Pricing with Bivariate Unspanned Stochastic Volatility Models

Joshua Wort

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

February 10, 2019

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy to the University of Cape Town. It has not before been submitted for any degree or examination.

Signed by candidate

Joshua Wort

February 10, 2019

Abstract

Unspanned stochastic volatility (USV) models have gained popularity in the literature. USV models contain at least one source of volatility-related risk that cannot be hedged with bonds, referred to as the unspanned volatility factor(s). Bivariate USV models are the simplest case, comprising of one state variable controlling the term structure and the other controlling unspanned volatility. This dissertation focuses on pricing with two particular bivariate USV models: the Log-Affine Double Quadratic (1,1) – or LADQ(1,1) – model of Backwell (2017) and the Linear-Rational Square Root (1,1) – or LRSQ(1,1) – model of Filipović *et al.* (2017). For the LADQ(1,1) model, we fully outline how an Alternating Directional Implicit finite difference scheme can be used to price options and implement the scheme to price caplets. For the LRSQ(1,1) model, we illustrate a semi-analytical Fourier-based method originally designed by Filipović *et al.* (2017) for pricing swaptions, but adjust it to price caplets. Using the above numerical methods, we calibrate each (1,1) model to both the British-pound yield curve and caps market. Although we cannot achieve a close fit to the implied volatility surface, we find that the parameters in the LADQ(1,1) model have direct control over the qualitative features of the volatility skew, unlike the parameters within the LRSQ(1,1) model.

Acknowledgements

Firstly, I would like to express my deep gratitude to my supervisor Dr Alex Backwell for his knowledgeable advice and guidance throughout.

To my parents, David and Michelle Wort, I thank you for the opportunities you have provided me and the constant support and encouragement along the way.

To my friends, thank you for the support. Thank you Emiel Zyde for running some MATLAB code for me.

Lastly, many thanks to FirstRand for the funding of my MPhil in Mathematical Finance degree.

Contents

List of Figures

Chapter 1

Introduction

An interest rate model describes the evolution of interest rates and their dependence on maturity. The dependence of interest rates on maturity is referred to as the term structure of interest rates. There are various different types of interest rates which are interconnected and representing the uncertainty around these future interest rates with a mathematical model can be extremely helpful in managing financial risk. Often there is only one underlying process modelled, from which the other quantities of interest can be determined (Buckova *et al.*, 2016). In this dissertation, we consider two approaches of modelling the term structure: specifying the short rate, originally introduced by Vasicek (1977), and directly specifying the state price density, which follows Constantinides (1992).

We introduce bivariate unspanned stochastic volatility models using the short rate process. The reason for this is that the short rate offers a natural characterisation of these models, as shown by Backwell (2017).

The short rate is the instantaneous spot rate. It is a continuous rate that drives the cash account (Brigo and Mercurio, 2007). A basic and fundamental class within general term structure models are one-factor short rate models. These are models where the short rate satisfies a stochastic differential equation (SDE) of general form

$$dr_t = \mu(t, r_t)dt + \sigma(t, r_t)dW_t,$$

where $\{W_t\}$ is a (standard, one-dimensional) Brownian motion, and where $\mu(t, r_t)$ and $\sigma(t, r_t)$ are the drift and volatility terms of the short rate, respectively, with some initial short rate r_0.

One-factor short rate models are useful to introduce and build intuitions about multi-factor models and often allow for useful comparisons; for instance, to highlight the contribution of additional factors. However, one-factor models do not take into account a fundamental stylised fact that interest rate volatility is stochastic in nature. One-factor short rate models can be given stochastic volatility by introducing an additional Brownian motion and a state variable that affects the volatility of

the short rate process. An example of such a model that exhibits stochastic volatility is the Fong and Vasicek (1991) model, where the short rate $\{r_t\}$ satisfies the following coupled dynamics:

$$dr_t = \kappa_r(\theta_r - r_t)dt + \sqrt{\nu_t}dW_t^1, \text{ and}$$
$$d\nu_t = \kappa_\nu(\theta_\nu - \nu_t)dt + \sigma_\nu\sqrt{\nu_t}dW_t^2,$$

where $\{W_t^1\}$ and $\{W_t^2\}$ are possibly correlated one-dimensional Brownian motions, and where κ_r, θ_r, κ_ν, θ_ν and σ_ν are constant parameters. This model exhibits stochastic volatility, due to the autonomous stochasticity feeding into the volatility process $\{\nu_t\}$ through the additional Brownian motion $\{W_t^2\}$.

The Fong and Vasicek (1991) model forms part of the affine model class, and like most affine models with stochastic volatility, this model predicts that it is able to extract spot rate volatility from bond prices alone (Collin-Dufresne *et al.*, 2009). This is because the volatility process will feature within the model's bond prices. A zero-coupon bond (ZCB) is defined as an asset which pays one unit at a future maturity time. In the Fong and Vasicek (1991) model, the time-t price of a ZCB maturing at time T, denoted by P_{tT}, is given by a function of the form

$$P_{tT} = P(t, T, r_t, \nu_t). \tag{1.1}$$

As we can see in Equation (1.1), one can gain exposure to the volatility process $\{\nu_t\}$ by buying a ZCB. As a result, sensitivity to the stochastic volatility state variable can be hedged with a bond portfolio. Using the terminology of Collin-Dufresne and Goldstein (2002), $\{\nu_t\}$ is said to be spanned.

However, Collin-Dufresne and Goldstein (2002) show evidence of *unspanned* sources of risk, i.e., there are sources of risk within fixed income derivatives that cannot be hedged with bonds. Through regression analysis Collin-Dufresne and Goldstein (2002) show that returns for cap straddles, which are particularly sensitive to volatility, are only slightly correlated with changes in bond prices. This intimates that an exposure to solely bonds will not enable one to hedge the return for cap straddles, and in turn volatility. Collin-Dufresne and Goldstein (2002) term this as unspanned stochastic volatility (USV). USV models comprise of at least one volatility-related state variable that is unspanned in nature, i.e., those state variables cannot be hedged with bonds and swaps (Backwell, 2017).

The literature contains several models that exhibit USV, though there are only two models involving one spanned and one unspanned volatility-related state variable. These models are named bivariate USV models and can also be referred to as (1,1) models. The two (1,1) models are: The Log-Affine Double Quadratic (1,1) – or LADQ(1,1) – model of Backwell (2017) and the LRSQ(1,1) model – the specific

linear-rational term structure model with a (1,1) structure outlined by Filipović *et al.* (2017). Note, we do not consider the Trolle and Schwartz (2008) model because, although the model can be specified with two Brownian motions (one driving interest rates and one driving unspanned volatility), it involves more than two state variables.

The focus of this dissertation is pricing with these (1,1) models. (1,1) models are the simplest, non-trivial USV models and are important for the same reasons that one factor short rate models are – we can compare (1,1) models to higher factor USV models in the same way one factor short rate models are compared to general term structure models (Backwell, 2017).

The dissertation begins, in Chapter 2, by firstly establishing the definition and the fundamental theorem of bivariate USV models from Backwell (2017). We then introduce the LADQ(1,1) model and LRSQ(1,1) model specification. The LRSQ(1,1) model follows a state price density approach but we show that we can transform it such that it is in terms of the short rate process. This enables us to better compare the (1,1) models.

In Chapter 3, we show how we can price derivatives in each of the (1,1) models. For the LADQ(1,1) model, we fully outline how a finite-difference scheme can be used to price options and implement the scheme to price caplets. For the LRSQ(1,1) model, we illustrate a semi-analytical Fourier-based method outlined by Filipović *et al.* (2017) and implement the method to price a caplet. This involves adjusting their method, which is designed for swaptions. Both numerical methods are validated by comparison to prices from Monte Carlo methods.

In Chapter 4, we use the above numerical methods to calibrate the model to cap prices written on British-pound LIBOR. Before each calibration, we pre-calibrate the model to the relevant yield curve. Finally, we consider the effect of the models' various parameters on the cap volatility surface.

Chapter 2

Bivariate USV models

Bivariate USV models – or (1,1) models – contain one spanned and one unspanned volatility-related state variables.

The (1,1) models[1] we discuss in this chapter are defined on a standard filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, where $\mathbb{P}$ is the real-world measure and the filration $\{\mathcal{F}_t\}$ is generated by a two-dimensional, independent Brownian motion $\{W_t^{\mathbb{P}}\}$. We also assume a risk-neutral measure $\mathbb{Q}$, equivalent to $\mathbb{P}$. The processes of the spanned and unspanned volatility-related state variables are defined for $t \in [0, S]$, where $S > 0$ is some finite time horizon. We assume the short rate $\{r_t\}$ and $\{u_t\}$ satisfy a pair of decoupled stochastic differential equations under $\mathbb{Q}$:

$$dr_t = \mu^r(t, r_t, u_t)\mathrm{dt} + \sigma^r(t, r_t, u_t)dW_t^{\mathbb{Q},1},$$
$$du_t = \mu^u(t, r_t, u_t)\mathrm{dt} + \sigma^u(t, r_t, u_t)dW_t^{\mathbb{Q},2},$$

where r_0 and u_0 are constants.

Following Backwell (2017), we define (1,1) models as any model where: i) $\sigma(t, r, u)$ is a non-trivial function of u, and ii) ZCB prices are given by the pricing function

$$P_{tT} = P(t, T, r_t).$$

A ZCB price is thus given by a pricing function which does not directly depend on the unspanned volatility-related state variable. Although $\{u_t\}$ drives the short-rate, and hence ZCB, volatility, it cannot be hedged with ZCBs, since its changes have no instantaneous effect on the yield curve.

Backwell (2017) provides a fundamental theorem of (1,1) models, which states that the following conditions are equivalent for a (1,1) model:

(A) The model has a (1,1) structure as defined by the above definition;

(B) The ZCB price is an affine function of the short rate:

$$P_{tT} = g(t, T) - f(t, T)r_t; \tag{2.1}$$

[1] We shall use the mathematical framework for (1,1) models proposed by Backwell (2017).

(C) The risk-neutral drift of the short rate is quadratic in the short rate:

$$\mu^r(t,r,u) = r^2 + r\alpha_1(t) + \alpha_2(t)$$

for deterministic functions $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$.

Note that the affine ZCB pricing function shown in condition (B) must not be confused with the standard exponential-affine functions found in affine term structure models. Condition (C) is helpful in enabling one to construct any (1,1) model.

Following Backwell (2017), the affine coefficient functions $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ in condition (B) can be further characterised as follows, provided the risk-neutral drift of the short rate admits two distinct real roots λ_1 and λ_2 (with $\lambda_1 < \lambda_2$),

$$f(t,T) = \frac{1}{\lambda_2 - \lambda_1}(\mathrm{e}^{-\lambda_1(T-t)} - \mathrm{e}^{-\lambda_2(T-t)}), \tag{2.2}$$

$$g(t,T) = \frac{1}{\lambda_2 - \lambda_1}(\lambda_2\mathrm{e}^{-\lambda_1(T-t)} - \lambda_1\mathrm{e}^{-\lambda_2(T-t)}), \tag{2.3}$$

for all $0 \leq t \leq T$.

2.1 The LADQ(1,1) model

In this section, we introduce the Log-Affine Double Quadratic – or LADQ(1,1) – model under the risk-neutral measure, $\mathbb{Q}$. This model is termed the Log-Affine Double Quadratic(1,1) model, because the model exhibits an affine bond price, and therefore log-affine yields, and due to the quadratic nature of both the short rate's drift and volatility terms.

The LADQ(1,1) model dynamics are specified as follows, under the risk-neutral measure $\mathbb{Q}$,

$$dr_t = (r_t - \lambda_1)(r_t - \lambda_2)dt + r_t(\bar{r} - r_t)u_t dW_t^{\mathbb{Q},1}, \tag{2.4}$$

$$du_t = \kappa(\theta - u_t)dt + \sigma\sqrt{u_t}dW_t^{\mathbb{Q},2}, \tag{2.5}$$

where $\{r_t\}$ is the short rate process which drives a cash account process $\{B_t\}$, with $B_t = \exp(\int_0^t r_s ds)$; $\{u_t\}$ is the unspanned volatility process, $\{W_t^{\mathbb{Q},1}\}$ and $\{W_t^{\mathbb{Q},2}\}$ are the two independent scalar components of Brownian motion $\{W_t^{\mathbb{Q}}\}$, and λ_1, λ_2, $\bar{r}$, κ, θ and σ are constant parameters.

The following parameter constraints ensure that Equation (2.4) admits a unique solution (Backwell, 2017):

$$0 < \lambda_1 < \bar{r} < \lambda_2, \tag{2.6}$$

$$0 < r_0 < \bar{r}. \tag{2.7}$$

These restrictions ensure that the volatility specification of the short rate, denoted by $\sigma^r(t,r,u) = r(\bar{r}-r)u$ bounds the short rate in the region $[0,\bar{r}]$. This is because if we ever have $r=0$, the short rate process becomes locally deterministic (as $\sigma^r(t,0,u)=0$) and has a positive drift of magnitude $\lambda_1\lambda_2$, hence the short rate is bounded from below by zero. And, if for some t, we have $r_t=\bar{r}$, the short-rate process has zero volatility (as $\sigma^r(t,\bar{r},u)=0$), but negative drift, hence the short rate has an upper bound $\bar{r}$. These arguments assume that the short rate process $\{r_t\}$ is continuous.

The short-rate volatility function also has the effect of giving the short rate a local-volatility structure, to accompany the unspanned volatility $\{u_t\}$. A volatility term which is expressed as a deterministic function of time and the state process is defined as exhibiting a local volatility structure. In our case, the state process is the short rate and the local volatility function, denoted $\sigma^{r,LV}(t,r,u)$, satisfies $\sigma^r(t,r,u)=\sigma^{r,LV}(t,r,u)u$, where $\sigma^{r,LV}(t,r,u)=r(\bar{r}-r)$.

