)ds\right] \\
&\leq \frac{1}{\varepsilon}C_{\omega'}\mathsf{E}\left[\tau_\varepsilon\sup_{0\leq u\leq T-t}\left(X^x_u - X^{x-\varepsilon}_u\right)\right] \leq \frac{1}{\varepsilon}C_{\omega'}\left(\mathsf{E}\tau^2_\varepsilon\right)^{1/2}\left(\mathsf{E}\sup_{0\leq u\leq T-t}\left(X^x_u - X^{x-\varepsilon}_u\right)^2\right)^{1/2} \\
&\leq \frac{1}{\varepsilon}C_{\omega'}C_L\,\varepsilon\left(\mathsf{E}\tau^2_\varepsilon\right)^{1/2} = C_\omega C_L\left(\mathsf{E}\tau^2_\varepsilon\right)^{1/2} \to 0
\end{aligned} \tag{3.20}
$$

where we used the mean value theorem and choice $\xi_s \in [X^{x-\varepsilon}_s, X^x_s]$, the fact that $|\omega'(\cdot)| \leq C_{\omega'}$ for some $C_{\omega'} > 0$, Holder inequality, then inequality (3.4) and that $\mathsf{E}\tau^2_\varepsilon \to 0$ as $\varepsilon \to 0$ by the dominated convergence theorem.

Thus, using (3.17)-(3.20) and taking the limits as $\varepsilon \to 0$ we have that

$$
C^A_x(t, x-) \geq G'(x) = 1 \tag{3.21}
$$

for $t \in [0, T)$. Thus, combining (3.15) and (3.21), we obtain (3.13).

5. Here we prove that the boundary b is continuous on $[0, T]$ and that $b(T-) = \max(K, x^*)$. The proof is provided in 3 steps and follows the approach proposed by De Angelis (2014).

(i) We first show that b is right-continuous. Let us fix $t \in [0, T)$ and take a sequence $t_n \downarrow t$ as $n \to \infty$. As b is decreasing, the right-limit $b(t+)$ exists and $(t_n, b(t_n))$ belongs to $\mathcal{E}$ for all $n \geq 1$. Recall that $\mathcal{E}$ is closed so that $(t_n, b(t_n)) \to (t, b(t+)) \in \mathcal{E}$ as $n \to \infty$ and we may conclude that $b(t+) \geq b(t)$. The fact that b is decreasing gives the reverse inequality and thus b is right-continuous as claimed.

(ii) Now we prove that b is also left-continuous. Assume that there exists $t_0 \in (0,T)$ such that $b(t_0-) > b(t_0)$. Let us set $x_1 = b(t_0)$ and $x_2 = b(t_0-)$ so that $x_1 < x_2$. For $\varepsilon \in (0,(x_2-x_1)/2)$ given and fixed, let $\varphi_\varepsilon : (0,\infty) \to [0,1]$ be a C^∞ - function satisfying (i) $\varphi_\varepsilon(x) = 1$ for $x \in [x_1+\varepsilon, x_2-\varepsilon]$ and (ii) $\varphi_\varepsilon(x) = 0$ for $x \in (0,x_1+\varepsilon/2] \cup [x_2-\varepsilon/2,\infty)$. Letting Ł_X^* denote the adjoint of Ł_X , recalling that $t \to C^A(t,x)$ is decreasing on $[0,T]$ and that $C_t^A + \text{Ł}_X C^A - rC^A = 0$ on $\mathcal{C}$, we find integrating by parts (twice) that

$$0 \geq \int_{x_1}^{x_2} \varphi(x) C_t^A(t_0-\delta,x)dx = -\int_{x_1}^{x_2} C^A(t_0-\delta,x)\left(\text{Ł}_X^*\varphi(x) - r\varphi(x)\right)dx \tag{3.22}$$

for $\delta \in (0,t_0 \wedge (\varepsilon/2))$ so that $\varphi_\varepsilon(x_2-\delta) = \varphi_\varepsilon'(x_2-\delta) = 0$ as needed. Letting $\delta \downarrow 0$ it follows using the dominated convergence theorem and integrating by parts (twice) that

$$\begin{aligned} 0 &\geq -\int_{x_1}^{x_2} C^A(t_0,x)\left(\text{Ł}_X^*\varphi(x) - r\varphi(x)\right)dx = -\int_{x_1}^{x_2} G(x)\left(\text{Ł}_X^*\varphi(x) - r\varphi(x)\right)dx \\ &= -\int_{x_1}^{x_2} \left(\text{Ł}_X G(x) - rG(x)\right)\varphi(x)dx = -\int_{x_1}^{x_2} H(x)\varphi(x)dx. \end{aligned} \tag{3.23}$$

Letting $\varepsilon \downarrow 0$ we obtain

$$0 \geq -\int_{x_1}^{x_2} H(x)dx > 0 \tag{3.24}$$

as $x \to H(x)$ is strictly negative on $(x_1,x_2]$. We thus have a contradiction and therefore we may conclude that b is continuous on $[0,T)$ as claimed.

(iii) To prove that $b(T-) = \max(K,x^*)$ we can use the same arguments as those in (ii) above with $t_0 = T$ and suppose that $b(T-) > \max(K,x^*)$.

6. The facts proved in paragraphs 1-5 above and standard arguments based on the strong Markov property (see, e.g., Peskir and Shiryaev (2006)) lead to the following free-boundary problem for the value function C^A and unknown boundary b :

$$C_t^A + \text{Ł}_X C^A - rC^A = 0 \quad \text{in } \mathcal{C} \tag{3.25}$$

$$C^A(t,b(t)) = G(b(t)) = b(t) - K \quad \text{for } t \in [0,T) \tag{3.26}$$

$$C_x^A(t,b(t)) = G'(b(t)) = 1 \quad \text{for } t \in [0,T) \tag{3.27}$$

$$C^A(t,x) > G(x) \quad \text{in } \mathcal{C} \tag{3.28}$$

$$C^A(t,x) = G(x) \quad \text{in } \mathcal{E} \tag{3.29}$$

where the continuation set $\mathcal{C}$ and the exercise set $\mathcal{E}$ are given by

$$\mathcal{C} = \{\, (t,x) \in [0,T) \times (0,\infty) : x < b(t) \,\} \tag{3.30}$$

$$\mathcal{E} = \{\, (t,x) \in [0,T) \times (0,\infty) : x \geq b(t) \,\}. \tag{3.31}$$

The following properties of C^A and b were also verified above:

$$C^A \text{ is continuous on } [0,T] \times (0,\infty) \tag{3.32}$$

$$(3.33)\qquad C^A \text{ is } C^{1,2} \text{ on } \mathcal{C}$$

$$(3.34)\qquad x \mapsto C^A(t,x) \text{ is increasing on } [0,\infty) \text{ for each } t \in [0,T]$$

$$(3.35)\qquad t \mapsto C^A(t,x) \text{ is decreasing on } [0,T] \text{ for each } x \in [0,\infty)$$

$$(3.36)\qquad t \mapsto b(t) \text{ is decreasing and continuous on } [0,T] \text{ with } b(T-) = \max(K, x^*).$$

4. The rational price of the American VIX call option

We will show in this section that the optimal exercise boundary b can be obtained as the unique solution to a nonlinear integral equation of Volterra type. We then provide the early exercise premium representation formula for the rational price C^A which decomposes it into the sum of the European VIX call price C^E and the early exercise premium which depends on the exercise boundary b .

1. We recall that we already showed how to compute the European call price in Section 2 above. Now we denote the following function:

$$(4.1)\qquad L(u,x,z) = -\mathsf{E}\big[e^{-ru} H(X_u^x) I(X_u^x \ge z)\big]$$

for $u \ge 0$ and $x,z > 0$. This function can be computed using the same idea as for the European call price. Using that $Y_t = g(X_t^x)$ is the mean-reverting square-root process and the random variable Y_t^y has non-central chi-squared density function $q(\widetilde{y};t,y)$, we have for f of $3/2$ - type that

$$(4.2)\qquad \begin{aligned} L(u,x,z) &= -\mathsf{E}\Big[e^{-ru} H\big(f(Y_u^{g(x)})\big) I(Y_u^{g(x)} \le g(z))\Big] \\ &= -e^{-ru} \int_0^{g(z)} H(f(\widetilde{y}))\, q(\widetilde{y};u,g(x))\, d\widetilde{y} \end{aligned}$$

for $u \ge 0$ and $x,z > 0$ as g is decreasing in this case. When f is of $1/2$ - type so that g is increasing, the function L can be computed as

$$(4.3)\qquad \begin{aligned} L(u,x,z) &= -\mathsf{E}\Big[e^{-ru} H\big(f(Y_u^{g(x)})\big) I(Y_u^{g(x)} \ge g(z))\Big] \\ &= -e^{-ru} \int_{g(z)}^{\infty} H(f(\widetilde{y}))\, q(\widetilde{y};u,g(x))\, d\widetilde{y} \end{aligned}$$

for $u \ge 0$ and $x,z > 0$.