If we ignore the unspanned volatility component by setting u_t to a constant, we see in the left-hand panel of Figure 2.1, the short rate's volatility is a negative quadratic in r. The local volatility is increasing in the region $[0,\frac{\bar{r}}{2}]$ and it would be an important model validation step to ensure $\bar{r}$ is set large enough, such that the short rate tends to remain in that region (Backwell, 2017). This is because the model would then satisfy the stylised fact that interest-rate volatility is positively level dependent, i.e., interest-rate volatility increases with rate levels (Piazzesi, 2010). The degree of the level dependence is also controlled by the value of $\bar{r}$: for a given short rate within the region $[0,\frac{\bar{r}}{2}]$, the larger the value of $\bar{r}$, the larger the level dependence[2] (see the right-hand panel of Figure 2.1).

The short rate process drift function $\mu(t,r,u)=\mu(r)=(r-\lambda_1)(r-\lambda_2)$ is a positive quadratic in r over the domain $[0,\bar{r}]$. The drift function is negative when $r>\lambda_1$, and positive when $r<\lambda_1$ (see the left hand panel of Figure 2.2). Therefore the drift function, in conjunction with the restriction in Equation (2.6) and the domain of the short rate, result in the short rate mean reverting to λ_1. The rate of mean reversion is controlled by λ_2; for a larger value of λ_2, the drift toward the mean λ_1 is larger in absolute value, as observed in the right-hand panel of Figure 2.2.

The unspanned volatility process has the simple effect of scaling the volatility of the short rate. The dynamics of $\{u_t\}$ in Equation (2.5) offer a standard mean-reverting interpretation, with θ controlling the level and κ controlling the rate of

[2] Note that the Brownian motions are uncorrelated. Although Backwell (2017) considers correlation, the level dependence from the local volatility function ensures that the model has an interaction between rates and volatility, without a need for a separate volatility parameter.

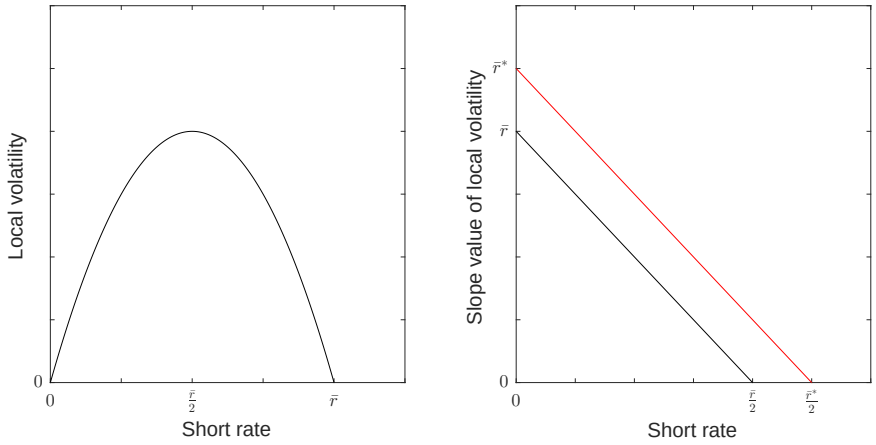


Fig. 2.1: In the left-hand panel, the local volatility term is plotted over the domain of the short rate. In the right-hand panel, the slope value of the local volatility term is plotted, with a larger value for $\bar{r}$ resulting in a larger slope value.

mean reversion. The values of these parameters will interact with the local volatility function, described above; for instance, a larger value of the long term mean θ, results in a local volatility term, which has smaller magnitude and is less level-dependent. Thus, the value of θ determines how the local volatility will manifest.

Recalling the price at time t of a T-dated ZCB given by Equation (2.1), the time-t continuously compounded yield of a T-dated ZCB is given by

$$y_{tT} = -\frac{\log P_{tT}}{T-t} = -\frac{1}{T-t}\log(g(t,T) - f(t,T)r_t),$$

where $f(t,T)$ and $g(t,T)$ are specified by Equations (2.2) and (2.3).

Backwell (2017) show that the long-term yields in the model are given as

$$\lim_{T\to\infty} y_{tT} = \lambda_1.$$

That is, long-term yields are given by the short rate's risk neutral long term mean (mean-reversion-level).

2.2 The LRSQ(1,1) model

In this section, we lay out the framework of linear-rational term structure models and introduce the Linear-Rational Square Root(1,1) – or LRSQ(1,1) – model. We also show the LRSQ(1,1) model specification in terms of the short rate, thereby allowing us to better understand the qualitative features of the parameters within the LRSQ(1,1) model.

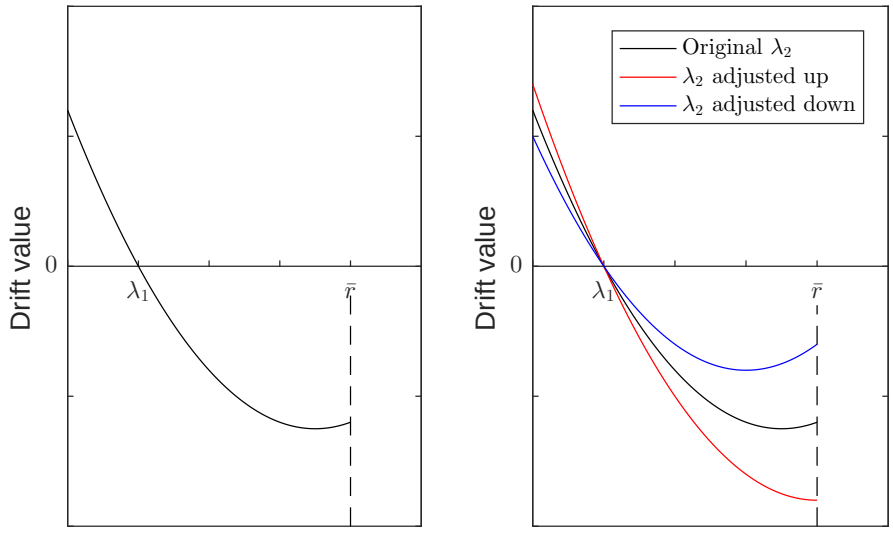


Fig. 2.2: In both panels, the risk-neutral short-rate drift term is plotted over the domain of the short rate. In the right-hand panel, the value for λ_2 is varied.

Linear-rational term structure models involve specifying a state price density, denoted $\{\zeta_t\}$; that is to specify a positive adapted process ζ_t such that the time-t price of any T-dated cash flow C_T is given by

$$\frac{1}{\zeta_t}\mathbb{E}^{\mathbb{P}}[\zeta_T C_T|\mathcal{F}_t], \tag{2.8}$$

where $\mathbb{E}^{\mathbb{P}}[\cdot|\mathcal{F}_t]$ denotes $\mathcal{F}_t$-conditional expectation under the real-world measure $\mathbb{P}$.

Specifically, a linear-rational term structure model consists of a multivariate factor process Z_t, which has linear drift, and a state price density ζ_t, which is a linear function of Z_t. The dynamics of Z_t are specified as

$$dZ_t = \kappa(\theta_Z - Z_t)dt + dM_t, \tag{2.9}$$

where $\kappa \in \mathbb{R}^{m\text{x}m}$, $\theta_Z \in \mathbb{R}^m$, and M_t is some m-dimensional martingale.

The state price density[3] is given by

$$\zeta_t = \mathrm{e}^{-\alpha t}(1 + \mathbf{1}^\top Z_t), \tag{2.10}$$

where $\mathbf{1}$ is a m-dimensional column vector of ones. Filipović *et al.* (2017) show that α is interpreted as the infinite-maturity ZCB yield, which makes it the analogue to λ_1 in the LADQ(1,1) model.

[3] Originally, Filipović *et al.* (2017) define the state price density of the form $\zeta_t = \mathrm{e}^{-\alpha t}(\phi + \psi^\top Z_t)$; however, without loss of generality, later specify $\phi = 1$ and $\psi = \mathbf{1}$. We rather work this parsimonious version of the state price density.

The linear drift of Z_t in Equation (2.9) implies that conditional expectations are of the following linear form:

$$\mathbb{E}^{\mathbb{P}}[Z_T|\mathcal{F}_t] = \theta_Z + \mathrm{e}^{-\kappa(T-t)}(Z_t - \theta_Z), \tag{2.11}$$

for $t \leq T$. Given this result and the specification of the state price density in Equation (2.10), ZCB prices and the short rate are linear-rational functions of Z_t; that is, they are ratios with the numerator and denominator each being linear in terms of Z_t. This is why these term structure models are referred to as linear-rational. By setting $C_T = 1$ in Equation (2.8), we obtain the price of a (default-free) ZCB as

$$P_{tT} = \mathrm{e}^{-\alpha(T-t)} \frac{1 + \mathbf{1}^\top \theta_Z + \mathbf{1}^\top \mathrm{e}^{-\kappa(T-t)}(Z_t - \theta_Z)}{1 + \mathbf{1}^\top Z_t}. \tag{2.12}$$

The short rate is obtained using the formula $r_t = -\frac{\partial}{\partial T} \log P_{tT}\big|_{T=t}$ and is shown by Filipović *et al.* (2017) to be given by

$$r_t = \alpha - \frac{\mathbf{1}^\top \kappa(\theta_Z - Z_t)}{1 + \mathbf{1}^\top Z_t}. \tag{2.13}$$

We shall now turn our attention to a specific example of a linear-rational term structure model: the Linear-Rational Square Root(1,1) – or LRSQ(1,1) – model. The LRSQ(1,1) model is based on a two-dimensional square-root diffusion process $\{X_t\}$, which takes values in $\mathbb{R}^2_+$, of the form

$$dX_t = (b - \beta X_t)dt + \mathrm{Diag}\left(\sigma_1 \sqrt{X_t^{(1)}}, \sigma_2 \sqrt{X_t^{(2)}}\right) dW_t^{\mathbb{P}}, \tag{2.14}$$

for some two-dimensional independent Brownian motion $\{W_t^{\mathbb{P}}\}$, and where $b \in \mathbb{R}^2$, $\beta \in \mathbb{R}^{2\mathrm{x}2}$, and volatility parameters, $\sigma_1, \sigma_2 > 0$.

The mean-reversion matrix β in Equation (2.14) is of the form

$$\beta = \begin{pmatrix} \kappa & 0 \\ 0 & \kappa \end{pmatrix}.$$

The constant drift term b in Equation (2.14) is of the form

$$b = \begin{pmatrix} \kappa(\theta_Z - \theta_U) \\ \kappa\theta_U \end{pmatrix},$$

where $\theta_Z \in \mathbb{R}$ and $\theta_U \in \mathbb{R}$.

Filipović *et al.* (2017) find it preferable to initially specify the state variables in the general square root form of Equation (2.14), and then apply a transform to isolate the unspanned component. The LRSQ(1,1) model consists of two factors in a

(1,1) structure. Though, it can be given in a general (m,n) specification (see Filipović *et al.* (2017) for the (m,n) specification including a general form of the transform).

We define $(Z_t, U_t) = SX_t$ as a linear transform of X_t, where $Z_t \in \mathbb{R}_+$ is defined as before but specified with a particular martingale, and is referred to as the term structure factor process; $U_t \in \mathbb{R}_+$ is the USV factor process, and S is a 2x2-matrix of the form

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Thus, the term structure factor and USV factor become $Z_t = X_t^{(1)} + X_t^{(2)}$ and $U_t = X_t^{(2)}$, respectively.

The implied dynamics of the joint factor process (Z_t, U_t) are

$$dZ_t = \kappa(\theta_Z - Z_t)dt + \sigma(Z_t, U_t)dW_t^{\mathbb{P}}, \tag{2.15}$$
$$dU_t = \kappa(\theta_U - U_t)dt + \sigma_2\sqrt{U_t}dW_t^{\mathbb{P},2}, \tag{2.16}$$

with dispersion function $\sigma(z, u) = (\sigma_1\sqrt{z-u}, \sigma_2\sqrt{u})$ and where κ, θ_Z, θ_U, σ_1, and σ_2 are constant parameters.

The restrictions on the LRSQ(1,1) model are:

$$\begin{aligned} \kappa(\theta_Z - \theta_U) &> 0, \\ \kappa\theta_U &> 0, \\ \sigma_1, \sigma_2 &> 0. \end{aligned}$$

In order to better compare the LRSQ(1,1) model with the LADQ(1,1) model, we can transform the LRSQ(1,1) model in terms of the short rate under the risk-neutral measure.

Filipović *et al.* (2017) show the market price of risk is given by $\lambda_t = -\frac{\sigma(Z_t,U_t)^\top}{1+Z_t}$. Using the Girsanov transformation, we change from the real-world measure $\mathbb{P}$ to the risk-neutral measure $\mathbb{Q}$, and get the following dynamics for the LRSQ(1,1) model:

$$\begin{aligned} dZ_t &= \left(\kappa(\theta_Z - Z_t) + \frac{\sigma(Z_t, U_t)\sigma(Z_t, U_t)^\top}{1 + Z_t}\right)dt + \sigma(Z_t, U_t)dW_t^{\mathbb{Q}}, \\ dU_t &= \left(\kappa(\theta_U - U_t) + \frac{\sigma_2^2 U_t}{1 + Z_t}\right)dt + \sigma_2\sqrt{U_t}dW_t^{\mathbb{Q},2}, \end{aligned} \tag{2.17}$$

where $\{W_t^{\mathbb{Q}}\}$ is a two-dimensional, independent Brownian motion under $\mathbb{Q}$.

Following Backwell (2017), we may now specify the LRSQ(1,1) model in terms of the short rate. The short rate is a function of Z_t given in Equation (2.13) and

thereby using Itô's lemma, we get the following dynamics for the short rate:

$$\begin{aligned} dr_t &= \frac{\kappa(1+\theta_Z)}{(1+Z_t)^2}dZ_t - \frac{\kappa(1+\theta_Z)}{(1+Z_t)^3}(dZ_t)^2 \\ &= \frac{\kappa(1+\theta_Z)}{(1+Z_t)^2}\left(\left(\kappa(\theta_Z - Z_t) + \frac{\sigma(Z_t,U_t)\sigma(Z_t,U_t)^\top}{1+Z_t}\right)dt + \sigma(Z_t,U_t)dW_t^{\mathbb{Q}}\right) \quad (2.18) \\ &\quad - \frac{\kappa(1+\theta_Z)}{(1+Z_t)^2}\sigma(Z_t,U_t)\sigma(Z_t,U_t)^\top dt \\ &= \frac{\kappa^2(\theta_Z - Z_t)(1+\theta_Z)}{(1+Z_t)^2}dt + \frac{\kappa(1+\theta_Z)}{(1+Z_t)^2}\sigma(Z_t,U_t)dW_t^{\mathbb{Q}} \\ &= \frac{\kappa^2(\theta_Z - Z_t)(1+\theta_Z)}{(1+Z_t)^2}dt + \frac{\kappa(1+\theta_Z)}{(1+Z_t)^2}\left[\sigma_1\sqrt{Z_t - U_t}, \sigma_2\sqrt{U_t}\right]dW_t^{\mathbb{Q}}. \quad (2.19) \end{aligned}$$

In Equation (2.18), the second-order Itô term cancels out exactly with the non-linear drift part in Equation (2.17), which is the Girsanov kernel multiplied by the volatility term of the term strucuture factor.