2. The main result of this section can now be stated as follows.

Theorem 4.1. *The price function C^A in (2.11) has the representation*

$$(4.4)\qquad C^A(t,x) = C^E(t,x) + \int_0^{T-t} L(u,x,b(t+u))du$$

for $t \in [0,T]$ and $x \in (0,\infty)$. The optimal exercise boundary b in (2.11) can be characterized as the unique solution to the nonlinear integral equation of Volterra type

$$(4.5)\qquad b(t) - K = C^E(t,b(t)) + \int_0^{T-t} L(u,b(t),b(t+u))du$$

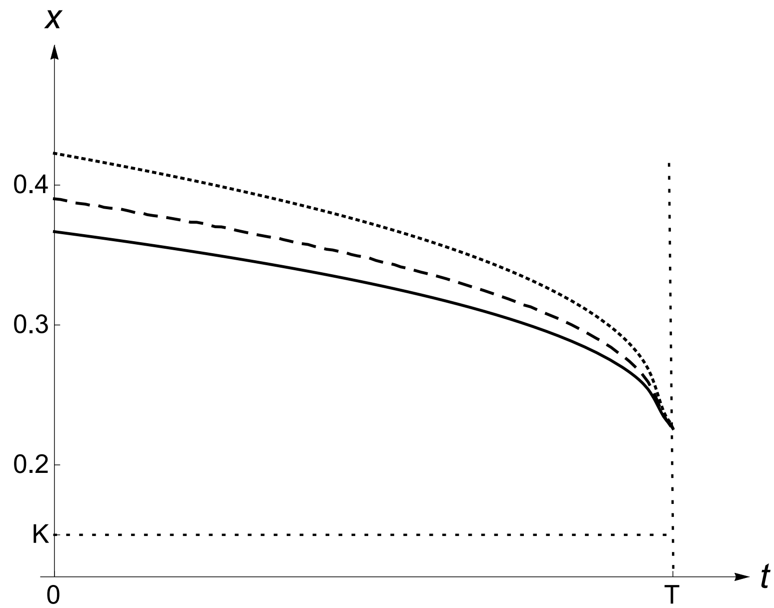

Figure 1. This figure plots the optimal exercise boundary b for models of $3/2$-type: $3/2$-model (solid), $\nu+1/2$-model with $\nu = 1.2$ (dotted), and mixture $(3/2, \nu{+}1/2)$-model with $\nu = 1.2$ and weights $w_j = 0.5$ (dashed). The parameter set is $T = 1, K = 0.15, r = 0.05$, and the coefficients for the $(3, 2)$-model are $\alpha = 2.94, \beta = 17.10, \kappa = 2.05$. For the $\nu{+}1/2$-model, the coefficient $\alpha = 3.64$ is adjusted to have the same attractor as for the $3/2$-model; for the mixture model the coefficient the α is 3.27. The parameters for the $3/2$-model were calibrated by Goard and Mazur (2013).

for $t \in [0, T]$ *in the class of continuous functions* $t \mapsto b(t)$ *with* $b(T) = \max(K, x^*)$ *(See Figures 1 and 2).*

Proof. (A) First, we clearly have that the following conditions hold: (i) C^A is $C^{1,2}$ on $\mathcal{C} \cup \mathcal{E}$; (ii) b is of bounded variation (due to monotonicity); (iii) $C^A_t + \text{Ł}_X C^A - rC^A$ is locally bounded; (iv) $C^A_{xx} = F_1 + F_2$ on $\mathcal{C} \cup \mathcal{E}$, where F_1 is non-negative and F_2 is continuous on $[0, T) \times (0, \infty)$; (v) $t \mapsto C^A_x(t, b(t)\pm)$ is continuous (recall (3.27)). Hence, we can apply the local time-space formula on curves (Peskir (2005a)) for $e^{-rs} C^A(t{+}s, X^x_s)$:

$$
\begin{aligned}
(4.6) \qquad & e^{-rs} C^A(t{+}s, X^x_s) \\
&= C^A(t, x) + M_s \\
&\quad + \int_0^s e^{-ru} \left(C^A_t + \text{Ł}_X C^A - rC^A\right)(t{+}u, X^x_u) I(X^x_u \neq b(t{+}u)) du \\
&\quad + \frac{1}{2} \int_0^s e^{-ru} \left(C^A_x(t{+}u, X^x_u{+}) - C^A_x(t{+}u, X^x_u{-})\right) I\big(X^x_u = b(t{+}u)\big) d\ell^b_u(X^x) \\
&= C^A(t, x) + M_s + \int_0^s e^{-ru} \left(\text{Ł}_X G - rG\right)(t{+}u, X^x_u) I(X^x_u \geq b(t{+}u)) du
\end{aligned}
$$

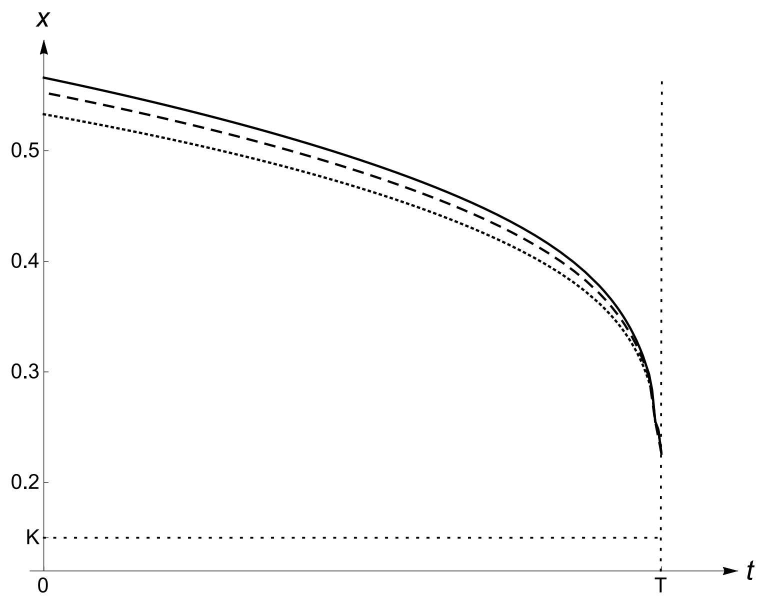

Figure 2. This figure plots the optimal exercise boundary b for models of $1/2$-type: $1/2$-model (solid), $1-1/(2\nu)$-model with $\nu = 0.8$ (dotted), and mixture $(1/2, 1-1/(2\nu))$-model with $\nu = 0.8$ and weights $w_j = 0.5$ (dashed). The parameter set is $T = 1$, $K = 0.15$, $r = 0.05$, and the coefficients for the $(1,2)$-model are $\alpha = 3$, $\beta = 0.68$, $\kappa = 1$. For the $1-1/(2\nu)$-model the coefficient $\alpha = 3.7$ is adjusted to have the same attractor as for the $1/2$-model; for the mixture model the coefficient α is 2.9.

$$= C^A(t,x) + M_s + \int_0^s e^{-ru} H(X_u^x) I(X_u^x \geq b(t+u))du$$

where we used (3.25), the smooth-fit condition (3.27), (3.29) and where $M = (M_s)_{s\geq 0}$ is the martingale part and $(\ell_t^b(X^x))_{t\geq 0}$ is the local time process of X^x at the boundary b

$$\ell_t^b(X^x) := Q - \lim_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon} \int_0^t I(b(t+u)-\varepsilon < X_u^x < b(t+u)+\varepsilon) d\langle X, X\rangle_u . \tag{4.7}$$

Now upon letting $s = T{-}t$, taking the expectation E, recalling the definition of C^E in (2.13), using the optional sampling theorem for M, rearranging terms and noting that $C^A(T,x) = G(x) = (x-K)^+$ for all $x > 0$, we get (4.4). The integral equation (4.5) is obtained by inserting $x = b(t)$ into (4.4) and using (3.26).

(B) Now we show that b is the unique solution to the equation (4.5) in the class of continuous functions $t \mapsto b(t)$. We note that monotonicity and the terminal value $b(T)$ are not needed for uniqueness. The proof is divided in several steps and is based on arguments similar to those employed by Du Toit and Peskir (2007) and originally derived by Peskir (2005b).

$(B.1)$ Let $c : [0,T] \to I\!R$ be a solution to equation (4.5) such that c is continuous with $c(T) = \max(K, x^*)$. We will show that c must be equal to the optimal exercise boundary b.

Now let us consider the function $U^c : [0,T) \to I\!R$ defined as follows

$$U^c(t,x) = C^E(t,x) + \int_0^{T-t} L(u,x,c(t+u))du \tag{4.8}$$

for $(t,x) \in [0,T] \times (0,\infty)$. Observe the fact that c solves the equation (4.5) means exactly that $U^c(t,c(t)) = G(c(t))$ for all $t \in [0,T]$. We will moreover show that $U^c(t,x) = G(x)$ for $x \in [c(t),\infty)$ with $t \in [0,T]$. This can be derived using the martingale property as follows: the Markov property of X implies that

$$e^{-rs}U^c(t+s,X_s^x) - \int_0^s e^{-ru}H(X_u^x)I(X_u^x \ge c(t+u))du = U^c(t,x) + N_s \tag{4.9}$$

where $(N_s)_{0\le s\le T-t}$ is a martingale under Q. On the other hand, we know from (3.10) that

$$e^{-rs}G(X_s^x) = G(x) + \int_0^s e^{-ru}H(X_u^x)I(X_u^x \ge K)du + M_s + \frac{1}{2}\int_0^s d\ell_u^K(X^x) \tag{4.10}$$

where $(M_s)_{0\le s\le T-t}$ is a continuous martingale under Q.