The USV factor $\{U_t\}$, which feeds into the volatility of $\{Z_t\}$, has risk-neutral dynamics of the form:

$$\begin{aligned} dU_t &= \kappa(\theta_U - U_t)dt + \sigma_2\sqrt{U_t}\left(dW_t^{\mathbb{Q},2} + \frac{\sigma_2\sqrt{U_t}}{1+Z_t}dt\right) \\ &= \left(\kappa(\theta_U - U_t) + \frac{\sigma_2^2 U_t}{1+Z_t}\right)dt + \sigma_2\sqrt{U_t}dW_t^{\mathbb{Q},2}. \quad (2.20) \end{aligned}$$

Equations (2.19) and (2.20) for the short rate and USV factor, respectively, can be simplified, using Equation (2.13), to

$$\begin{aligned} dr_t &= (r_t - \alpha)(r_t - \alpha - \kappa)dt \\ &\quad + \frac{(r_t - \alpha - \kappa)^2}{\kappa(1+\theta_Z)}\left[\sigma_1\sqrt{\frac{\alpha - \kappa\theta_Z - r_t}{r_t - \alpha - \kappa} - U_t}, \sigma_2\sqrt{U_t}\right]dW_t^{\mathbb{Q}}, \quad (2.21) \end{aligned}$$

and

$$dU_t = \left(\kappa(\theta_U - U_t) - \frac{\sigma_2^2 U_t(r_t - \alpha - \kappa)}{\kappa(1+\theta_Z)}\right)dt + \sigma_2\sqrt{U_t}dW_t^{\mathbb{Q},2}.$$

Equation (2.21) is driven by two uncorrelated Brownian motions. We can rather rewrite the above spanned and unspanned components in a decoupled fashion such that each factor is driven by only one Brownian motion. Following Backwell (2017), the dynamics of the LRSQ(1,1) model in terms of the short rate is specified

in the following form

$$dr_t = (r_t - \alpha)(r_t - \alpha - \kappa)dt + \frac{(r_t - \alpha - \kappa)^2}{\kappa(1+\theta_Z)}\sqrt{\frac{\sigma_1^2(\alpha - \kappa\theta_Z - r_t)}{r_t - \alpha - \kappa} + (\sigma_2^2 - \sigma_1^2)U_t}dW_t^{\mathbb{Q},1}, \tag{2.22}$$

and

$$dU_t = \left(\kappa(\theta_U - U_t) - \frac{\sigma_2^2 U_t(r_t - \alpha - \kappa)}{\kappa(1+\theta_Z)}\right)dt + \sigma_2\sqrt{U_t}d\hat{W}_t^{\mathbb{Q},2},$$

with correlation, ρ_t, between the Brownian motions $\hat{W}_t^{\mathbb{Q},1}$ and $\hat{W}_t^{\mathbb{Q},2}$ given as

$$\rho_t = \frac{\sigma_2\sqrt{U_t}}{\sqrt{\frac{\sigma_1^2(\alpha-\kappa\theta_Z-r_t)}{r_t-\alpha-\kappa} + (\sigma_2^2 - \sigma_1^2)U_t}}. \tag{2.23}$$

The LRSQ(1,1) short rate is endogenously bounded. Its lower bound follows from the process $\{Z_t\}$ being non-negative. As a result, if we let $(Z_t = 0)$ in Equation (2.13), we see that the short rate is bounded below by $\alpha - \kappa\theta_Z$. The upper bound of the short rate is given by taking the supremum of the short rate given in Equation (2.13) over the widest possible range $(-1, \infty)$ such that the state price density is positive. That is,

$$\sup_{z\in(-1,\infty)} \alpha - \frac{\kappa(\theta_Z - z)}{1+z} = \lim_{z\to\infty} \alpha - \frac{\kappa(\theta_Z - z)}{1+z} = \alpha + \kappa.$$

Note $\alpha + \kappa$ is analogous to λ_2 in the LADQ(1,1) model. Unlike the LADQ(1,1) model, the LRSQ(1,1) model endogenously has an upper bound for the short rate, without specifying an ad-hoc restriction.

As expected from Condition (C) of the fundamental theorem of (1,1) models, the risk-neutral short-rate drift term in Equation (2.22) is a quadratic in the short rate. From the drift function we can see that α and $\alpha + \kappa$ are analogous to λ_1 and λ_2 in the LADQ(1,1) model, respectively. Comparing to the LADQ(1,1) case in Section 2.1, the short rate mean reverts to α and the rate of mean reversion is controlled by $\alpha + \kappa$.

Chapter 3

Option Pricing

In this chapter, we consider the pricing of interest-rate derivatives, such as caplets, with the (1,1) models introduced in Chapter 2. For the LRSQ(1,1) model, we consider the Fourier pricing method of Filipović *et al.* (2017). The Fourier method cannot be applied to the LADQ(1,1) model, as the dynamics do not admit a characteristic function, and so a two-dimensional finite difference scheme, following Backwell (2017), will be developed. Monte-Carlo pricing would be too inefficient in general, but is also implemented to validate the Fourier and finite-difference methods.

3.1 Implementing the ADI scheme for the LADQ(1,1) model

Finite difference methods enable one to simply and accurately find the numerical solution for partial differential equations (PDEs) by approximating the partial derivatives with finite differences and discretising the problem space. Finite difference methods thereby firstly require a PDE, paired with initial and boundary conditions.

A term structure PDE governs the dynamics of option values depending on interest rates. We initially consider a European put option on a ZCB (which will enable the handling of the interest rate caps we consider in Chapter 4).

The payoff at maturity of a ZCB put option is given by $\Phi(r)$, and the no arbitrage time-t price of such a claim is a function of time and the state variables and denoted by $F(t, r_t, u_t)$. The value of this option must offer an expected rate of return equal to the short rate, under the risk-neutral measure. Equating the two and reversing time shall produce the desired PDE (Backwell, 2017).

The expected rate of return can be computed using Itô's lemma and the LADQ(1,1) model dynamics given by Equations (2.4) and (2.5). A simple application of Itô's

lemma produces the following dynamics of the ZCB put option

$$dF(t,r_t,u_t) = [F_t + (r_t-\lambda_1)(r_t-\lambda_2)F_r + \kappa(\theta-u_t)F_u + \frac{1}{2}(r_t(\bar{r}-r_t)u_t)^2F_{rr} + \frac{1}{2}\sigma^2 u_t F_{uu}]dt + r_t(\bar{r}-r_t)u_t F_r dW_t^{\mathbb{Q},1} + \sigma\sqrt{u_t}dW_t^{\mathbb{Q},2},$$

where partial derivatives are denoted in shorthand (for example, $F_r = \frac{\partial F}{\partial r}$).

From standard option pricing theory, the risk neutral drift of the option price process must equal the short rate multiplied by the option price, and by reversing time by setting $\tau = T - t$, the following PDE is produced:

$$F_\tau - (r-\lambda_1)(r-\lambda_2)F_r - \kappa(\theta-u)F_u - \frac{1}{2}(r(\bar{r}-r)u)^2F_{rr} - \frac{1}{2}\sigma^2 uF_{uu} + rF = 0. \quad (3.1)$$

Since the PDE is time reversed, the terminal payoff of the ZCB put at maturity is equivalent to the initial condition of the PDE. The closed form bond price apparent in (1,1) models enables us to explicitly represent a ZCB put as a T-dated claim that depends only on the short rate. Consider, for example, a put option written on a S-year ZCB, with expiry in T-year's time (where $T < S$), which pays out $(K - P_{TS})^+ = (K - g(T,S) + f(T,S)r_T)^+ = \Phi(r_T)$ at time T. As a result the initial condition, denoted by $\Phi(r_0)$, is in terms of the initial short rate.

The above initial boundary value problem (3.1) has two spatial domains and can be solved using the Alternating Directional Implicit (ADI) scheme, which was originally outlined by Peaceman and Rachford (1955). The ADI scheme has a number of advantages over alternative finite difference schemes. First, explicit difference methods, which explicitly compute the values at time step $m+1$ using the values at time step m, have poor stability problems. While implicit difference methods, which implicitly compute the values at time step m given the values at time step $m+1$ (i.e., give the values as a solution to a set of simultaneous equations), have superior stability properties, they are difficult to implement in two or more dimensions. The ADI scheme is a better alternative as it improves the stability of the explicit scheme while remaining computationally tractable, through introducing an intermediary half time-step. An added benefit is that the ADI scheme typically simplifies the discretised PDE, for each half-time step, to a tridiagonal system of equations (Duffy, 2013). This can be computed efficiently with the Thomas algorithm (Higham, 2002).

In the first half time-step of the scheme, the PDE is solved implicitly in the first spatial direction and explicitly in the second spatial direction, and vice-versa for the next half time-step. Duffy (2013) proves that if there are no mixed derivative terms in the PDE with two spatial domains, the ADI scheme is unconditionally stable, which is the case for the above PDE.

To employ the ADI scheme to the above initial boundary value problem, the time and state-variable domains must first be discretised. We divide the short rate axis into equally spaced nodes a distance δ_r apart, and do the same for both the unspanned volatility axis and time axis, using a spacing of δ_u and δ_τ, respectively.

We must work within a finite state space, hence it is necessary to truncate the unspanned volatility domain with an upper bound $\bar{u}$. We are only concerned with the values of $F(\tau, r, u)$ at the mesh points $(m\delta_\tau, i\delta_r, j\delta_u)$ for $0 \leq m \leq \frac{T}{\delta_\tau}$, $0 \leq i \leq \frac{\bar{r}}{\delta_r}$, and $0 \leq j \leq \frac{\bar{u}}{\delta_u}$. As a convenience we use the shorthand notation $F_m^{i,j} = F(m\delta_\tau, i\delta_r, j\delta_u)$.

Finally, the boundary conditions need to be specified. The boundary conditions here describe the behavior of the problem along the edges of the domain $[0, \bar{r}]$ x $[0, \bar{u}]$.

In the r-direction, we can use Neumann boundary conditions, instead of the typical Dirichlet boundary conditions which specify the value at the boundary. Neumann boundary conditions, on the other hand, specify the value of the derivative along the boundary. In the r-direction, the slope of the pricing surface of the option tends to the slope of the terminal payoff surface, as $r \to 0$ and $r \to \infty$. When $r = 0$, the put option is far out-the-money (as low interest rates imply high bond prices), and so we expect $\frac{\partial \Phi}{\partial r} = 0$. When $r = \bar{r}$, the ZCB put option will be deep in-the-money and we can ignore the 'positive part' of the terminal payoff function. By expanding the payoff function and taking the partial derivative with respect to r, we get the following

$$\begin{aligned}\frac{\partial \Phi}{\partial r} &= \frac{\partial}{\partial r}\left(K - g(T,S) + f(T,S)r\right) \\ &= f(T,S).\end{aligned}$$

Therefore, the Neumann boundary conditions in the r-direction are given by

$$\frac{\partial}{\partial r}F(\tau, 0, u) = 0 \quad \text{and} \quad \frac{\partial}{\partial r}F(\tau, \bar{r}, u) = f(T,S).$$

These boundary conditions can be numerically approximated using a forward difference and backward difference approximation, respectively. A forward / backward difference approximation uses the nearest point(s) after / before the given point at which the derivative is to be found.

We define a column vector for the option price as

$$F_m^{i,\cdot} = \begin{bmatrix} F_m^{i,1} \\ \vdots \\ F_m^{i,\frac{\bar{u}}{\delta_u}-1} \end{bmatrix},$$

for all $0 \leq i \leq \frac{\bar{r}}{\delta_r}$ and, similarly, we define a row vector for the option price as

$$F_m^{\cdot,j} = \left[F_m^{1,j}, \cdots, F_m^{\frac{\bar{r}}{\delta_r}-1,j} \right],$$

for all $0 \leq j \leq \frac{\bar{u}}{\delta_u}$.

The above r-boundary conditions are specified in column vector form as

$$\begin{aligned} F_{m+\frac{1}{2}}^{0,\cdot} &= F_{m+\frac{1}{2}}^{1,\cdot}, \\ F_{m+\frac{1}{2}}^{\frac{\bar{r}}{\delta_r},\cdot} &= F_{m+\frac{1}{2}}^{\frac{\bar{r}}{\delta_r}-1,\cdot} + \delta_r f(T,S). \end{aligned}$$

In the u-direction, we cannot use the same reasoning as for the r-direction, since the unspanned volatility is independent of the terminal payoff surface. A Neumann boundary condition can still be approximated by the use of extrapolation. The value of the derivative along the $u = 0$ and $u = \bar{u}$ boundaries can be approximated by the slope of the pricing surface between the two nearest interior points which precede the boundary point, i.e., we assume a 'continued slope'.