For $x \in [c(t),\infty)$ with $t \in [0,T]$ given and fixed, consider the stopping time

$$\sigma_c = \inf\ \{\ 0 \le s \le T-t : c(t+s) \ge X_s^x\ \} \tag{4.11}$$

under Q. Using that $U^c(t,c(t)) = G(c(t))$ for all $t \in [0,T]$ and $U^c(T,x) = G(x)$ for all $x > 0$, we see that $U^c(t+\sigma_c, X_{\sigma_c}^x) = G(X_{\sigma_c}^x)$. Hence from (4.9) and (4.10) using the optional sampling theorem we find:

$$\begin{aligned} U^c(t,x) &= \mathsf{E}e^{-r\sigma_c}U^c(t+\sigma_c,X_{\sigma_c}^x) - \mathsf{E}\int_0^{\sigma_c} e^{-ru}H(X_u^x)I(X_u^x \ge c(t+u))du \\ &= \mathsf{E}e^{-r\sigma_c}G(X_{\sigma_c}^x) - \mathsf{E}\int_0^{\sigma_c} e^{-ru}H(X_u^x)du = G(t,x) \end{aligned} \tag{4.12}$$

as $X_u^x \in (c(t+u),\infty)$ and $\ell_u^K(X^x) = 0$ for all $u \in [0,\sigma_c)$. This proves that $U^c(t,x) = G(x)$ for $x \in [c(t),\infty)$ with $t \in [0,T]$ as claimed.

$(B.2)$ We show that $U^c(t,x) \le C^A(t,x)$ for all $(t,x) \in [0,T] \times (0,\infty)$. For this consider the stopping time

$$\tau_c = \inf\ \{\ 0 \le s \le T-t : X_s^x \ge c(t+s)\ \} \tag{4.13}$$

under Q with $(t,x) \in [0,T] \times (0,\infty)$ given and fixed. The same arguments as those following (4.11) above show that $U^c(t+\tau_c, X_{\tau_c}^x) = G(X_{\tau_c}^x)$. Inserting τ_c instead of s in (4.9) and using the optional sampling theorem, we get

$$U^c(t,x) = \mathsf{E}e^{-r\tau_c}U^c(t+\tau_c,X_{\tau_c}^x) = \mathsf{E}e^{-r\tau_c}G(X_{\tau_c}^x) \le C^A(t,x) \tag{4.14}$$

proving the claim.

$(B.3)$ We show that $c \ge b$ on $[0,T]$. For this, suppose that there exists $t \in [0,T)$ such that $b(t) < c(t)$ and choose a point $x \in [c(t),\infty)$ and consider the stopping time

$$\sigma = \inf\ \{\ 0 \le s \le T-t : b(t+s) \ge X_s^x\ \} \tag{4.15}$$

under Q . Inserting σ instead of s in (4.6) and (4.9) and using the optional sampling theorem, we obtain

$$\mathsf{E}e^{-r\sigma}C^A(t+\sigma, X_\sigma^x) = C^A(t,x) + \mathsf{E}\int_0^\sigma e^{-ru}H(X_u^x)du \tag{4.16}$$

$$\mathsf{E}e^{-r\sigma}U^c(t+\sigma, X_\sigma^x) = U^c(t,x) + \mathsf{E}\int_0^\sigma e^{-ru}H(X_u^x)I\big(X_u^x \geq c(t+u)\big)du. \tag{4.17}$$

As $U^c \leq C^A$ and $C^A(t,x) = U^c(t,x) = G(x)$ for $x \in [c(t), \infty)$ with $t \in [0,T]$, it follows from (4.16) and (4.17) that

$$\mathsf{E}\int_0^\sigma e^{-ru}H(X_u^x)I\big(c(t+u) \geq X_u^x\big)du \geq 0. \tag{4.18}$$

Due to the fact that H is negative above $\max(K, x^*)$, we see by the continuity of b and c that (4.19) is not possible, so that we arrive at a contradiction. Hence, we can conclude that $b(t) \geq c(t)$ for all $t \in [0,T]$.

$(B.4)$ We show that c must be equal to b . For this, let us assume that there exists $t \in [0,T)$ such that $c(t) < b(t)$. Choose an arbitrary point $x \in (c(t), b(t))$ and consider the optimal stopping time τ^* from (2.11) under Q . Inserting τ^* instead of s in (4.6) and (4.9), and using the optional sampling theorem, gives

$$\mathsf{E}e^{-r\tau^*}G(X_{\tau^*}^x) = C^A(t,x) \tag{4.19}$$

$$\mathsf{E}e^{-r\tau^*}G(X_{\tau^*}^x) = U^c(t,x) + \mathsf{E}\int_0^{\tau^*} e^{-ru}H(X_u^x)I\big(X_u^x \geq c(t+u)\big)du \tag{4.20}$$

where we use that $C^A(t+\tau^*, X_{\tau^*}^x) = G(X_{\tau^*}^x) = U^c(t+\tau^*, X_{\tau^*}^x)$ upon recalling that $c \leq b$ and $U^c = G$ either above c or at T . As $U^c \leq C^A$ we have from (4.19) and (4.20) that

$$\mathsf{E}\int_0^{\tau^*} e^{-ru}H(X_u^x)I\big(X_u^x \geq c(t+u)\big)du \geq 0. \tag{4.21}$$

Due to the fact that H is negative above $\max(K, x^*)$, we see from (4.21) by continuity of b and c that such a point (t,x) cannot exist. Thus, c must be equal to b and the proof of the theorem is complete.

$\square$

Remark 4.2. *The integral equation* (4.5) *can be easily solved numerically via a backwards induction scheme based on a discretization of the integral with respect to time (for details see, e.g., Chapter 8 in Detemple (2006)). Note that in order to implement the algorithm, it is crucial to know the distribution of* Y_t *and the value of* $b(T) = \max(K, x^*)$ *. See Figures 1 and 2 for illustrations of the optimal exercise boundary* b *for models in Examples 2.1-2.6.*

Remark 4.3. *Numerical computations using the EEP formula* (4.4) *show that the American call price function* C^A *fails to be convex with respect to* x *under the* $3/2$ *-model at* $t = 0$ *(see Figure 3), unlike, e.g., in the geometric Brownian motion and the mean-reverting log process models (see Detemple and Osakwe (2000)). We note that the European call price function under the* $3/2$ *-model is not convex either, which was also pointed out by Goard and Mazur (2013). In contrast, Figure 4 shows that, for the chosen set of parameters, the American call price function is convex in* x *at* $t = 0$ *under the* $1/2$ *-model.*

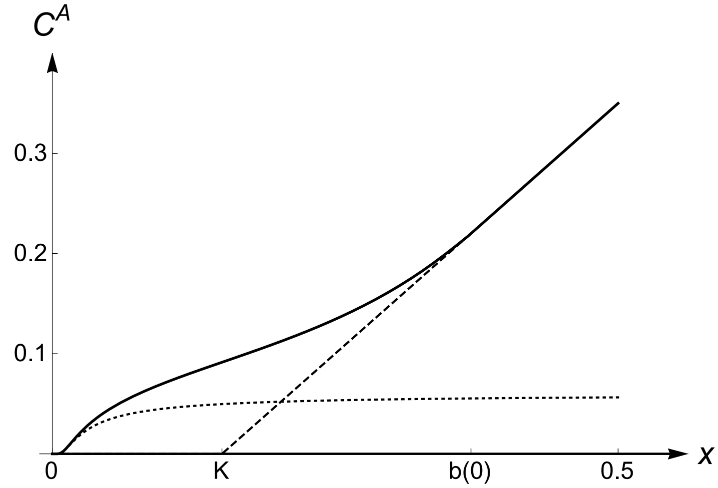

Figure 3. This figure plots the price functions of the American $C^A(0,x)$ (solid) and the European $C^E(0,x)$ (dotted) call prices for the $3/2$-model against x at $t=0$. The dashed line corresponds to the payoff function $(x-K)^+$. The graph shows that both functions are not convex with respect to x. The parameter set, as for Figure 1, is $T=1, K=0.15, r=0.05$, $\alpha=2.94, \beta=17.10, \kappa=2.05$.

5. The American VIX put option

In this section, we will briefly discuss the pricing problem for the American VIX put under model (2.3) with f of types $(A1)$ and $(A2)$:

$$P^A(t,x) = \sup_{0\le\tau\le T-t} \mathsf{E} e^{-r\tau}\widetilde{G}(X_\tau^x) \tag{5.1}$$

for $t\in[0,T)$ and $x>0$ where $\widetilde{G}=(K-x)^+$. The rational price of European VIX put is given by

$$P^E(t,x) = e^{-r(T-t)}\mathsf{E}(K-X_{T-t}^x)^+ \tag{5.2}$$

for $t\in[0,T)$ and $x>0$. The latter can be computed in the same way as the European call in Section 2. The methodology for the American put option is very similar to the one for the call option, thus we omit an analysis and only state the main result. As for the call option here we impose the **Assumption R** on the function h.

We define the function

$$\widetilde{L}(u,x,z) = -\mathsf{E}\big[e^{-ru}\widetilde{H}(X_u^x)I(X_u^x\le z)\big] \tag{5.3}$$

for $u\ge 0$ and $x,z>0$, which can be computed for f of 3/2 type as follows

$$\widetilde{L}(u,x,z) = -e^{-ru}\int_{g(z)}^\infty \widetilde{H}(f(\widetilde{y}))\,q(\widetilde{y};u,g(x))\,d\widetilde{y} \tag{5.4}$$

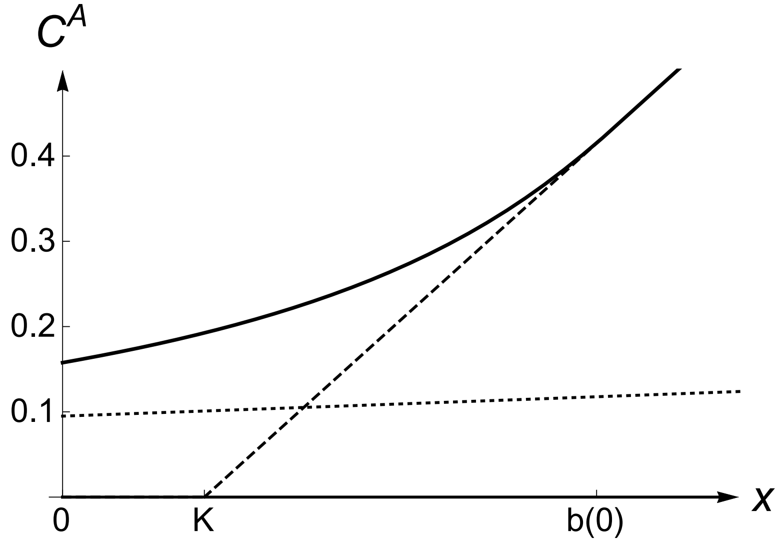

Figure 4. This figure plots the price functions of the American $C^A(0,x)$ (solid) and the European $C^E(0,x)$ (dotted) call prices for the $1/2$-model against x at $t=0$. The dashed line corresponds to the payoff function $(x-K)^+$. This graph shows that the American call price function is convex with respect to x at $t=0$ for the parameter set $T=1, K=0.15, r=0.05$, $\alpha=2.94, \beta=17.10, \kappa=2.05$.

and for f of $1/2$ type as

$$\widetilde{L}(u,x,z) = -e^{-ru}\int_0^{g(z)} \widetilde{H}(f(\widetilde{y}))\, q(\widetilde{y};u,g(x))\, d\widetilde{y}. \tag{5.5}$$

We now state the theorem on the rational price and optimal exercise boundary of the American VIX put. The proof is similar to the proof of Theorem 4.1.

Theorem 5.1. *The optimal exercise strategy in* (5.1) *is given by*

$$\widetilde{\tau} = \inf\ \{\ 0 \le s \le T-t : X_s^x \le \widetilde{b}(t+s)\ \} \tag{5.6}$$

where the optimal exercise boundary $\widetilde{b}$ *satisfies* $0 < \widetilde{b}(t) < \min(K, x^*)$ *for* $t \in [0,T)$ *and* $\widetilde{b}$ *is increasing on* $[0,T)$. *The price function* P^A *in* (5.1) *has the representation*

$$P^A(t,x) = P^E(t,x) + \int_0^{T-t} \widetilde{L}(u,x,b(t+u))du \tag{5.7}$$

for $t \in [0,T]$ *and* $x \in (0,\infty)$. *The exercise boundary* $\widetilde{b}$ *in* (5.1) *can be characterized as the unique solution to the nonlinear integral equation*

$$K - \widetilde{b}(t) = P^E(t,\widetilde{b}(t)) + \int_0^{T-t} \widetilde{L}(u,\widetilde{b}(t),\widetilde{b}(t+u))du \tag{5.8}$$

for $t \in [0,T]$ *in the class of continuous functions* $t \mapsto \widetilde{b}(t)$ *with* $\widetilde{b}(T) = \min(K, x^*)$.

6. Pricing the American VIX call under the generalized mixture model

In this section, we study the pricing of American VIX calls when the VIX is modelled as the sum of two processes: generalized 3/2- and 1/2-types. In other words, the process X is a function of a CIR process Y where this function is the sum of functions of $(A1)$ and $(A2)$ types. This can be seen as the generalization of the model introduced by Grasselli (2015), where the VIX is $a/\sqrt{Y}+b\sqrt{Y}$ and follows a $(1,0)$ - mixture model in our terminology. The process Y, which is assumed to be observed, will represent the underlying factor for the optimal stopping problem. We will show that, under certain assumptions, there exists a pair of optimal exercise boundaries that can be obtained as the unique solution to a system of coupled integral equations. The latter can be computed numerically by backward induction. We then provide the early exercise premium representation formula for the option price which decomposes it into the sum of a European part and an early exercise premium that depends on the pair of exercise boundaries.

6.1. The generalized mixture model

1. Consider a mean-reverting square-root process (Feller or CIR process) under a risk neutral measure Q,

$$dY_t = (\beta - \alpha Y_t)dt - \kappa\sqrt{Y_t}dB_t \tag{6.1}$$

for $t > 0$ where B is a standard Q-Brownian motion started at 0 and $\alpha, \beta, \kappa > 0$ are constant parameters such that $\beta \geq \kappa^2/2$ (Feller condition).

Now we take a function $f(y) := f_1(y) + f_2(y)$ where f_1 is of $A1$-type and f_2 is of $A2$-type and consider the VIX model

$$X_t = f(Y_t) \tag{6.2}$$

for $t > 0$. Defining the processes $X_{1t} = f_1(Y_t)$ (generalized 3/2-model) and $X_{2t} = f_2(Y_t)$ (generalized 1/2-model), then we obtain the alternative characterization of X

$$X_t = X_{1t} + X_{2t} \tag{6.3}$$

which means that X is the mixture of generalized 3/2-type and 1/2-type of models. Throughout the section, we will mostly use (6.2).

We adopt the standard assumption that the process Y is observed (e.g., Grasselli (2015)). Note that f converges to $+\infty$ as Y goes to 0 or $+\infty$. We also assume the following

Assumption M: There exists $y_{\min}$ such that f is strictly decreasing (increasing) for $y < y_{\min}$ $(y > y_{\min})$.

Remark 6.1. *By differentiating the function f one can see that **Assumption M** is equivalent to the following condition: $-f_2'(y)/f_1'(y) < 1$ if and only if $y < y_{\min}$ for some $y_{\min}$.*

Mixing the functions f_1 and f_2 from Examples 2.1-2.6, we can naturally consider the following models

Example 6.2. *((3/2, 1/2) -mixture model): Let $f(y) = a/y + by$ for positive constants a, b. Then $f'(y) = -a/y^2 + b, f''(y) = 2a/y^3$. The process exhibits reverting behavior if $\beta > \kappa^2$. It may have multiple attractors. The elasticity of variance is a non-linear function of the underlying factor.*

Example 6.3. *($(\nu+1/2, 1-1/(2\mu))$ -mixture model): Let $f(y) = a/y^\nu + by^\mu$ where $\nu > 0$, $\mu \in (0,1]$ and a, b are positive constants. Then $f'(y) = -a\nu/y^{\nu+1} + b\mu y^{\mu-1}, f''(y) = a\nu(\nu+1)/y^{\nu+2} + b\mu(\mu-1)y^{\mu-2}$. The $(3/2, 1/2)$ -mixture model is obtained when $\nu = \mu = 1$. The $(1, 0)$ -mixture model examined by Grasselli (2015) is obtained when $\nu = \mu = 1/2$.*

Example 6.4. *($(\nu_j+1/2, j = 1, ..., n, 1-1/(2\mu_i), i = 1, ..., m)$ -mixture model): Let $f(y) = \sum_j \omega_j/y^{\nu_j} + \sum_i \widehat{\omega}_i y^{\mu_i}$ where $\nu_j \geq 1, \omega_j > 0$ for $j = 1, ..., n$ and $\mu_i \in (0,1], \widehat{\omega}_i > 0$ for $i = 1, ..., n$ so that $f'(y) = -\sum_j \omega_j \nu_j/y^{\nu_j+1} + \sum_i \widehat{\omega}_i \mu_i y^{\mu_i-1}$ and $f''(y) = \sum_j \omega_j \nu_j(\nu_j+1)/y^{\nu_j+2} + \sum_i \widehat{\omega}_i \mu_i(\mu_i-1)y^{\mu_i-2}$.*

Examples 6.1-6.3 satisfy **Assumption M** as shown next for Example 6.3 which is the most general one. Indeed, it is enough to show that the derivative f' changes sign only once from negative to positive. Let us assume that $\mu_1 < \mu_i$ for $i = 2, ..., m$, then we rewrite f' as

$$f'(y) = \frac{-\sum_j \omega_j \nu_j/y^{\nu_j+\mu_1} + \widehat{\omega_1}\mu_1 + \sum_{i\geq 2} \widehat{\omega}_i \mu_i y^{\mu_i-\mu_1}}{y^{1-\mu_1}} \tag{6.4}$$

for $y > 0$ and note that the numerator is strictly increasing and varies from $-\infty$ to $+\infty$ and the denominator is strictly positive. Therefore f' changes sign a single time and the proof of the initial claim is complete.

2. Under the model (6.1)-(6.2) the rational price C^A of the American VIX call at time $t = 0$ is the value function of the following optimal stopping problem

$$C^A = \sup_{0\leq\tau\leq T} \mathsf{E} e^{-r\tau}(f(Y_\tau) - K)^+ \tag{6.5}$$

where the supremum is taken over all stopping times τ of Y and the expectation E is taken under a risk neutral measure Q.