The u-boundary row vectors are specified as

$$\begin{aligned} F_m^{\cdot,0} &= 2F_m^{\cdot,1} - F_m^{\cdot,2}, \\ F_m^{\cdot,\frac{\bar{u}}{\delta_u}} &= 2F_m^{\cdot,\frac{\bar{u}}{\delta_u}-1} - F_m^{\cdot,\frac{\bar{u}}{\delta_u}-2}. \end{aligned}$$

We may now numerically approximate the PDE (3.1). Between time m and $m + \frac{1}{2}$, we solve the system implicitly in the r-direction and explicitly in the u-direction. Using a forward-difference approximation for the time derivative and central-difference approximation[1] for the spatial derivatives, we obtain the following expression:

$$\begin{aligned} &\frac{F_{m+\frac{1}{2}}^{i,j} - F_m^{i,j}}{\delta_\tau/2} - (i\delta_r - \lambda_1)(i\delta_r - \lambda_2)\frac{F_{m+\frac{1}{2}}^{i+1,j} - F_{m+\frac{1}{2}}^{i-1,j}}{2\delta_r} - \kappa(\theta - j\delta_u)\frac{F_m^{i,j+1} - F_m^{i,j-1}}{2\delta_u} \\ &- \frac{1}{2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2\frac{F_{m+\frac{1}{2}}^{i+1,j} - 2F_{m+\frac{1}{2}}^{i,j} + F_{m+\frac{1}{2}}^{i-1,j}}{\delta_r^2} - \frac{1}{2}\sigma^2(j\delta_u)\frac{F_m^{i,j+1} - 2F_m^{i,j} + F_m^{i,j-1}}{\delta_u^2} \\ &+ i\delta_r F_m^{i,j} = 0, \end{aligned}$$

[1] A central difference approximation uses the nearest point(s) before and after the given point at which the derivative is to be found. Note, some central difference approximations, for example, the second order one, include the value at the given point for the approximation.

which can be rewritten as

$$\begin{aligned}
&F^{i,j}_{m+\frac{1}{2}} - \frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}) \\
&- \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}) = \\
&\kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F^{i,j+1}_m - F^{i,j-1}_m) + \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(F^{i,j+1}_m - 2F^{i,j}_m + F^{i,j-1}_m) \\
&+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F^{i,j}_m.
\end{aligned} \tag{3.2}$$

The above system (3.2) may be rewritten in vector/matrix form[2] as

$$\begin{aligned}
F^{\cdot,j}_{m+\frac{1}{2}} = (\mathbf{G}^j_{m,1})^{-1}[&\gamma_3\kappa(\theta - j\delta_u)(F^{\cdot,j+1}_m - F^{\cdot,j}_m) \\
&+ \sigma^2\gamma_4(j\delta_u)(F^{\cdot,j+1}_m - 2F^{\cdot,j}_m + F^{\cdot,j-1}_m) \\
&+ (\mathbf{1} - \mathbf{H}_1\frac{\delta_r\delta_\tau}{2}) \cdot F^{\cdot,j}_m + \mathbf{b}^j_m],
\end{aligned}$$

for $0 < j < \frac{\bar{u}}{\delta_u}$, where

$$\mathbf{G}^j_{m,1} = \mathbf{I} - \gamma_1\mathbf{D}_1\mathbf{D}_2\mathbf{S}_1 - \gamma_2\mathbf{D}_3^2\mathbf{D}_4^2\mathbf{T}_1(j\delta_u)^2. \tag{3.3}$$

In the above Equation (3.3), we have that $\mathbf{I}$, $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_3$, $\mathbf{D}_4$, $\mathbf{S}_1$and $\mathbf{T}_1$ are all matrices of size $(\bar{r}/\delta_r - 1)$ x $(\bar{r}/\delta_r - 1)$. $\mathbf{I}$ is the identity matrix,

$$\mathbf{D}_2 = \begin{bmatrix} \delta_r & 0 & \cdots & 0 \\ 0 & 2\delta_r & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \bar{r} - \delta_r \end{bmatrix},$$

$$\begin{aligned}
\mathbf{D}_1 &= \mathbf{D}_2 - \lambda_1\mathbf{I}, \\
\mathbf{D}_3 &= \mathbf{D}_2 - \lambda_2\mathbf{I}, \\
\mathbf{D}_4 &= \bar{r}\mathbf{I} - \mathbf{D}_2,
\end{aligned}$$

$$\mathbf{S}_1 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{T}_1 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -2 & 1 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}.$$

[2] The proof of this derivation for the system in the first half time step is given in Appendix A.1.

$\mathbf{1}$, $\mathbf{H}_1$ and $\mathbf{b}_m^j$ are all vectors of size $(\bar{r}/\delta_r - 1)$ x 1, where $\mathbf{1}$ is a vector of ones,

$$\mathbf{H}_1 = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ \frac{\bar{r}}{\delta_r} - 1 \end{bmatrix}, \quad \mathbf{b}_m^j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b_1 + b_2(j\delta_u)^2 \end{bmatrix},$$

with b_1 and b_2 defined as follows:

$$b_1 = \gamma_1(\bar{r} - \delta_r - \lambda_1)(\bar{r} - \delta_r - \lambda_2)\delta_r f(T,S),$$
$$b_2 = \gamma_2(\bar{r} - \delta_r)^2\delta_r^3 f(T,S),$$

and where $\gamma_1 = \frac{\delta_\tau}{4\delta_r}$, $\gamma_2 = \frac{\delta_\tau}{4\delta_r^2}$, $\gamma_3 = \frac{\delta_\tau}{4\delta_u}$ and $\gamma_4 = \frac{\delta_\tau}{4\delta_u^2}$.

$\mathbf{G}_{m,1}^j$ is a tridiagonal matrix, which allows its inverse to be efficiently computed using the Thomas algorithm (Higham, 2002).

Then, between time $m + \frac{1}{2}$ and $m + 1$, we solve the system (3.1) explicitly in the r-direction and implicitly in the u-direction. Using a forward-difference approximation for the time derivative and central-difference approximations for the spatial derivatives, we obtain the following expression:

$$\begin{aligned}
&F_{m+1}^{i,j} - \kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F_{m+1}^{i,j+1} - F_{m+1}^{i,j-1}) \\
&- \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(F_{m+1}^{i,j+1} - 2F_{m+1}^{i,j} + F_{m+1}^{i,j-1}) = \\
&\frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F_{m+\frac{1}{2}}^{i+1,j} - F_{m+\frac{1}{2}}^{i-1,j}) \\
&+ \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F_{m+\frac{1}{2}}^{i+1,j} - 2F_{m+\frac{1}{2}}^{i,j} + F_{m+\frac{1}{2}}^{i-1,j}) \\
&+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F_{m+\frac{1}{2}}^{i,j},
\end{aligned}$$

for $1 < j < \frac{\bar{u}}{\delta_u} - 1$. For $j = 1$ and $j = \frac{\bar{u}}{\delta_u} - 1$, we cannot use a central-difference approximation for the second derivative with respect to the unspanned volatility factor. The central-difference approximation for the second derivative contradicts the assumption of the continued slope, and as such results in an approximation of 0 for the second derivative which is clearly incorrect. Instead, we use a forward-difference approximation around $F_m^{i,j-1}$ when $j = 1$ and a backward-difference approximation around $F_m^{i,j+1}$ when $j = \frac{\bar{u}}{\delta_u} - 1$. We then obtain the following ex-

pression when $j = 1$:

$$
\begin{aligned}
&F^{i,j}_{m+1} - \kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F^{i,j+1}_{m+1} - F^{i,j-1}_{m+1}) \\
&- \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(-F^{i,j+2}_{m+1} + 4F^{i,j+1}_{m+1} - 5F^{i,j}_{m+1} + 2F^{i,j-1}_{m+1}) = \\
&\frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}) \\
&+ \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}) \\
&+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F^{i,j}_{m+\frac{1}{2}},
\end{aligned}
$$

and when $j = \frac{\bar{u}}{\delta_u} - 1$:

$$
\begin{aligned}
&F^{i,j}_{m+1} - \kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F^{i,j+1}_{m+1} - F^{i,j-1}_{m+1}) \\
&- \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(-F^{i,j-2}_{m+1} + 4F^{i,j-1}_{m+1} - 5F^{i,j}_{m+1} + 2F^{i,j+1}_{m+1}) = \\
&\frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}) \\
&+ \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}) \\
&+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F^{i,j}_{m+\frac{1}{2}}.
\end{aligned}
$$

The above second system can be rewritten in vector/matrix form[3] as

$$
\begin{aligned}
F^{i,\cdot}_{m+1} = [&\gamma_1(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,\cdot}_{m+\frac{1}{2}} - F^{i-1,\cdot}_{m+\frac{1}{2}}) \\
&+ \gamma_2(i\delta_r(\bar{r} - i\delta_r)\delta_u)^2||\mathbf{H}_2||^2 \cdot (F^{i+1,\cdot}_{m+\frac{1}{2}} - 2F^{i,\cdot}_{m+\frac{1}{2}} + F^{i-1,\cdot}_{m+\frac{1}{2}}) \\
&+ (1 - i\delta_r\frac{\delta_\tau}{2})F^{i,\cdot}_{m+\frac{1}{2}}]\mathbf{G}_2^{-1},
\end{aligned}
$$

where $||\cdot||$ is defined as the norm of a vector, and

$$
\mathbf{G}_2 = \mathbf{I} - \gamma_3\kappa\mathbf{S}_2^{'}\mathbf{D}_5 - \gamma_4\sigma^2\mathbf{T}_2^{'}\mathbf{D}_6,
$$

where $\mathbf{I}$, $\mathbf{D}_5$, $\mathbf{D}_6$, $\mathbf{S}_2$ and $\mathbf{T}_2$ are all of size $(\bar{u}/\delta_u - 1)$ x $(\bar{u}/\delta_u - 1)$. $\mathbf{I}$ is again an identity matrix,

$$
\mathbf{D}_6 = \begin{bmatrix}
\delta_u & 0 & \cdots & 0 \\
0 & 2\delta_u & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & \bar{u} - \delta_u
\end{bmatrix},
$$

[3] We provide the proof for this derivation of the second system in the Appendix A.2.

$$\mathbf{D}_5 = \theta\mathbf{I} - \mathbf{D}_6,$$

$$\mathbf{S}_2 = \begin{bmatrix} -2 & 2 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \ddots & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 1 & 0 \\ \vdots & \ddots & \ddots & -1 & 0 & 1 \\ 0 & \cdots & 0 & 0 & -2 & 2 \end{bmatrix}, \quad \mathbf{T}_2 = \begin{bmatrix} -1 & 2 & -1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & -1 & 2 & -1 \end{bmatrix}, \text{ and}$$

$\mathbf{H}_2$ is a vector of size $(\bar{u}/\delta_u - 1)$ x 1, and is defined as

$$\mathbf{H}_2 = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ \frac{\bar{u}}{\delta_u} - 1 \end{bmatrix}.$$

Note that $\mathbf{G}_2$ is not a tridiagonal matrix[4], hence as such we cannot use the Thomas algorithm. There is still a computational efficiency as we can compute the inverse of $\mathbf{G}_2$ only once as it does not depend on i.

We now implement the above initial boundary problem for a specific example. Let us consider a European put option written on a two-year ZCB, with maturity in one year's time. The payoff at time $T = 1$ is $(K - P_{12})^+ = (K - g(1,2) + f(1,2)r_1)^+$. The left-hand panel of Figure 3.1 plots the pricing surface of the ZCB put option obtained from the ADI scheme, for an at-the-money (ATM) strike. The ATM strike is calculated assuming an initial short rate, $r_0 = 0.03$. While, in the right-hand panel, we verify the ADI scheme by comparing its estimates to the 99% Monte-Carlo confidence interval, over a range of strikes whilst fixing the initial unspanned volatility value.

The 99% Monte-Carlo (MC) confidence interval is defined as the MC estimate of the realised discounted payoffs, shifted up and down by three times the estimated standard deviation. The MC estimates are calculated using an Euler-Maruyama discretisation of the LADQ(1,1) dynamics, with 200 time steps for the one-year. The Euler-Maruyama discretisation of the LADQ(1,1) model dynamics, given by Equations (2.4) and (2.5), is given by

$$\hat{r}_i = \begin{cases} r_0 & i = 0, \\ \hat{r}_{i-1} + (\hat{r}_{i-1} - \lambda_1)(\hat{r}_{i-1} - \lambda_2)\Delta t + \hat{r}_{i-1}(\bar{r} - \hat{r}_{i-1})\hat{u}_{i-1}\Delta W_t^{\mathbb{Q},1} & \text{otherwise.} \end{cases}$$

[4] If we separately represented the system for $j = 1$ and $j = \frac{\bar{u}}{\delta_u} - 1$ and were left with the intermediate points, then we would have a tridiagonal matrix and be able to use the Thomas algorithm.

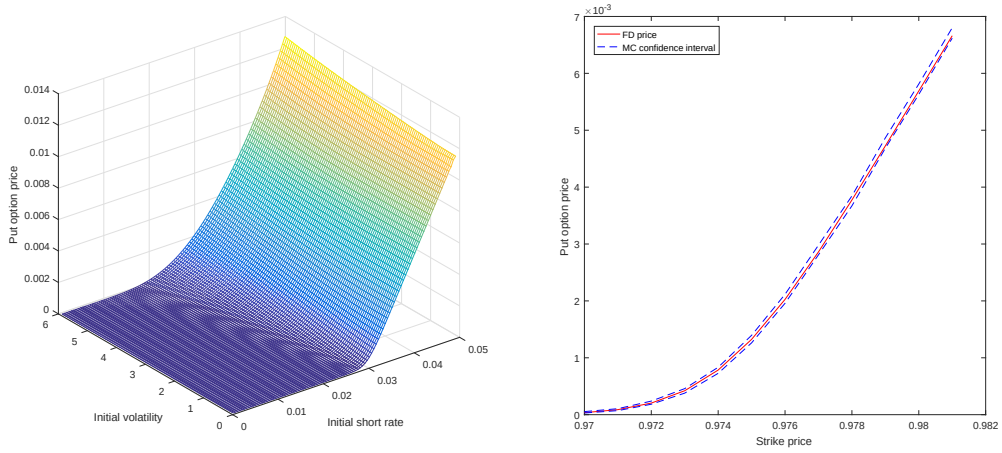


Fig. 3.1: Finite-difference (FD) prices of European put options written on a two-year ZCB, with expiry of one year, are plotted in both panels. The left-hand panel illustrates the whole FD pricing surface for one put struck ATM. The right-hand panel plots prices of puts for various strikes, whilst fixing $u_0 = 1.5$, and also plots 99% Monte-Carlo (MC) confidence bounds.

and

$$\hat{u}_i = \begin{cases} u_0 & i = 0, \\ \max\left(\hat{u}_{i-1} + \kappa(\theta - \hat{u}_{i-1})\Delta t + \sigma\sqrt{\hat{u}_{i-1}}\Delta W_t^{\mathbb{Q},2}, 0\right) & \text{otherwise.} \end{cases}$$

Note, the max function has to be included to ensure that the unspanned volatility component is positive for all time steps. For each strike value, we implement 5000 such paths and estimate the discount factor using the trapezoidal rule.