As the process Y is time-homogeneous Markov, we will study the problem (6.5) in the Markovian setting and hence we introduce dependence on time t and the initial value of Y

$$C^A(t, y) = \sup_{0\leq\tau\leq T-t} \mathsf{E} e^{-r\tau} G(Y_\tau^y) \tag{6.6}$$

for $t \in [0, T)$ and $y > 0$ where Y^y represents the process Y started from $Y_0^y = y$ and the payoff function G is given by

$$G(y) := (f(y) - K)^+ \tag{6.7}$$

for $y > 0$.

3. The rational price function of the European VIX call is

$$C^E(t, y) = e^{-r(T-t)} \mathsf{E}(f(Y_{T-t}^y) - K)^+ \tag{6.8}$$

for $t \in [0,T)$ and $y > 0$. We note that given **Assumption M**, there are unique points $K_* \le K^*$ such that $f(y) \ge K$ when $y \le K_*$ or $y \ge K^*$. As the random variable Y_t^y has non-central chi-squared density function $q(\widetilde{y};t,y)$, one can compute C^E numerically using

$$C^E(t,y) = e^{-r(T-t)} \left[\int_0^{K_*} (f(\widetilde{y}) - K)\, q(\widetilde{y};u,y)\, d\widetilde{y} + \int_{K^*}^{\infty} (f(\widetilde{y}) - K)\, q(\widetilde{y};u,y)\, d\widetilde{y} \right] \tag{6.9}$$

for $t \in [0,T)$ and $y > 0$.

6.2. The free-boundary problem for the American VIX call

In this section, we reduce the problem (6.6) to a free-boundary problem which will be tackled using again the local time-space calculus (Peskir (2005a)). The continuity of G and standard arguments show that the continuation and exercise regions read

$$\mathcal{C} = \{\, (t,y) \in [0,T)\times[0,\infty) : C^A(t,y) > G(y) \,\} \tag{6.10}$$

$$\mathcal{E} = \{\, (t,y) \in [0,T)\times[0,\infty) : C^A(t,y) = G(y) \,\} \tag{6.11}$$

and the optimal stopping time in (6.6) is given by

$$\tau = \inf \{\ 0 \le s \le T{-}t : (t{+}s, Y_s^y) \in \mathcal{E}\ \}. \tag{6.12}$$

The process (6.1) also satisfies the conditions of Theorem 37 of Chapter V, Section 7 in Protter (1990) so that

$$\left[\mathsf{E} \sup_{0 \le u \le T} (Y_u^x {-} Y_u^y)^2 \right]^{1/2} \le C_L\, |x-y| \tag{6.13}$$

for $x, y > 0$ and some constant $C_L > 0$. We will use this estimate for the proof of the smooth-fit property.

1. First, we show that the price function C^A is continuous on $[0,T) \times (0,\infty)$. We have

$$\begin{aligned} 0 \le C^A(t,x) - C^A(t,y) &\le \sup_{0 \le \tau \le T-t} \mathsf{E} e^{-r\tau} \left(f(Y_\tau^x) {-} f(Y_\tau^y) \right)^+ \\ &\le \mathsf{E} \sup_{0 \le u \le T-t} \left(X_{1u}^{f_1(x)} {+} X_{2u}^{f_2(x)} {-} X_{1u}^{f_1(y)} {-} X_{2u}^{f_2(y)} \right)^+ \\ &\le \mathsf{E} \sup_{0 \le u \le T-t} \left(X_{1u}^{f_1(x)} {-} X_{1u}^{f_1(y)} \right)^+ + \mathsf{E} \sup_{0 \le u \le T-t} \left(X_{2u}^{f_2(x)} {-} X_{2u}^{f_2(y)} \right)^+ \end{aligned} \tag{6.14}$$

for $x \ge y$ and $t \in [0,T)$ where we used that $\sup(f) - \sup(g) \le \sup(f-g)$, $(x-K)^+ - (y-K)^+ \le (x-y)^+$ for $x, y, K \in I\!R$, and the representation (6.3). Using the continuity of f_1 and f_2 and the same arguments for processes X_1 and X_2 as in paragraph 1 of Section 3, shows that $y \mapsto C^A(t,y)$ is continuous uniformly over $t \in [0,T]$. The proof that $t \mapsto C^A(t,y)$ is continuous on $[0,T]$ for each $y \ge 0$ fixed is also analogous to the one in paragraph 1 of Section 3 and thus we omit it. Combining both facts establishes the continuity of C^A on $[0,T) \times (0,\infty)$.

2. Now we derive some initial insights into the structure of exercise region $\mathcal{E}$.

(i) We first calculate the function $H(y) := (\text{Ł}_Y G - rG)(y)$ for $y \in (0,\infty)$ (which is the instantaneous benefit of waiting to exercise) where

$$\text{Ł}_Y = (\beta - \alpha y)\,\frac{d}{dy} + \frac{\kappa^2 y}{2}\,\frac{d^2}{dy^2} \tag{6.15}$$

is the infinitesimal generator of Y . As $G(y) = (f(y) - K)^+$ we have that

$$H(y) = h(y) I(y \le K_* \text{ or } y \ge K^*) \tag{6.16}$$

for $y \in (0,\infty)$ where

$$h(y) = (\beta - \alpha y)\, f'(y) + \frac{\kappa^2 y}{2} f''(y) - r f(y) + rK \tag{6.17}$$

for $y > 0$. The following condition is imposed on the model

Assumption R': There exist $y_* < y^*$ such that $H(y) \ge 0$ if and only if $\min(y_*, K_*) \le y \le \max(y^*, K^*)$.

Numerical computations show that the models in Examples 6.1-6.3 satisfy this assumption for wide range of parameters.

(ii) We now use the Ito-Tanaka's formula and the definition of H to obtain

$$\begin{aligned} \mathsf{E} e^{-r\tau} G(Y_\tau^y) = G(y) + \mathsf{E} \int_0^\tau e^{-rs} H(Y_s^y) ds \\ + \frac{1}{2} \mathsf{E} \int_0^\tau e^{-rs} (-f'(K_*)) d\ell_s^{K_*}(Y^y) + \frac{1}{2} \mathsf{E} \int_0^\tau e^{-rs} f'(K^*) d\ell_s^{K^*}(Y^y) \end{aligned} \tag{6.18}$$

for $y \in (0,\infty)$ and any stopping time τ of the process Y where $(\ell_s^K(X))_{s \ge 0}$ is the local time process of X at levels $K \in \{K_*, K^*\}$

$$\ell_s^K(X^x) := Q - \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^s I(K - \varepsilon < X_u^x < K + \varepsilon)\, d\,\langle X, X \rangle_u \tag{6.19}$$

and $d\ell_s^K(X^x)$ refers to the integration with respect to the continuous increasing function $s \mapsto \ell_s^K(X^x)$.

Equation (6.18) and **Assumption R'** show that it is not optimal to exercise the call option when $\min(y_*, K_*) \le Y_t \le \max(y^*, K^*)$ as $H(Y_t) \ge 0$ in this region and thus both integral terms on the right-hand side of (6.18) are non-negative. This fact can be also explained in the particular case where $K_* \le Y_t \le K^*$ as follows: if the option holder exercises between K_* and K^* the payoff is null, however there is a positive probability of receiving a strictly positive payoff in future.

Another implication of (6.18) is that the exercise region is non-empty for all $t \in [0,T)$, as for small $y \downarrow 0$ and large $y \uparrow \infty$ the integrand H is negative and the local time terms are zero, and thus due to the insufficient time to compensate the negative H , it is optimal to stop at once.

3. Next we prove further properties of the exercise region $\mathcal{E}$ and define the optimal exercise boundaries.

(i) Using the same arguments as in Section 3 we can show that $\mathcal{E}$ is right-connected.

(ii) Now let us take $t>0$ and $x>y>\max(K^*,y^*)$ such that $(t,y)\in\mathcal{E}$. Then by right-connectedness of the exercise region we have that $(s,y)\in\mathcal{E}$ as well for any $s>t$. If we now run the process (t,Y_t) from (t,x) we cannot hit the level $\max(K^*,y^*)$ before exercise (as $x>y$), thus the local time terms in (6.18) are 0 and the integrand H is negative (by **Assumption R'**). Therefore, it is optimal to exercise at (t,x), which establishes up-connectedness of the exercise region $\mathcal{E}$ when $y>\max(K^*,y^*)$. Exploiting the same arguments, we show down-connectedness of the exercise region $\mathcal{E}$ when $y<\min(K_*,y_*)$

(iii) From (i) - (ii) and paragraph 2 (ii) above we can conclude that there exist a pair of optimal exercise boundaries $b_*:[0,T]\to(0,\infty)$ and $b^*:[0,T]\to(0,\infty)$ such that

$$\tau=\inf\ \{\ 0\le s\le T-t : Y_s^y\le b_*(t+s) \text{ or } Y_s^y\ge b^*(t+s)\ \} \tag{6.20}$$

is optimal in (6.6) and $0<b_*(t)<\min(K_*,y_*)<\max(K^*,y^*)<b^*(t)<\infty$ for $t\in[0,T)$. Moreover, b_* is increasing and b^* is decreasing on $[0,T)$.

4. Now we prove that the smooth-fit condition along the boundaries b_* and b^* holds

$$C_y^A(t,b_*(t)+)=C_y^A(t,b_*(t)-)=G'(b_*(t))=f'(b_*(t)) \tag{6.21}$$

$$C_y^A(t,b^*(t)-)=C_y^A(t,b^*(t)+)=G'(b^*(t))=f'(b^*(t)) \tag{6.22}$$

for all $t\in[0,T)$. We will only prove (6.22) below, as the proof for the lower boundary b_* is similar and can be omitted.