We can now easily extend this to pricing caps (or caplets). Firstly the payoff of a caplet on the simple interest rate $L(T, S)$ over the period $[T, S]$, expiring at time T and struck at rate K, is

$$\tau(L(T, S) - K)^+ \quad \text{at time } S,$$

where $\tau = S - T$. This is equivalent to a payoff at expiration of

$$\begin{aligned} & P_{TS}(L(T,S) - K)^+\tau \\ &= P_{TS}(\frac{P_{TT}}{P_{TS}} - 1 - K\tau)^+ \\ &= (1 - (1 + K\tau)P_{TS})^+ \\ &= (1 + K\tau)\left(\frac{1}{1 + K\tau} - P_{TS}\right)^+ \\ &= (1 + K\tau)\left(\frac{1}{1 + K\tau} - g(T,S) + f(T,S)r_T\right)^+ . \end{aligned} \tag{3.4}$$

Thus, a caplet struck at rate K is equivalent to $(1 + K\tau)$-many put options on the ZCB P_{TS}, with strike $\frac{1}{1+K\tau}$ and expiry at time T.

The adjustments required to price a caplet are to first change the initial condition to the above Equation (3.4) whilst also reversing time. As a result, the boundary condition for $r = \bar{r}$ is also adjusted. The Neumann boundary condition becomes

$$\frac{\partial}{\partial r}F(\tau, \bar{r}, u) = (1 + K\tau)f(T,S).$$

The value for b_1 and b_2 also need to be adjusted by a factor of $(1 + K\tau)$.

Using the above adjustments we can price a cap which is a portfolio of caplets with increasing expiries. Though finite difference methods are relatively slow, there is an applicable efficiency when pricing a cap. We can just price the longest-dated caplet and the intermediary solutions provide the prices for the shorter dated caplets (Backwell, 2017). This is because the payoff of a caplet depends on the time until maturity, τ (not the specific period $[T, S]$), and this is the same for each caplet within a cap.

3.2 Implementing the Fourier method for the LRSQ(1,1) model

In this section, we show how we can directly price a caplet (and thus a cap) using the Fourier method within the LRSQ(1,1) model. We apply the pricing method of a swaption outlined by Filipović *et al.* (2017) to the LRSQ(1,1) model, by noting that a caplet represents a one-period swaption. The payoff of a caplet on the simple interest rate $L(T_0, T_1)$ expiring at time T_0, struck at rate K, is

$$\tau(L(T_0, T_1) - K)^+ \quad \text{at time } T_1,$$

where $\tau = T_1 - T_0$. This is equivalent to a payoff at expiration, denoted by C_{T_0}, of

$$\begin{aligned} C_{T_0} &= (1 - (1 + K\tau)P_{T_0,T_1})^+ \\ &= \left(\sum_{i=0}^{1} c_i P_{T_0,T_i}\right)^+, \end{aligned} \tag{3.5}$$

with $c_0 = 1$ and $c_1 = -(1 + K\tau)$. Note that thus far, we have not implemented features of the LRSQ(1,1) model. Throughout this pricing method we use the state price density approach given by Filipović *et al.* (2017), as it is simpler and enables us to use the exponential affine transform method. Within the class of linear-rational term structure models, by applying the fundamental pricing formula (2.8), the time-T_0 price of a T_i-dated ZCB is

$$P_{T_0,T_i} = \frac{1}{\zeta_{T_0}} \mathbb{E}^{\mathbb{P}} \left[\zeta_{T_i} | \mathcal{F}_{T_0}\right].$$

Using the above result, Equations (2.10) and (2.11), we can express Equation (3.5) as

$$C_{T_0} = p_{\text{caplet}}(Z_{T_0})^+ / \zeta_{T_0},$$

where $p_{\text{caplet}}(z)$ is the explicit linear function of z given by

$$p_{\text{caplet}}(z) = \sum_{i=0}^{1} c_i \mathrm{e}^{-\alpha T_i}(1 + \theta + \mathrm{e}^{-\kappa(T_i - T)}(z - \theta)). \tag{3.6}$$

Let Cpl^K_{0,T_0,T_1} denote the time-0 price of a caplet written on a forward rate $f(T_0, T_1)$ struck at rate K. The time-0 price of the caplet is then obtained by once again applying the fundamental pricing formula (2.8), which yields

$$\text{Cpl}^K_{0,T_0,T_1} = \frac{1}{\zeta_0} \mathbb{E}^{\mathbb{P}}[\zeta_{T_0} C_{T_0}] = \frac{1}{\zeta_0} \mathbb{E}^{\mathbb{P}}[p_{\text{caplet}}(Z_{T_0})^+]. \tag{3.7}$$

To compute the price of a caplet, one needs to evaluate the expectation on the right-hand side of (3.7). The distribution of Z_{T_0} is unknown; however, there is an efficient numerical approach based on Fourier transform methods. Following the theorem[5] of Filipović *et al.* (2017), we define

$$\hat{q}(x) = \mathbb{E}[\exp(x p_{\text{caplet}}(Z_{T_0}))], \tag{3.8}$$

for $x \in \mathbb{C}$. Let μ be any positive constant such that $\hat{q}(\mu) < \infty$. Then the caplet price is given by

[5] Proof of the theorem is given by Filipović *et al.* (2017) and is easily applied to a caplet since a caplet is a one-period payer swaption. In Appendix B.1, we show the proof applied for a caplet.

$$\text{Cpl}^{K}_{0,T_0,T_1} = \frac{1}{\zeta_0 \pi} \int_0^{\infty} \text{Re}\left[\frac{\hat{q}(\mu + i\lambda)}{(\mu + i\lambda)^2}\right] d\lambda. \tag{3.9}$$

Computing $\hat{q}(\mu + i\lambda)$ initially looks difficult; however, given the specification of the LRSQ(1,1) model, it amounts to solving a system of ordinary differential equations. This is because the transformation (Z_t, U_t) is a linear transform of a square-root diffusion process, meaning the function in Equation (3.9) can be expressed using the exponential-affine transform formula (which can be found originally in Duffie *et al.* (2000)). We apply here the exponential affine transform formula applied for the dynamics of X_t in the LRSQ(1,1) model. We have reproduced the exponential affine transform formula in Appendix B.2.

Note that $\hat{q}(\mu+i\lambda)$ can be expressed as an expectation under the real-world measure $\mathbb{P}$ of the form $\mathbb{E}^{\mathbb{P}}[\mathrm{e}^{u+v^T X_{T_0}}]$, where $u \in \mathbb{C}$ and $v \in \mathbb{C}^2$, so that the exponential affine transform formula can be applied.

Given Equation (3.6), the exponent in the right hand side of Equation (3.8) is

$$\begin{aligned}(\mu + i\lambda)p_{\text{caplet}}(Z_{T_0}) &= (\mu + i\lambda)[\mathrm{e}^{-\alpha T_0}(1 + Z_{T_0}) - \mathrm{e}^{-\alpha T_1}(1 + K\tau)(1 + (1 - \mathrm{e}^{-\alpha\tau})\theta \\ &\quad + \mathrm{e}^{-\kappa T_\tau} Z_{T_0})] \\ &= u + v^T X_{T_0},\end{aligned}$$

where

$$\begin{aligned}u &= (\mu + i\lambda)\mathrm{e}^{-\alpha T_0}[1 - \mathrm{e}^{-\alpha\tau}(1 + K\tau)(1 + (1 - \mathrm{e}^{-\alpha\tau})\theta)], \quad \text{and} \\ v &= (\mu + i\lambda)\mathrm{e}^{-\alpha T_0}[1 - \mathrm{e}^{-(\alpha+\kappa)\tau}(1 + K\tau)]\mathbf{1},\end{aligned}$$

with $\mathbf{1} = (1, 1)^\top$.

We now apply the exponential-affine transform formula and get the following result

$$\hat{q}(\mu + i\lambda) = \mathbb{E}^{\mathbb{P}}[\mathrm{e}^{u+v^T X_{T_0}}] = \mathrm{e}^{\Phi(T_0)+\Psi(T_0)^T X_0},$$

where $\Phi : \mathbb{R}_+ \to \mathbb{C}, \Psi : \mathbb{R}_+ \to \mathbb{C}^2$ solve the system

$$\begin{aligned}\Phi'(\tau) &= [\kappa(\theta - \theta_U), \kappa\theta_U]\Psi(\tau), \\ \Psi_i'(\tau) &= -[\kappa, 0]\Psi(\tau) + \frac{1}{2}\sigma_i^2 \Psi_i(\tau)^2, \quad i = 1, 2,\end{aligned}$$

with initial conditions $\Phi(0) = u$ and $\Psi(0) = v$.

We may now implement the above Fourier-based method to price a particular caplet. Let us consider a caplet written on the simple interest rate $L(1, 2)$ with maturity in one year's time. In Figure 3.2, we verify the Fourier-based method by comparing its estimates to the 99% Monte-Carlo confidence interval, over a range of strikes whilst fixing the initial term structure value, Z_0, and initial USV factor U_0.

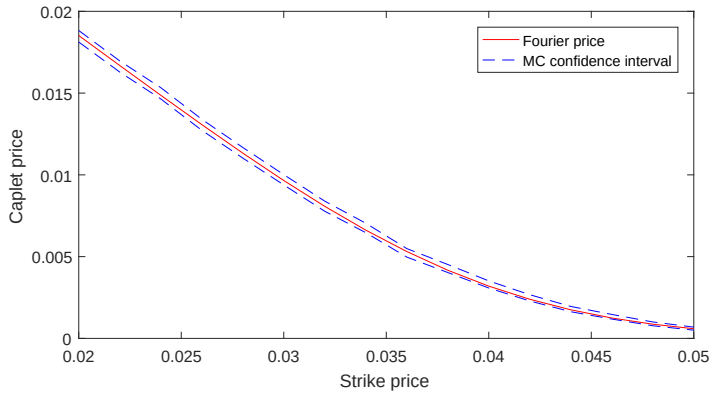


Fig. 3.2: Fourier-based method prices a caplet written on a forward rate $f(1,2)$, with expiry of one year, is plotted for varying strikes, whilst fixing $Z_0 = 0.12$ and $U_0 = 0.04$. We verify the Fourier-based method by plotting 99% Monte-Carlo (MC) confidence bounds.

The MC estimates are calculated using an Euler-Maruyama discretisation of the LRSQ(1,1) dynamics, with 100 time steps per year. The Euler-Maruyama discretisation of the LRSQ(1,1) model dynamics, given by Equations (2.15) and (2.16), is of the following form:

$$\hat{Z}_i = \begin{cases} Z_0 & i = 0, \\ \hat{Z}_{i-1} + \kappa(\theta_Z - \hat{Z}_{i-1})\Delta t + \sigma_1\sqrt{\max(\hat{Z}_{i-1} - \hat{U}_{i-1}, 0)\Delta t}\Delta W_t^{\mathbb{P},1} & \\ +\sigma_2\sqrt{\hat{U}_{i-1}\Delta t}\Delta W_t^{\mathbb{P},2} & \text{otherwise.} \end{cases}$$

and

$$\hat{U}_i = \begin{cases} U_0 & i = 0, \\ \max(\hat{U}_{i-1} + \kappa(\theta_U - \hat{U}_{i-1})\Delta t + \sigma_2\sqrt{\hat{U}_{i-1}\Delta t}\Delta W_t^{\mathbb{P},2}, 0) & \text{otherwise.} \end{cases}$$

Note, the max function needs to be included in the Euler-Maruyama discretisation of the term structure process Z_t to ensure we don't have a complex value for the value of the term structure factor at any time during its path. We also include a max function for the USV factor to ensure the USV factor remains positive throughout its path, since $U_t \in \mathbb{R}_+$.

Chapter 4

Calibration

In this chapter, we initially pre-calibrate each (1,1) model to the British-pound yield curve. We then vary the remaining parameters within each (1,1) model specification given in Chapter 2, such that the model-implied difference cap prices match the difference cap prices written on British-pound LIBOR as closely as possible. This is done by minimising the sum of squared errors. Following calibration, we consider the effect of the parameters within each model on the cap volatility surface.

4.1 Data

The dataset we work with is based on caps written on British-pound LIBOR on the trading day 1 December 2014. On this day, we have a bootstrapped yield curve up to a maturity of ten years and a term structure of implied volatilities for the one-year, two-year, three-year, four-year and five-year caps, each struck at 1%, 1.5% and 2%. A cap-implied volatility is the volatility that equates the Black formula to the market price of the cap. It is another way to quote the cap price.

The implied volatilities are converted into cap prices, using the relevant yield curve and the Black formula. The one-year cap, for each strike, consists of four quarterly caplets with payout dates in three, six, nine and twelve months' time. The two-year cap comprises of eight such adjacent, quarterly caplets and so on, with the five-year comprising of twenty quarterly caplets. We separate the caps into difference caps to ensure that each cap does not overlap each other and we may then weight each instrument equally when we calibrate. A difference cap is constructed by taking the difference between each consecutive cap and comprises of four quarterly caplets. For instance, the three-year difference cap is calculated by taking the difference between the three-year cap price and two-year cap price, such that the first quarterly payout begins in two and a quarter years.

We calibrate to the three-year, four-year and five-year difference caps volatility skew for each (1,1) model. We attempted to include the earlier caplets, but

had difficulty in matching the short-term prices (with the one model significantly overpricing and the other underpricing). Matching short-term interest rate option prices, consistently with longer-term prices, is known to be a potentially difficult problem (see, for instance, Moraleda and Vorst (1997)), but we do not attempt to address here.

4.2 Calibrating to both the yield curve and caps market

In this section, we use the numerical methods outlined in Chapter 3 to calibrate each (1,1) model to the above caps dataset. Before calibration, we pre-calibrate the model to the relevant yield curve, since interest-rate derivatives are typically hedged with yield-curve instruments, such as swaps. When we then calibrate, we need to price caps consistently with the yield curve.

Following Backwell (2017), we use the technique originally described by Brigo and Mercurio (2001), where we shift the time-homogeneous short rate by the addition of a deterministic function, denoted by $\varphi(t)$. The original LADQ(1,1) model specification, given by Equations (2.4) and (2.5), now describes the unshifted process, denoted by $\{r_t^*\}$. The actual, shifted short rate is defined as

$$r_t = r_t^* + \varphi(t)$$

for all $t \in [0, S]$. The deterministic function $\varphi(\cdot)$ is set such that the initial term structure of the model fits with the market prevailing term structure.

Let$\{P_{0T}^{\lambda}\}$ denote the time-0 market-observed term structure of ZCB prices. The market ZCB prices can be related to the unshifted models' ZCB prices with the deterministic function according to the following:

$$P_{0T}^{\lambda} = \mathrm{e}^{-\int_0^T \varphi(s)ds}(g(0,T) - f(0,T)r_0^*), \tag{4.1}$$

where $T > 0$.

Under the shifted model, Brigo and Mercurio (2001) show that ZCB put options, denoted by ZCP_{tTS}^{K}, can be computed in terms of the unshifted model's prices with

$$ZCP_{tTS}^{K} = \mathrm{e}^{-\int_t^S \varphi(s)ds}\widetilde{ZCP}_{tTS}^{K\exp(\int_T^S \varphi(s)ds)}, \tag{4.2}$$

where $\widetilde{ZCP}_{tTS}$ is the put option price under the unshifted model with an adjusted strike. We shall refer to the scaling of the unshifted model price given by $\mathrm{e}^{-\int_t^S \varphi(s)ds}$ as the scaling adjustment; and the strike adjustment is given by $\mathrm{e}^{\int_T^S \varphi(s)ds}$.

The above Equation (4.2) can be computed without explicitly knowing the deterministic function, $\varphi(t)$ (Backwell, 2017). This is because using Equation (4.1),

both the scaling adjustment and the strike adjustment can be expressed as

$$\mathrm{e}^{-\int_t^S \varphi(s)ds} = \frac{P_{0S}^{\lambda}\widetilde{P}_{0t}}{\widetilde{P}_{0S}P_{0t}^{\lambda}} \quad \text{and} \quad \mathrm{e}^{\int_T^S \varphi(s)ds} = \frac{\widetilde{P}_{0S}P_{0T}^{\lambda}}{P_{0S}^{\lambda}\widetilde{P}_{0T}}, \tag{4.3}$$

respectively, if the market prices $\{P_{0t}\}$, $\{P_{0t}\}$ and $\{P_{0S}\}$ are known. $\{\widetilde{P}_{0T}\}$ denotes the ZCB prices under the unshifted model, i.e., are given by the closed-form ZCB affine pricing formula.

Before applying the deterministic / time-dependent shift, we pre-calibrate by setting r_0 and λ_1 (in the affine coefficient functions) to the prevailing short-end and long-end of the yield curve, respectively. This ensures that we shift the model as little as possible, which seems desirable.

The remaining parameters within the model are calibrated to the difference caps volatility skew. In order to price the caps consistently with the yield curve, we use the numerical methods outlined in Chapter 2 to price caps whilst taking into account the adjustments in Equation (4.2) and (4.3).

Recall, the payoff of a caplet written on a forward rate $L(T,S)$ with expiry at time T, struck at rate K is

$$(1+K\tau)\left(\frac{1}{1+K\tau} - P_{TS}\right)^+ \quad \text{at time } T,$$

where $\tau = S - T$. Adjustments now need to be made. The scaling adjustment scales the number of portfolios of puts and the strike adjustment is multiplied to the strike (when observing the cap in the form of a ZCB put). Thus, the payoff of the adjusted caplet, with the same features as above, is

$$(1+K\tau)\mathrm{e}^{-\int_0^S \varphi(s)ds}\left(\left(\frac{1}{1+K\tau}\right)\mathrm{e}^{\int_T^S \varphi(s)ds} - P_{TS}\right)^+ \quad \text{at time } T. \tag{4.4}$$

4.2.1 LADQ(1,1) model calibration

In this section, we calibrate the LADQ(1,1) model to both the yield curve and the difference caps volatility skew.

We firstly set r_0 and λ_1 to the short-end and long-end of the yield curve, respectively. The shortest-term ZCB in our dataset is the three-month ZCB, as such we set

$$r_0 = -\frac{1}{0.25}\log P_{0,0.25},$$

and the longest-term ZCB is the 10-year ZCB, hence we set

$$\lambda_1 = -\frac{1}{10}\log P_{0,10}.$$

We also pre-calibrate λ_2. Recall, in Figure 2.2, λ_2 controls the rate of mean reversion, hence it should also dictate the speed of convergence between the short-end and long-end of the yield curve. This means that λ_2 controls the curvature of the yield curve, and thus we set λ_2 to match the shape of the intermediate yields in our dataset.

In order to avoid an overparameterisation to the term structure of the implied volatilities, we set the initial unspanned volatility, $u_0 = \theta$, to the level of its mean reversion.

The remaining four parameters ($\bar{r}$, κ, θ and σ) are calibrated to the three-year, four-year and five-year difference caps. We use the finite difference scheme outlined in Section 2.1 to price caplets, which we group into difference caps. However, as revealed above we need to adjust the caplet prices by both the scaling adjustment and the strike adjustment. Using Equations (4.3) and (4.4), the terminal payoff of a caplet, in terms of the short rate, is

$$(1+K\tau)\left(\frac{P^{\lambda}_{0S}}{\widetilde{P}_{0S}}\right)\left(\left(\frac{1}{1+K\tau}\right)\left(\frac{\widetilde{P}_{0S}P^{\lambda}_{0T}}{P^{\lambda}_{0S}\widetilde{P}_{0T}}\right)-g(T,S)+f(T,S)r_T\right)^{+}. \tag{4.5}$$

The adjustments to be made to the finite difference scheme (assuming the adjustments from a ZCB put to a caplet have been completed) are:

(i) Change the initial condition to the above Equation (4.5).

(ii) Adjust the Neumann boundary condition for $r = \bar{r}$ to:

$$\frac{\partial}{\partial r}F(\tau,\bar{r},u) = (1+K\tau)\left(\frac{P^{\lambda}_{0S}}{\widetilde{P}_{0S}}\right)f(T,S).$$

(iii) The values for b_1 and b_2 need to be adjusted by a factor of $\left(\frac{P^{\lambda}_{0S}}{\widetilde{P}_{0S}}\right)$.

Note that the efficient method for pricing a cap is no longer viable anymore. This is because the payoffs of the various caplets differ due to their different strike adjustments.

In Figure 4.1, the resultant fit is plotted in terms of implied volatilities, and the resultant parameters are given in the caption. For each difference cap, their is an associated volatility skew. A volatility skew is a plot of the implied volatility relative to varying strikes. The fit to the market implied volatility skew is reasonably close for the four- and five-year difference caps, although the model cannot quite accommodate the degree of skew for the four-year difference cap. The model struggles to match the three-year difference cap volatility skew.

4.2.2 LRSQ(1,1) model calibration

In this subsection, we calibrate the LRSQ(1,1) model to both the yield curve and the difference cap volatility skew.

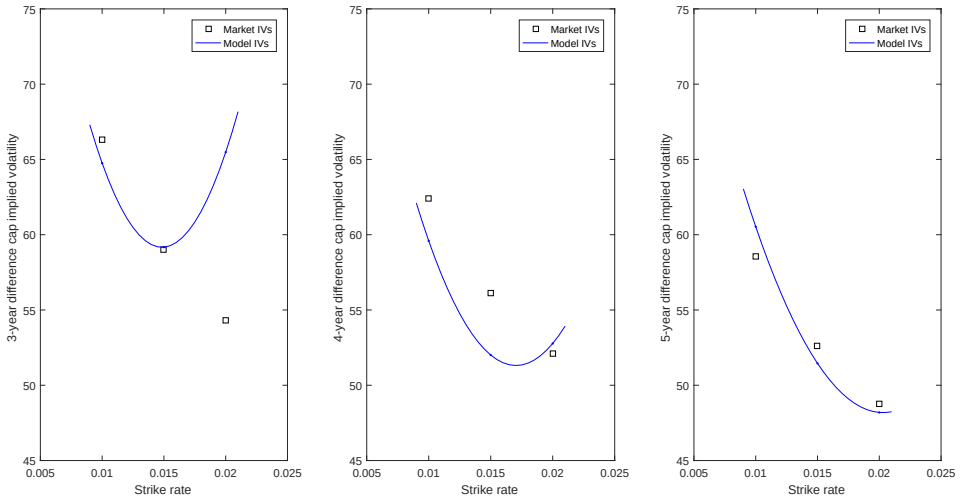


Fig. 4.1: The various panels show the model-implied difference cap-implied volatility as a function of strike price, using the parameters established by the calibration procedure outlined above. The resultant parameters are: $\lambda_1 = 0.0185$, $\lambda_2 = 0.7133$, $\bar{r} = 0.5546$, $\kappa = 0.1449$, $\theta = 2.2198$, $\sigma = 2.1001$, $r_0 = 0.0055$ and $u_0 = 2.2198$. Sum of squared differences $= 1.2003 \times 10^{-6}$.

We pre-calibrate the LRSQ(1,1) model by setting r_0 to the short-end of the yield curve and α, the analogue to λ_1 in the LADQ(1,1) model, is set to the long-end of the yield curve. Unlike, the LADQ(1,1) model, we do not though pre-calibrate κ which as we recall controls the curvature of the yield curve. This is because contrary to its analogue λ_2 in the LADQ(1,1) model, it also forms part of the risk-neutral drift term of the unspanned volatility as well as the volatility term of the short rate.

We once again avoid overparameterisation to the term structure of implied volatilities, by setting $U_0 = \theta_U$.

The remaining five parameters (κ, θ_Z, θ_U, σ_1 and σ_2), are calibrated to the three-year, four-year and five-year difference caps volatility skew. We use the Fourier-based method to price caplets, but need to include the adjustments. Equation (4.4) based on a simple interest rate $L(T_0, T_1)$ is equivalent to

$$\begin{aligned} &\mathrm{e}^{-\int_0^{T_1}\varphi(s)ds}\left(\mathrm{e}^{\int_{T_0}^{T_1}\varphi(s)ds} - (1+K\tau)P_{TS}\right)^+ \\ &= \mathrm{e}^{-\int_0^{T_1}\varphi(s)ds}\left(\sum_{i=0}^{1} c_i P_{T_0,T_i}\right)^+, \end{aligned} \tag{4.6}$$

with $c_0 = \mathrm{e}^{-\int_{T_0}^{T_1}\varphi(s)ds}$ and $c_1 = (1+K\tau)$.

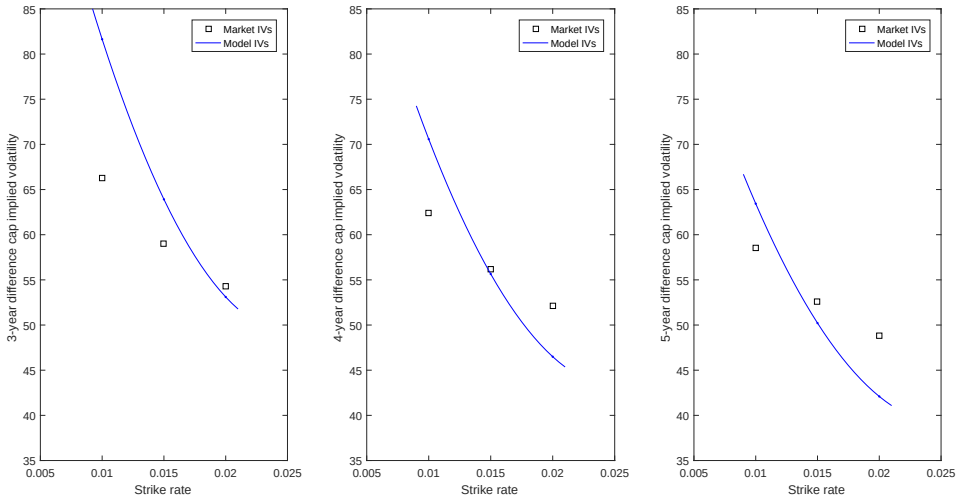


Fig. 4.2: The various panels show the model-implied difference cap-implied volatility as a function of strike price, using the parameters established by the calibration procedure outlined above. The resultant parameters are: $\alpha = 0.0185$, $\kappa = 0.0951$, $\theta_Z = 5.0107$, $\theta_U = 2.9850$, $\sigma_1 = 0.0556$, $\sigma_2 = 0.3048$, $r_0 = 0.0055$ and $U_0 = 2.2198$. Sum of squared differences $= 2.7753 \times 10^{-6}$.

By comparing the above Equation (4.6) to Equation (3.5), we recognise that the caplet payoff is scaled by $\mathrm{e}^{-\int_0^{T_1}\varphi(s)ds}$ and the value for c_0 has changed. Thus, the adjustments to the pricing method outlined in Section 3.2 are:

(i) Multiply Equation (3.9) by $\mathrm{e}^{-\int_0^{T_1}\varphi(s)ds} = \left(\frac{P^{\lambda}_{0,T_1}}{\widetilde{P}_{0,T_1}}\right)$.

(ii) Adjust initial conditions u and v to the exponential-affine transform formula, such that:

$$u = (\mu + i\lambda)\mathrm{e}^{-\alpha T_0}\left[\left(\frac{\widetilde{P}_{0S}P^{\lambda}_{0T}}{P^{\lambda}_{0S}\widetilde{P}_{0T}}\right) - \mathrm{e}^{-\alpha\tau}(1+K\tau)(1+(1-\mathrm{e}^{-\alpha\tau})\theta)\right]$$

$$v = (\mu + i\lambda)\mathrm{e}^{-\alpha T_0}\left[\left(\frac{\widetilde{P}_{0S}P^{\lambda}_{0T}}{P^{\lambda}_{0S}\widetilde{P}_{0T}}\right) - \mathrm{e}^{-(\alpha+\kappa)\tau}(1+K\tau)\right]\mathbf{1}.$$

In Figure 4.2, the resultant fit is plotted in terms of implied volatilities, and the resultant parameters are given in the caption. Unlike the LADQ(1,1) model where the skews were too curved, the shape of the model-fitted implied volatilities is similar to the market. However, the model produces a strongly downward sloping skew which slopes more than the market.