(i) First, let us fix a point $(t,y)\in[0,T)\times(0,\infty)$ lying on the boundary b^* so that $y=b^*(t)$. Then, we have

$$\frac{C^A(t,y)-C^A(t,y-\varepsilon)}{\varepsilon}\le\frac{G(y)-G(y-\varepsilon)}{\varepsilon} \tag{6.23}$$

and taking the limit as $\varepsilon\downarrow 0$, we get

$$\limsup_{\varepsilon\downarrow 0}\frac{C^A(t,y)-C^A(t,y-\varepsilon)}{\varepsilon}\le G'(y)=f'(y). \tag{6.24}$$

(ii) To prove the reverse inequality, we set $\tau_\varepsilon=\tau_\varepsilon(t,y-\varepsilon)$ as an optimal stopping time for $C^A(t,y-\varepsilon)$. Using that Y is a regular diffusion and $t\mapsto b^*(t)$ is decreasing, we have that $\tau_\varepsilon\to 0$ as $\varepsilon\to 0$ Q-a.s. By the comparison theorem for solutions of SDEs and noting that

$$\begin{aligned}(6.25)\quad & G(Y_{\tau_\varepsilon}^y)-G(Y_{\tau_\varepsilon}^{y-\varepsilon})\\ &=\big(f(Y_{\tau_\varepsilon}^y)-f(Y_{\tau_\varepsilon}^{y-\varepsilon})\big)\,I\left(f(Y_{\tau_\varepsilon}^{y-\varepsilon})\ge K\right)+(f(Y_{\tau_\varepsilon}^y)-K)I\left(f(Y_{\tau_\varepsilon}^y)\ge K\ge f(Y_{\tau_\varepsilon}^{y-\varepsilon})\right)\\ &\ge\big(f(Y_{\tau_\varepsilon}^y)-f(Y_{\tau_\varepsilon}^{y-\varepsilon})\big)\,I\left(f(Y_{\tau_\varepsilon}^{y-\varepsilon})\ge K\right)\end{aligned}$$

we obtain

$$\frac{1}{\varepsilon}\Big(C^A(t,y)-C^A(t,y-\varepsilon)\Big) \tag{6.26}$$

$$\geq \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(f(Y^y_{\tau_\varepsilon})-f(Y^{y-\varepsilon}_{\tau_\varepsilon})\right) I\left(f(Y^{y-\varepsilon}_{\tau_\varepsilon})\geq K\right)\right]$$
$$= \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(f(Y^y_{\tau_\varepsilon})-f(Y^{y-\varepsilon}_{\tau_\varepsilon})\right)\right] - \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(f(Y^y_{\tau_\varepsilon})-f(Y^{y-\varepsilon}_{\tau_\varepsilon})\right) I\left(f(Y^{y-\varepsilon}_{\tau_\varepsilon})\leq K\right)\right].$$

Then the second term on the right-hand side of (6.26) goes to 0 as $\varepsilon\to 0$ as

$$\begin{aligned}
0 &\leq \frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(f(Y^y_{\tau_\varepsilon})-f(Y^{y-\varepsilon}_{\tau_\varepsilon})\right) I\left(f(Y^{y-\varepsilon}_{\tau_\varepsilon})\leq K\right)\right]\\
&\leq \frac{1}{\varepsilon}\left(\mathsf{E}\left(f(Y^y_{\tau_\varepsilon})-f(Y^{y-\varepsilon}_{\tau_\varepsilon})\right)^2\right)^{1/2}\left(Q(f(Y^{y-\varepsilon}_{\tau_\varepsilon})\leq K)\right)^{1/2}\\
&= \frac{1}{\varepsilon}\left(\mathsf{E}\left(f'(\xi)(Y^y_{\tau_\varepsilon}-Y^{y-\varepsilon}_{\tau_\varepsilon})\right)^2\right)^{1/2}\left(Q(f(Y^{y-\varepsilon}_{\tau_\varepsilon})\leq K)\right)^{1/2}\\
&\leq \frac{1}{\varepsilon}C_{f'}\left(\mathsf{E}\sup_{0\leq u\leq T-t}\left(Y^y_u-Y^{y-\varepsilon}_u\right)^2\right)^{1/2}\left(Q(f(Y^{y-\varepsilon}_{\tau_\varepsilon})\leq K)\right)^{1/2}\\
&\leq C_{f'}C_L\left(Q(f(Y^{y-\varepsilon}_{\tau_\varepsilon})\leq K)\right)^{1/2}\to 0
\end{aligned}\tag{6.27}$$

where we used Holder inequality, the mean value theorem by taking $\xi\in[Y^{y-\varepsilon}_{\tau_\varepsilon}, Y^y_{\tau_\varepsilon}]$, the facts that $|f'(y)|\leq C_{f'}$ for some constant $C_{f'}>0$ and any $y\geq b_*(0)>0$, that $\xi\geq Y^{y-\varepsilon}_{\tau_\varepsilon}\geq b_*(t+\tau_\varepsilon)>b_*(0)$, the inequality (6.13) and that the latter probability goes to zero because $y>K^*$. Now, we turn to the first term on the right-hand side of (6.26). Using Ito's formula we have

$$\frac{1}{\varepsilon}\mathsf{E}\left[e^{-r\tau_\varepsilon}\left(f(Y^y_{\tau_\varepsilon})-f(Y^{y-\varepsilon}_{\tau_\varepsilon})\right)\right] = \frac{f(y)-f(y-\varepsilon)}{\varepsilon}+\frac{1}{\varepsilon}\mathsf{E}\left[\int_0^{\tau_\varepsilon} e^{-rs}\left(\omega(Y^y_s)-\omega(Y^{y-\varepsilon}_s)\right)ds\right]\tag{6.28}$$

where $\omega(y):=\big(\beta-\alpha y\big)f'(y)+\frac{1}{2}\kappa^2 y f''(y)-rf(y)$ for $y>0$. We show that the second term of (6.28) goes to 0 as $\varepsilon\to 0$

$$\begin{aligned}
0 &\leq \frac{1}{\varepsilon}\mathsf{E}\left|\int_0^{\tau_\varepsilon} e^{-rs}(\omega(Y^y_s)-\omega(Y^{y-\varepsilon}_s))ds\right| \leq \frac{1}{\varepsilon}\left[\mathsf{E}\int_0^{\tau_\varepsilon} e^{-rs}|\omega'(\xi_s)|(Y^y_s-Y^{y-\varepsilon}_s)ds\right]\\
&\leq \frac{1}{\varepsilon}C_{\omega'}\mathsf{E}\left[\tau_\varepsilon\sup_{0\leq u\leq T-t}\left(Y^y_u-Y^{y-\varepsilon}_u\right)\right] \leq \frac{1}{\varepsilon}C_{\omega'}\left(\mathsf{E}\tau_\varepsilon^2\right)^{1/2}\left(\mathsf{E}\sup_{0\leq u\leq T-t}\left(Y^y_u-Y^{y-\varepsilon}_u\right)^2\right)^{1/2}\\
&\leq \frac{1}{\varepsilon}C_{\omega'}C_L\,\varepsilon\left(\mathsf{E}\tau_\varepsilon^2\right)^{1/2} = C_{\omega'}C_L\left(\mathsf{E}\tau_\varepsilon^2\right)^{1/2}\to 0
\end{aligned}\tag{6.29}$$

where we used the mean value theorem and picking $\xi_s\in[Y^{y-\varepsilon}_s, Y^y_s]$, the facts that $|\omega'(y)|\leq C_{\omega'}$ for some $C_{\omega'}>0$ and all $y\geq b_*(0)>0$, that $\xi_s\geq Y^{y-\varepsilon}_s\geq b_*(t)>b_*(0)$ for $s\in[0,\tau_\varepsilon]$, Holder inequality, the inequality (6.13) and that $\mathsf{E}\tau_\varepsilon^2\to 0$ as $\varepsilon\to 0$ by the dominated convergence theorem.

Thus, using (6.26)-(6.29) and taking the limits as $\varepsilon\to 0$ we have that

$$\liminf_{\varepsilon\downarrow 0}\frac{C^A(t,y)-C^A(t,y-\varepsilon)}{\varepsilon}\geq G'(y)=f'(y)\tag{6.30}$$

for $t\in[0,T)$. Thus, combining (6.24) and (6.30) we obtain (6.22).

5. Using similar arguments as in paragraph 5 of Section 3, we can prove that the boundaries b_* and b^* are continuous on $[0,T]$ and that $b_*(T-) = \min(K_*, y_*)$ and $b^*(T-) = \max(K^*, y^*)$.