4.3 Parameter effects of (1,1) models

In this section, we examine the parameter effects of each (1,1) model on the implied volatility skew. Volatilities do not move independently and depend on three degrees of freedom: i) level, ii) slope and iii) curvature. Within each model, we analyse which parameter(s) control each of these degrees of freedom of the implied volatility skew. Finally, we examine the parameters controlling the term structure of implied volatilities.

4.3.1 LADQ(1,1) model: Parameter effects

In this section, we analyse the parameter effects within the LADQ(1,1) model. The difference cap volatility skews that are implied by the LADQ(1,1) model are displayed in Figure 4.3, where we vary each calibrated parameter separately. This enables us to inspect the effect each parameter has on the shape of the volatility skew.

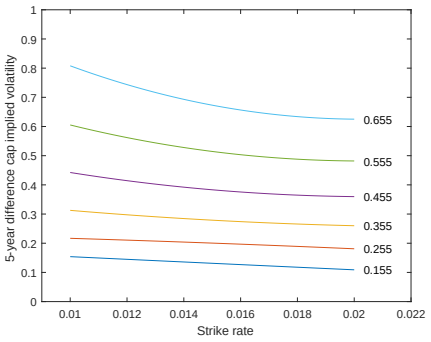
Recall from Section 2.1, θ determines how the local volatility will manifest, with a higher value of θ increasing the value of the total volatility. One would therefore expect that θ controls the level of the volatility skew. This is confirmed in panel(c) of Figure 4.3. It does appear that for lower values of θ, the volatility skew is perfectly linearly decreasing, whereas for higher values, the volatility skew becomes more curved in nature. Hence, θ has some minor influence over the curvature of the volatility skew.

The volatility parameter, σ, of the volatility process, controls the curvature of the volatility skew (Hagan *et al.*, 2002). This is confirmed by panel (d) of Figure 4.3. The curvature is limited, with values of σ higher than approximately 2.5 resulting in not much of a change to the shape of the volatility skew.

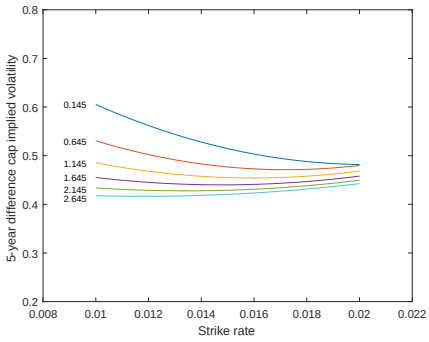
In panel (a) of Figure 4.3, as $\bar{r}$ increases, the level increases and slope becomes steeper of the volatility skew. This feature is indeed consistent with the interpretation that followed from the model specification, since recall, $\bar{r}$ increases the slope of the local volatility term whilst also scaling the level of the local volatility term, assuming a constant unspanned volatility term.

The rate of mean reversion parameter, κ, also controls the slope of the volatility skew, where the lower the value for κ, the larger the slope of the volatility skew (in absolute terms), and is confirmed in panel (b) of Figure 4.3.

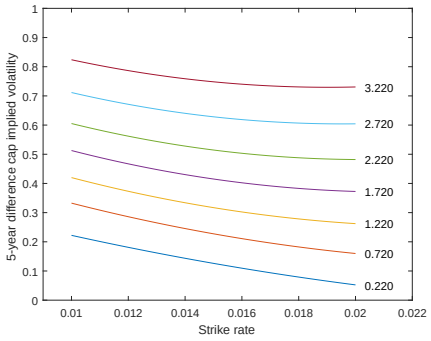
Thus, it appears that $\bar{r}$ and κ control the slope of the volatility skew. The higher the value for $\bar{r}$ and the lower the value for κ, the steeper the slope of the volatility skew (see Figure 4.4 for confirmation). However, this also results in the level of the volatility skew increasing, which could be offset by decreasing the value for θ.



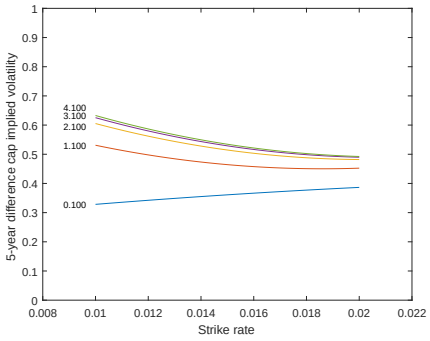
(a) $\bar{r}$ ranging from 0.1546 to 0.6546.



(b) κ ranging from 0.1449 to 2.6449.



(c) θ ranging from 0.2198 to 3.2198.



(d) σ ranging from 0.1001 to 4.1001.

Fig. 4.3: 5-year difference cap volatility skews implied by the LADQ(1,1) model, where in each panel a single parameter is varied.

Regarding the term structure of implied volatility, Figure 4.5 plots difference cap volatility curves implied by the LADQ(1,1) model, depending on the maturity of the difference cap, struck at the 1.5% strike rate. These graphs suggest u_0 and θ control the short-end and long-end of the volatility, respectively, and κ determines how fast the short volatilities converge to the long volatilities.

4.3.2 LRSQ(1,1) model: Parameter effects

Examples of the 5-year difference cap volatility skews that are implied by the LRSQ(1,1) model are displayed in Figure 4.6, where we vary each calibrated parameter separately. The volatility skews are also decreasing as strikes increase.

The graphs in Figure 4.6 suggest that the parameters have similar effects on the volatility skew. As a result, it is difficult to map specific parameters to specific features of the volatility skew. According to the interpretation of the parameters, θ_U and σ_2 control the level and curvature of the volatility skew, respectively. This

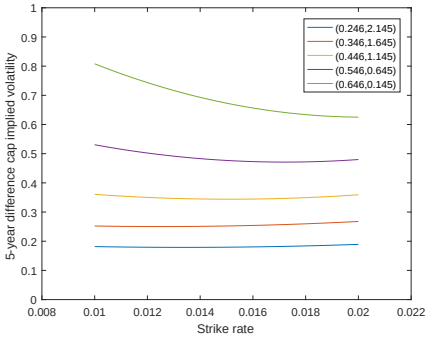
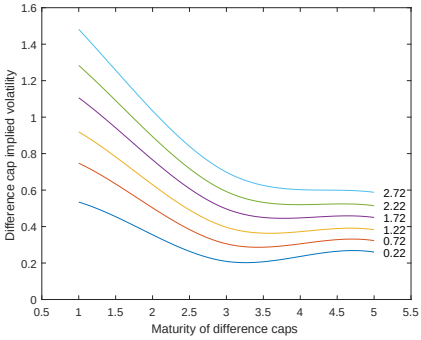
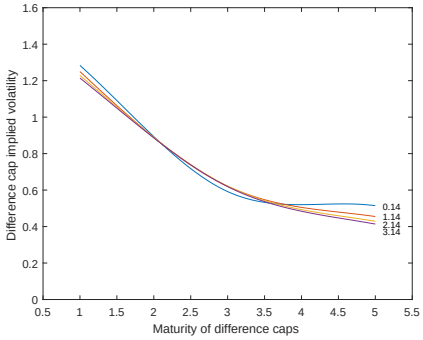


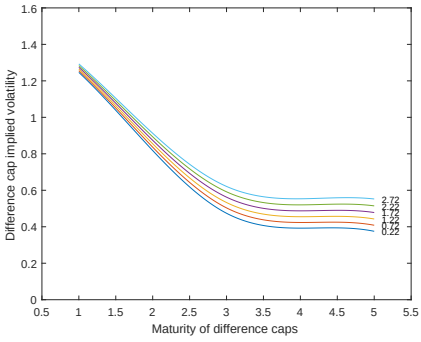
Fig. 4.4: 5-year difference cap volatility skew implied from the LADQ(1,1) model. The values for $\bar{r}$ and κ are varied resulting in differing slopes for the volatility skew.



(a) u_0 ranging from 0.2198 to 2.7198.



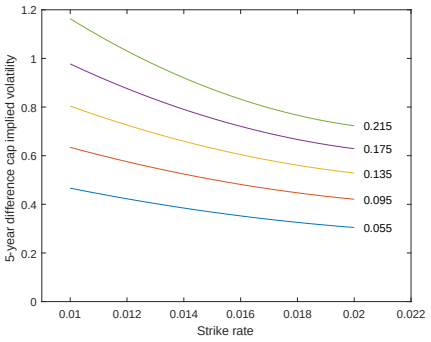
(b) κ ranging from 0.1449 to 2.6449.



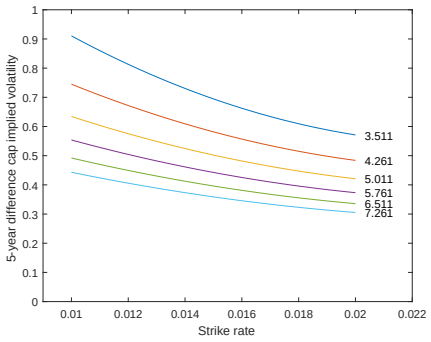
(c) θ ranging from 0.2198 to 2.7198.

Fig. 4.5: Difference cap volatility curves implied by the LADQ(1,1) model, where in each panel a single parameter is varied.

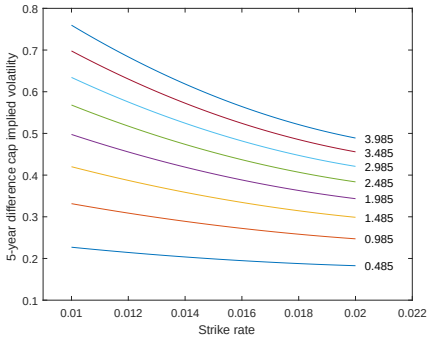
indeed appears the case in panel (c) and (e) of Figure 4.6. However, the graphs indicate that these parameters do not solely control one degree of freedom of the volatility skew. The larger the value for θ_U, the higher the level as well as the



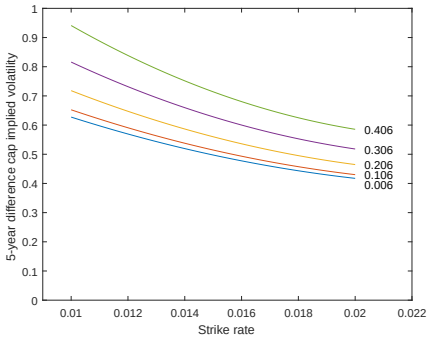
(a) κ ranging from 0.0551 to 0.2151.



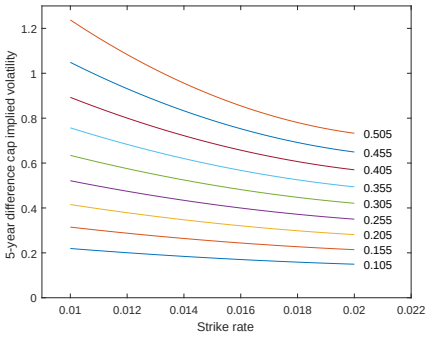
(b) θ_Z ranging from 3.5107 to 7.2607.



(c) θ_U ranging from 0.4850 to 3.9850.



(d) σ_1 ranging from 0.0056 to 0.4056.

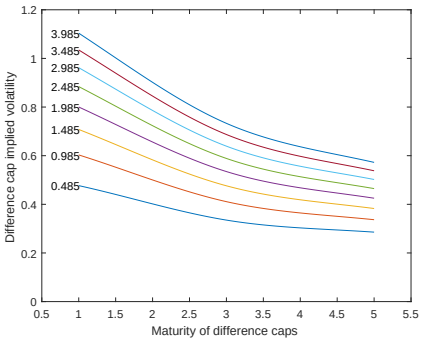


(e) σ_2 ranging from 0.1048 to 0.5048.

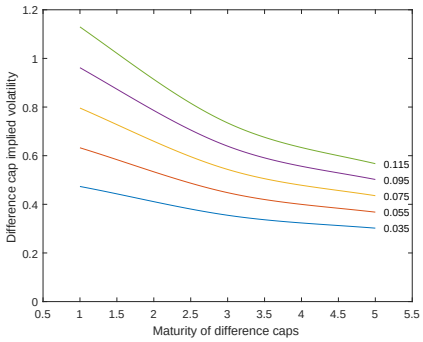
Fig. 4.6: 5-year difference cap volatility skews implied by the LRSQ(1,1) model, where in each panel a single parameter is varied.

steeper the slope of the volatility skew. σ_2 has control over all three degrees of freedom, with a higher value for σ_2 resulting in higher level, steeper slope and larger convexity of the volatility skew.

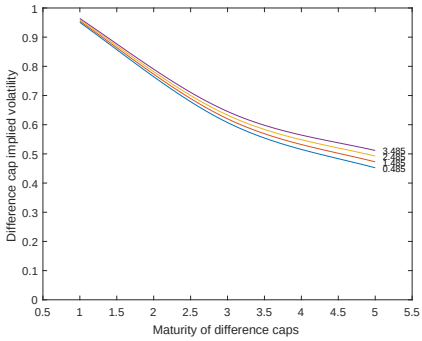
It appears that the remaining parameters, κ, θ_Z and σ_1, predominately control



(a) U_0 ranging from 0.4850 to 3.9850.



(b) κ ranging from 0.0351 to 0.1151.



(c) θ_U ranging from 0.4850 to 3.4850.

Fig. 4.7: Difference cap volatility curves implied by the LADQ(1,1) model, where in each panel a single parameter is varied.

the level of the volatility skew, with minor control over the other two degrees of freedom.

Regarding the term structure of implied volatility, Figure 4.7 plots difference cap volatility curves implied by the LRSQ(1,1) model, depending on the maturity of the difference cap. Consistent with the interpretation of the parameters, these graphs suggest U_0 and θ_U control the short-end and long-end of the volatility, respectively. κ should dictate how fast the short volatilities converge to the long volatilities, though this does not seem to be the case in panel (b) of Figure 4.7.