6. The facts proved in paragraphs 1-5 above and standard arguments based on the strong Markov property (see e.g. Peskir and Shiryaev (2006)) lead to the following free-boundary problem for the value function C^A and unknown boundaries b_* and b^* :

$$C^A_t + \mathbb{L}_Y C^A - rC^A = 0 \quad \text{in } \mathcal{C} \tag{6.31}$$

$$C^A(t, b_*(t)) = G(b_*(t)) = f(b_*(t)) - K \quad \text{for } t \in [0,T) \tag{6.32}$$

$$C^A(t, b^*(t)) = G(b^*(t)) = f(b^*(t)) - K \quad \text{for } t \in [0,T) \tag{6.33}$$

$$C^A_y(t, b_*(t)) = G'(b_*(t)) = f'(b_*(t)) \quad \text{for } t \in [0,T) \tag{6.34}$$

$$C^A_y(t, b^*(t)) = G'(b^*(t)) = f'(b^*(t)) \quad \text{for } t \in [0,T) \tag{6.35}$$

$$C^A(t,y) > G(y) \quad \text{in } \mathcal{C} \tag{6.36}$$

$$C^A(t,y) = G(y) \quad \text{in } \mathcal{E} \tag{6.37}$$

where the continuation set $\mathcal{C}$ and the exercise set $\mathcal{E}$ are given by

$$\mathcal{C} = \{\, (t,y) \in [0,T) \times (0,\infty) : b_*(t) < y < b^*(t) \,\} \tag{6.38}$$

$$\mathcal{E} = \{\, (t,y) \in [0,T) \times (0,\infty) : y \le b_*(t) \text{ or } y \ge b^*(t) \,\}. \tag{6.39}$$

The following properties of C^A , b_* and b^* were also verified above

$$C^A \text{ is continuous on } [0,T] \times (0,\infty) \tag{6.40}$$

$$C^A \text{ is } C^{1,2} \text{ on } \mathcal{C} \tag{6.41}$$

$$t \mapsto C^A(t,y) \text{ is decreasing on } [0,T] \text{ for each } y \in [0,\infty) \tag{6.42}$$

$$t \mapsto b_*(t) \text{ is increasing and continuous on } [0,T] \text{ with } b_*(T-) = \min(K_*, x_*) \tag{6.43}$$

$$t \mapsto b^*(t) \text{ is decreasing and continuous on } [0,T] \text{ with } b^*(T-) = \max(K^*, x^*). \tag{6.44}$$

7. We recall that we already showed how to compute the European VIX call price in Section 6.1 above. Now define the function

$$L(u, y, z_1, z_2) = -\mathsf{E}\big[e^{-ru} H(Y^y_u) I(Y^y_u \le z_1 \text{ or } Y^y_u \ge z_2)\big] \tag{6.45}$$

for $u \ge 0$ and $y, z_1, z_2 > 0$. Using that the random variable Y^y_t has non-central chi-squared density function $q(\widetilde{y}; t, y)$, we have

$$L(u, y, z_1, z_2) = -e^{-ru} \int_0^{z_1} H(\widetilde{y})\, q(\widetilde{y}; u, y)\, d\widetilde{y} - e^{-ru} \int_{z_2}^{\infty} H(\widetilde{y})\, q(\widetilde{y}; u, y)\, d\widetilde{y} \tag{6.46}$$

for $u \ge 0$ and $y, z_1, z_2 > 0$.

Theorem 6.5. *The price function* C^A *in* (6.6) *has the representation*

$$C^A(t,y) = C^E(t,y) + \int_0^{T-t} L(u, y, b_*(t+u), b^*(t+u))\, du \tag{6.47}$$

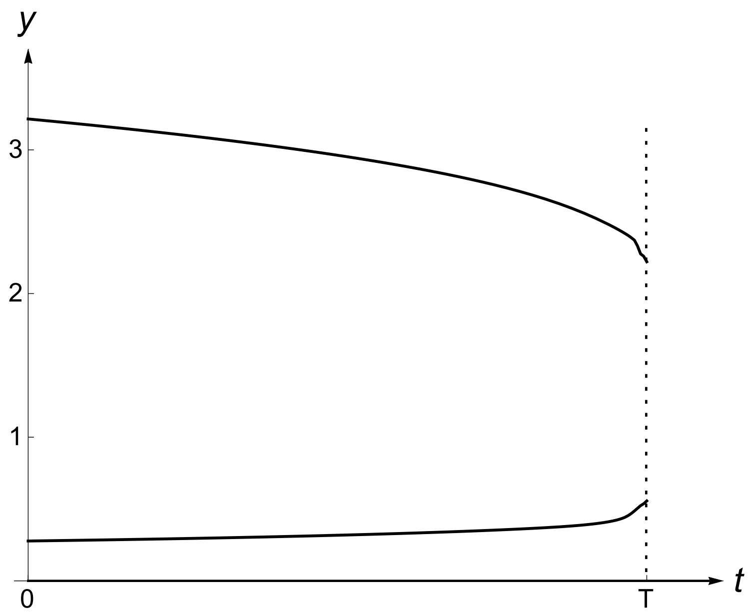

Figure 5. This figure plots the optimal exercise boundaries b_* (lower) and b^* (upper) for the process Y in the $(3/2, 1/2)$ -mixture model. The parameter set is $T = 1$ year, $\alpha = \kappa = 1, \beta = 2, r = 0.05, K = 0.15, a = b = 0.07$.

for $t \in [0, T]$ *and* $y \in (0, \infty)$ *. The optimal exercise boundaries* b_* *and* b^* *in* (6.6) *can be characterized as the unique solution to the coupled nonlinear integral equations of Volterra type*

$$f(b_*(t)) - K = C^E(t, b_*(t)) + \int_0^{T-t} L(u, b_*(t), b_*(t+u), b^*(t+u))du \tag{6.48}$$

$$f(b^*(t)) - K = C^E(t, b^*(t)) + \int_0^{T-t} L(u, b^*(t), b_*(t+u), b^*(t+u))du \tag{6.49}$$

for $t \in [0, T]$ *in the class of continuous functions* $b_*(t)$ *and* $b^*(t)$ *with* $b_*(T) = \min(K_*, y_*)$ *and* $b^*(T) = \max(K^*, y^*)$ *(See Figures 5 and 6).*

Proof. (A) First, we clearly have that the conditions for the local time-space formula on curves (Peskir (2005a)) hold (in the relaxed form) for $e^{-rs} C^A(t+s, Y_s^y)$ so that

$$\begin{aligned} e^{-rs} C^A(t+s, Y_s^y) & \\ = C^A(t, y) + M_s & \\ + \int_0^s e^{-ru} \left(C_t^A + \mathrm{L}_Y C^A - rC^A \right)(t+u, Y_u^y) I(X_u^x \neq \{b_*(t+u), b^*(t+u)\}) du & \\ + \frac{1}{2} \int_0^s e^{-ru} \left(C_y^A(t+u, Y_u^y+) - C_y^A(t+u, Y_u^y-) \right) I\big(Y_u^y = b_*(t+u)\big) d\ell_u^{b_*}(Y^y) & \\ + \frac{1}{2} \int_0^s e^{-ru} \left(C_y^A(t+u, Y_u^y+) - C_y^A(t+u, Y_u^y-) \right) I\big(Y_u^y = b^*(t+u)\big) d\ell_u^{b^*}(Y^y) & \end{aligned} \tag{6.50}$$

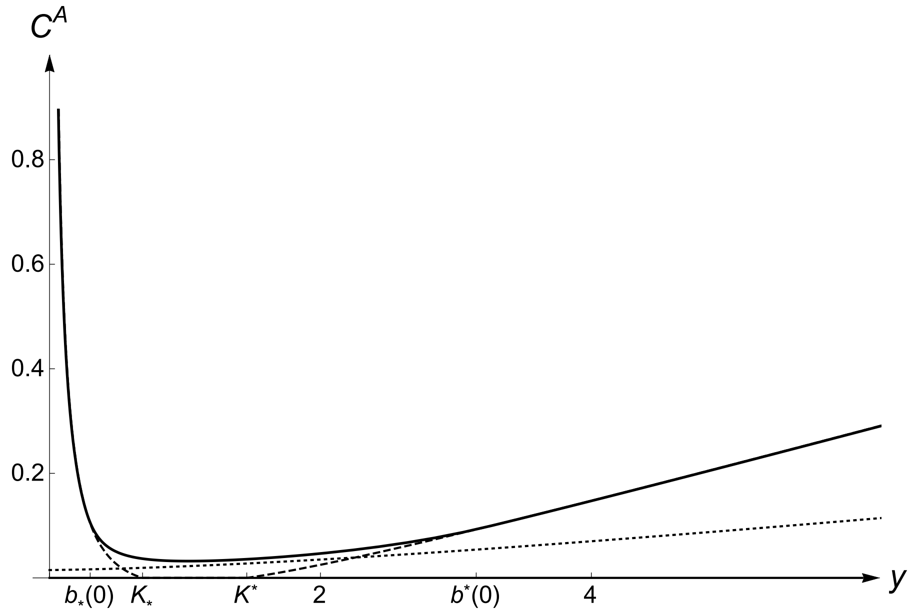

Figure 6. This figure plots the price functions of the American $C^A(0,y)$ (solid) and the European $C^E(0,y)$ (dotted) call price functions for the $(3/2,1/2)$ - mixture model against y at $t=0$. The dashed line corresponds to the pay-off function $(f(y)-K)^+$. The parameter set, as for Figure 5, is $T=1$ year, $\alpha=\kappa=1, \beta=2, r=0.05, K=0.15, a=b=0.07$. For this set of parameters, the figure shows the convexity of the American call price with respect to y.

$$
\begin{aligned}
&= C^A(t,y) + M_s \\
&\qquad + \int_0^s e^{-ru}\,(\mathrm{Ł}_Y G - rG)\,(t+u, Y_u^y) I(Y_u^y \le b_*(t+u) \,\text{or}\, Y_u^y \ge b^*(t+u))du \\
&= C^A(t,y) + M_s + \int_0^s e^{-ru} H(Y_u^y) I(Y_u^y \le b_*(t+u) \,\text{or}\, Y_u^y \ge b^*(t+u))du
\end{aligned}
$$

where we used (6.31) and the smooth-fit conditions (6.34)-(6.35), (6.37) and where $M = (M_s)_{s\ge0}$ is the martingale term, $(\ell_t^b(X^x))_{t\ge0}$ is the local time process of X^x at the boundaries $b \in \{b_*, b^*\}$:

$$
\ell_t^b(X^x) := Q - \lim_{\varepsilon\downarrow0} \frac{1}{2\varepsilon} \int_0^t I(b(t+u)-\varepsilon < X_u^x < b(t+u)+\varepsilon) d\,\langle X, X\rangle_u \,. \tag{6.51}
$$

Now upon letting $s=T-t$, taking the expectation E, recalling the definition of C^E in (6.9), using the optional sampling theorem for M, rearranging terms and noting that $C^A(T,y) = G(y) = (f(y)-K)^+$ for all $y>0$, we get (6.47). The system of integral equations (6.48)-(6.50) is obtained by substituting $x=b_*(t)$ and $x=b^*(t)$ into (6.47) and using (6.32) and (6.33), respectively.

(B) The proof of that the pair (b_*, b^*) is the unique solution to the system (6.48)-(6.50) in the class of continuous functions $t \mapsto b_*(t)$ and $t \mapsto b^*(t)$ follows from arguments similar to those employed in Theorem 3.1 in Section 3.

□

Remark 6.6. *The results of this section might be seen as the generalization of the results in Sections 2-4 if we slightly change the model and take* $f = f_1 + f_2$ *where* f_1 *is of* (A1) *-type or zero function, and* f_2 *is of* (A2) *-type or zero. Then if* $f_1 \equiv 0$ *(thus* f *is of 1/2-type), we have* $K_* = 0$ *,* $b_* = 0$ *and a single boundary* b^* *for* Y *which can be translated into the boundary* $f(b^*)$ *for the VIX process* X *. If now* $f_2 \equiv 0$ *(i.e.* f *is of 3/2-type), we have* $K^* = \infty$ *,* $b^* = \infty$ *and a single boundary* b_* *for* Y *which can be transformed into the boundary* $f(b_*)$ *for* X *.*

Appendix

Here we show that the models in Examples 2.1-2.6 satisfy **Assumption R** under some conditions for parameters when needed.

1. ($3/2$ -model) When $\beta > \kappa^2$ we get

$$h(x) = x(\alpha - r) - \left(\beta - \kappa^2\right) x^2 + rK$$

with $x^* = \frac{\alpha - r + \sqrt{(\alpha-r)^2 + 4(\beta-\kappa^2) rK}}{2(\beta-\kappa^2)} > 0.$

2. ($\nu+1/2$ -model) When $\beta > \frac{1}{2}\kappa^2(\nu+1)$ we obtain that

$$h(x) = \nu\left(\left(\alpha - \tfrac{r}{\nu}\right) x - \left(\beta - \tfrac{1}{2}\kappa^2(\nu+1)\right) x^{1+1/\nu}\right) + rK$$

is a strictly concave function for $x > 0$ with $h(+\infty) = -\infty$. The threshold x^* is the unique positive root of $\nu\left(\left(\alpha - \frac{r}{\nu}\right) x - \left(\beta - \frac{1}{2}\kappa^2(\nu+1)\right) x^{1+1/\nu}\right) + rK = 0$.

3. (mixture $\nu_j + 1/2$, $j = 1, ..., n$ model) We were not able to verify analytically the **Assumption R** for this model, however numerical results strongly support the claim that this assumption is satisfied when $\beta > \frac{1}{2}\kappa^2(\nu_j+1)$ for any $j = 1, ..., n$.

4. ($1/2$ -model) We have that

$$h(x) = \beta - \alpha x - r(x - K)$$

with $x^* = (\beta + rK)/(\alpha + r)$.

5. ($1 - 1/(2\nu)$ -model) When $\beta + \frac{1}{2}\kappa^2(\nu - 1) > 0$ (which is satisfied under Feller condition $\beta > \kappa^2/2$ we imposed throughout the paper) we have that

$$h(x) = \nu x^{1-1/\nu}\left(\beta + \tfrac{1}{2}\kappa^2(\nu - 1)\right) - (r + \nu\alpha) x + rK$$

is a strictly decreasing function for $x > 0$ with $h(+\infty) = -\infty$. The threshold x^* is the unique positive root of $\nu x^{1-1/\nu}\left(\beta + \frac{1}{2}\kappa^2(\nu - 1)\right) - (r + \nu\alpha) x + rK = 0$.

6. (mixture $1 - 1/(2\nu_j)$, $j = 1, ..., n$ model) When $\beta + \frac{1}{2}\kappa^2(\nu_j - 1) > 0$ for any $j = 1, ..., n$ (which also holds under Feller condition), we have that

$$h(x) = \sum_j \omega_j \nu_j g^{\nu_j - 1}(x)\left(\beta + \tfrac{1}{2}\kappa^2(\nu_j - 1)\right) - \alpha \sum_j \omega_j \nu_j g^{\nu_j}(x) - r(x - K)$$

is a strictly decreasing function for $x > 0$ with $h(+\infty) = -\infty$. The threshold x^* is the unique positive root of $\sum_j \omega_j \nu_j g^{\nu_j - 1}(x)\left(\beta + \frac{1}{2}\kappa^2(\nu_j - 1)\right) - \alpha \sum_j \omega_j \nu_j g^{\nu_j}(x) - r(x - K) = 0$.

References

[1] Broadie, M. *and* Jain, A. (2008). Pricing and Hedging Volatility Derivatives. *J. Derivat.* 15 (7–24).

[2] Cox, J., Ingersoll, J. *and* Ross, S. (1985), A Theory of the Term Structure of Interest Rates, *Econometrica* 53 (385–408).

[3] De Angelis, T. (2014). A Note on the Continuity of Free-Boundaries in Finite-Horizon Optimal Stopping Problems for One-Dimensional Diffusions. *SIAM J. Control Optim.* 53 1 (167–184).

[4] Detemple, J. *and* Osakwe, C. (2000). The Valuation of Volatility Option. *Eur. Finance Rev.* 4 (21–50).

[5] Detemple, J. (2006). *American-Style Derivatives.* Chapman & Hall/CRC.

[6] Drimus, G. (2011). Options on Realized Variance by Transform Methods: a Non-Affine Stochastic Volatility Model. *Quantitative Finance* (1–16).

[7] Du Toit, J. *and* Peskir, G. (2007). The Trap of Complacency in Predicting the Maximum. *Ann. Probab.* 35 (340–365).

[8] Elliott, R., Siu, T. *and* Chan, L. (2007). Pricing Volatility Swaps under Heston's Stochastic Volatility Model with Regime Switching. *Appl. Math. Finance* 14 (41–62).

[9] Gatheral, J. (2008). *Consistent Modeling of SPX and VIX options.* Presented at the Fifth World Congress of the Bachelier Finance Society, London. http://www.math.nyu.edu/fellows-fin-math/gatheral/Bachelier2008.pdf.

[10] Goard, J. *and* Mazur, M. (2013). Stochastic Volatility Models and the Pricing of VIX Options. *Math. Finance* 23 (439-458).

[11] Grasselli, M. (2015). The 4/2 Stochastic Volatility Model. To appear in *Math. Finance.*

[12] Grünbichler, A. *and* Longstaff, F. (1996). Valuing Futures and Options on Volatility. *J. Banking Finance* 20 (985–1001).

[13] Heston, S.L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Rev. Finan. Stud.* 6 (327-343).

[14] Heston, S.L. (1997). A Simple New Formula for Options with Stochastic Volatility. Working Paper, Washington University.

[15] Liu, H-K. (2015). Properties of American Volatility Options in the Mean-Reverting 3/2 Volatility Model. *SIAM J. Financial Math.* 1 (53–65).

[16] Mencia, J. *and* Sentana, E. (2013). Valuation of VIX derivatives. *J. Financ. Econ.* 108 (367–391).

[17] PESKIR, G. (2005a). A Change-of-Variable Formula with Local Time on Curves. *J. Theoret. Probab.* 18 (499–535).

[18] PESKIR, G. (2005b). On the American Option Problem. *Math. Finance* 15 (169–181).

[19] PESKIR, G. *and* SHIRYAEV, A. N. (2006). *Optimal Stopping and Free-Boundary Problems.* Lectures in Mathematics, ETH Zürich, Birkhäuser.

[20] PLATEN, E. (1997). A Non-Linear Stochastic Volatility Model. Working Paper, Australian National University.

[21] PROTTER, P. (1990). *Stochastic Integration and Differential Equations.* Springer-Verlag, Berlin.

[22] SEPP, A. (2008). Pricing Options on Realized Variance in Heston Model with Jumps in Returns and Volatility, *J. Computat. Finance* 11 (33–70).

[23] STEIN, E. M. *and* STEIN, J. C. (1991). Stock Price Distributions with Stochastic Volatility: an Analytical Approach. *Rev. Finan. Stud.* 4 (727–752).

[24] WHALEY, R. E. (1993). Derivatives on Market Volatility: Hedging Tools Long Overdue. *J. Derivat.* 1 (71–84).