Chapter 5

Conclusion

This dissertation introduced the concept of USV, and outlined the pricing of interest rate derivatives with the two bivariate USV models in the literature. For the LRSQ(1,1) model, an efficient semi-analytical Fourier-based method was used. Since the LADQ(1,1) dynamics do not admit a characteristic function, an ADI finite difference scheme was implemented. Using these numerical pricing methods, we calibrated each (1,1) model to the British-pound yield curve and caps market.

Both (1,1) models struggled to fit to the market implied volatility surface, even when we excluded the short-term caps. The parameters in the LADQ(1,1) model had direct control over: i) level, ii) slope and iii) curvature of the implied volatility skew (but not in a way that matched how the skew changes over expiries), whereas the parameters in the LRSQ(1,1) model had limited direct control. We propose that future research examines reparameterising the LRSQ(1,1) model such that there is more direct control over the shape of the volatility skew.

A more thorough empirical investigation could also follow from Section 4.2, whereby each (1,1) model is calibrated to a series of volatility skews to examine the stability of the resultant parameters.

Bibliography

Backwell, A. (2017). Bivariate unspanned stochastic volatility models, *SSRN Working Paper* .

Brigo, D. and Mercurio, F. (2001). A deterministic-shift extension of analytically tractable and time-homogeneous short rate models, *Finance and Stochastics* **5**(3): 369–387.

Brigo, D. and Mercurio, F. (2007). *Interest rate models-theory and practice: with smile, inflation and credit*, Springer Science & Business Media.

Buckova, Z., Stehlikova, B. and Sevcovic, D. (2016). Numerical and analytical methods for bond pricing in short rate convergence models of interest rates, *arXiv preprint arXiv:1607.04968* .

Collin-Dufresne, P. and Goldstein, R. (2002). Do bonds span the fixed income markets? theory and evidence for unspanned stochastic volatility, *The Journal of Finance* **57**(4): 1685–1730.

Collin-Dufresne, P., Goldstein, R. and Jones, C. (2009). Can interest rate volatility be extracted from the cross section of bond yields?, *Journal of Financial Economics* **94**(1): 47–66.

Constantinides, G. (1992). A theory of the nominal term structure of interest rates, *The Review of Financial Studies* **5**(4): 531–552.

Duffie, D., Pan, J. and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**(6): 1343–1376.

Duffy, D. (2013). *Finite difference methods in financial engineering : a partial differential equation approach*, John Wiley & Sons Ltd.

Filipović, D., Larsson, M. and Trolle, A. (2017). Linear-rational term structure models, *The Journal of Finance* **72**(2): 655–704.

Fong, H. G. and Vasicek, O. A. (1991). Fixed-income volatility management, *The Journal of Portfolio Management* **17**(4): 41–46.

Hagan, P., Kumar, D., Lesniewski, A. and Woodward, D. (2002). Managing smile risk, *The Best of Wilmott* pp. 249–296.

Higham, N. (2002). Accuracy and stability of numerical algorithms, **80**.

Moraleda, J. and Vorst, T. (1997). Pricing american interest rate claims with humped volatility models, *Journal of Banking & Finance* **21**(8): 1131–1157.

Peaceman, D. W. and Rachford, H. H. (1955). The numerical solution of parabolic and elliptic differential equations, *Journal of the Society for Industrial and Applied Mathematics* **3**(1): 28–41.

Piazzesi, M. (2010). Affine term structure models, *Handbook of Financial Econometrics* **1**: 691–766.

Trolle, A. and Schwartz, E. (2008). A general stochastic volatility model for the pricing of interest rate derivatives, *The Review of Financial Studies* **22**(5): 2007–2057.

Vasicek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**(2): 177–188.

Appendix A

ADI scheme in vector / matrix form

A.1 ADI scheme for first half-time step

In the first-half time step of the ADI scheme, the PDE (3.1) is solved implicitly in the r-spatial direction and explictly in the u-spatial direction. We have the following form of the term structure PDE, where we have organised the implicit terms on the left-hand side (LHS) of the equation and the explicit terms on the right-hand side (RHS) of the equation,

$$\begin{aligned}
&F^{i,j}_{m+\frac{1}{2}} - \frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}) \\
&- \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}) = \\
&\kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F^{i,j+1}_m - F^{i,j-1}_m) + \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(F^{i,j+1}_m - 2F^{i,j}_m + F^{i,j-1}_m) \\
&+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F^{i,j}_m
\end{aligned} \tag{A.1}$$

We now show how the above system may be rewritten in column vector / matrix form. The LHS terms require some careful reasoning to rewrite in column vector / matrix form.

We first examine the term

$$F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}. \tag{A.2}$$

For $i = 1$, Term (A.2) equals

$$F^{2,j}_{m+\frac{1}{2}} - F^{0,j}_{m+\frac{1}{2}} = F^{2,j}_{m+\frac{1}{2}} - F^{1,j}_{m+\frac{1}{2}}$$

by using the Neumann boundary condition at $r = 0$.

For $i = \frac{\bar{r}}{\delta_r} - 1$, Term (A.2) equals

$$F^{\frac{\bar{r}}{\delta_r},j}_{m+\frac{1}{2}} - F^{\frac{\bar{r}}{\delta_r}-2,j}_{m+\frac{1}{2}} = F^{\frac{\bar{r}}{\delta_r}-1,j}_{m+\frac{1}{2}} - F^{\frac{\bar{r}}{\delta_r}-2,j}_{m+\frac{1}{2}} + \delta_r f(T,S)$$

by using the Neumann boundary condition at $r = \frac{\bar{r}}{\delta_r} - 1$. Hence, Term (A.2) in

column vector form is

$$\mathbf{S}_1 F^{\cdot,j}_{m+\frac{1}{2}} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \delta_r f(T,S) \end{bmatrix},$$

where

$$\mathbf{S}_1 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}.$$

We now examine the term

$$F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}. \tag{A.3}$$

For $i = 1$, Term (A.3) equals

$$F^{2,j}_{m+\frac{1}{2}} - 2F^{1,j}_{m+\frac{1}{2}} + F^{0,j}_{m+\frac{1}{2}} = F^{2,j}_{m+\frac{1}{2}} - F^{1,j}_{m+\frac{1}{2}}$$

For $i = \frac{\bar{r}}{\delta_r} - 1$, Term (A.3) equals

$$F^{\frac{\bar{r}}{\delta_r},j}_{m+\frac{1}{2}} - 2F^{\frac{\bar{r}}{\delta_r}-1,j}_{m+\frac{1}{2}} + F^{\frac{\bar{r}}{\delta_r}-2,j}_{m+\frac{1}{2}} = -F^{\frac{\bar{r}}{\delta_r}-1,j}_{m+\frac{1}{2}} + F^{\frac{\bar{r}}{\delta_r}-2,j}_{m+\frac{1}{2}} + \delta_r f(T,S)$$

Hence, Term (A.3) in column vector form is

$$\mathbf{T}_1 F^{\cdot,j}_{m+\frac{1}{2}} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \delta_r f(T,S) \end{bmatrix},$$

where

$$\mathbf{T}_1 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -2 & 1 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}.$$

The LHS of system (A.1) can now be expressed in column vector / matrix form as

$$\left(\mathbf{I} - \gamma_1 \mathbf{D}_1 \mathbf{D}_2 \mathbf{S}_1 - \gamma_2 \mathbf{D}_3^2 \mathbf{D}_4^2 \mathbf{T}_1 (j\delta_u)^2\right) F^{\cdot,j}_{m+\frac{1}{2}} - \mathbf{b}^j_m \tag{A.4}$$

where $\mathbf{I}$, $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_3$, $\mathbf{D}_4$, $\mathbf{S}_1$and $\mathbf{T}_1$ are all matrices of size $(\bar{r}/\delta_r - 1)$ x $(\bar{r}/\delta_r - 1)$. $\mathbf{I}$ is the identity matrix, $\gamma_1 = \frac{\delta_\tau}{4\delta_r}$, $\gamma_2 = \frac{\delta_\tau}{4\delta_r^2}$,

$$\mathbf{D}_2 = \begin{bmatrix} \delta_r & 0 & \cdots & 0 \\ 0 & 2\delta_r & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \bar{r} - \delta_r \end{bmatrix},$$

$$\begin{aligned} \mathbf{D}_1 &= \mathbf{D}_2 - \lambda_1 \mathbf{I}, \\ \mathbf{D}_3 &= \mathbf{D}_2 - \lambda_2 \mathbf{I}, \\ \mathbf{D}_4 &= \bar{r}\mathbf{I} - \mathbf{D}_2, \text{ and} \end{aligned}$$

$$b_m^j = \left(\gamma_1(\bar{r} - \delta_r - \lambda_1)(\bar{r} - \delta_r - \lambda_2) + \gamma_2(\bar{r} - \delta_r)^2\delta_r^2(j\delta_u)^2\right) \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \delta_r f(T,S) \end{bmatrix}. \tag{A.5}$$

Let us define:

$$\begin{aligned} b_1 &= \gamma_1(\bar{r} - \delta_r - \lambda_1)(\bar{r} - \delta_r - \lambda_2)\delta_r f(T,S), \text{ and} \\ b_2 &= \gamma_2(\bar{r} - \delta_r)^2\delta_r^3 f(T,S), \end{aligned}$$

then we may simplify (A.5) to

$$b_m^j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b_1 + b_2(j\delta_u)^2 \end{bmatrix}.$$

We define $\gamma_3 = \frac{\delta_\tau}{4\delta_u}$, $\gamma_4 = \frac{\delta_\tau}{4\delta_u^2}$ and

$$\mathbf{H}_1 = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ \frac{\bar{r}}{\delta_r} - 1 \end{bmatrix},$$

such that the RHS of system (A.1) can be expressed in column vector / matrix form as

$$\begin{aligned} &\gamma_3\kappa(\theta - j\delta_u)(F_m^{\cdot,j+1} - F_m^{\cdot,j}) + \sigma^2\gamma_4(j\delta_u)(F_m^{\cdot,j+1} - 2F_m^{\cdot,j} + F_m^{\cdot,j-1}) \\ &+ (\mathbf{1} - \mathbf{H}_1\frac{\delta_r\delta_\tau}{2}) \cdot F_m^{\cdot,j} \end{aligned} \tag{A.6}$$

where $\mathbf{1}$ is a vector of size $\left(\frac{\bar{r}}{\delta_r} - 1\right)$ x 1. By defining

$$\mathbf{G}_{m,1}^j = \mathbf{I} - \gamma_1\mathbf{D}_1\mathbf{D}_2\mathbf{S}_1 - \gamma_2\mathbf{D}_3^2\mathbf{D}_4^2\mathbf{T}_1(j\delta_u)^2,$$

and setting the LHS (A.4) equal to the RHS (A.6), the first-half time step of the system (A.1) in column vector / matrix form is given as

$$\begin{aligned} F^{\cdot,j}_{m+\frac{1}{2}} = (\mathbf{G}^{j}_{m,1})^{-1}[&\gamma_3\kappa(\theta - j\delta_u)(F^{\cdot,j+1}_{m} - F^{\cdot,j}_{m}) \\ &+ \sigma^2\gamma_4(j\delta_u)(F^{\cdot,j+1}_{m} - 2F^{\cdot,j}_{m} + F^{\cdot,j-1}_{m}) \\ &+ (\mathbf{1} - \mathbf{H}_1\frac{\delta_r\delta_\tau}{2}) \cdot F^{\cdot,j}_{m} + \mathbf{b}^{j}_{m}], \end{aligned}$$

for $0 < j < \frac{\bar{u}}{\delta_u}$.

A.2 ADI scheme for second half-time step

In the second-half time step of the ADI scheme, the PDE (3.1) is solved implicitly in the u-spatial direction and explictly in the r-spatial direction. We have the following form of the term structure PDE, where we have organised the implicit terms on the left-hand side (LHS) of the equation and the explicit terms on the right-hand side (RHS) of the equation,

$$\begin{aligned} &F^{i,j}_{m+1} - \kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F^{i,j+1}_{m+1} - F^{i,j-1}_{m+1}) \\ &- \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(F^{i,j+1}_{m+1} - 2F^{i,j}_{m+1} + F^{i,j-1}_{m+1}) = \\ &\frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}) \\ &+ \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}) \\ &+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F^{i,j}_{m+\frac{1}{2}}, \end{aligned} \tag{A.7}$$

for $1 < j < \frac{\bar{u}}{\delta_u} - 1$.

At the boundaries the central difference approximation for the second order derivative with respect to the unspanned volatility component cannot be used. We instead use a second-order forward difference approximation when $j = 1$, and get the following expression for the discretised PDE:

$$\begin{aligned} &F^{i,j}_{m+1} - \kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F^{i,j+1}_{m+1} - F^{i,j-1}_{m+1}) \\ &- \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(-F^{i,j+2}_{m+1} + 4F^{i,j+1}_{m+1} - 5F^{i,j}_{m+1} + 2F^{i,j-1}_{m+1}) = \\ &\frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}) \\ &+ \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}) \\ &+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F^{i,j}_{m+\frac{1}{2}}, \end{aligned} \tag{A.8}$$

and we use a second-order backward difference approximation when $j = \frac{\bar{u}}{\delta_u} - 1$, and get the following expression for the discretised PDE:

$$\begin{aligned}
&F^{i,j}_{m+1} - \kappa\frac{\delta_\tau}{4\delta_u}(\theta - j\delta_u)(F^{i,j+1}_{m+1} - F^{i,j-1}_{m+1}) \\
&- \sigma^2\frac{\delta_\tau}{4\delta_u^2}(j\delta_u)(-F^{i,j-2}_{m+1} + 4F^{i,j-1}_{m+1} - 5F^{i,j}_{m+1} + 2F^{i,j+1}_{m+1}) = \\
&\frac{\delta_\tau}{4\delta_r}(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,j}_{m+\frac{1}{2}} - F^{i-1,j}_{m+\frac{1}{2}}) \\
&+ \frac{\delta_\tau}{4\delta_r^2}(i\delta_r(\bar{r} - i\delta_r)j\delta_u)^2(F^{i+1,j}_{m+\frac{1}{2}} - 2F^{i,j}_{m+\frac{1}{2}} + F^{i-1,j}_{m+\frac{1}{2}}) \\
&+ (1 - (i\delta_r)\frac{\delta_\tau}{2})F^{i,j}_{m+\frac{1}{2}}.
\end{aligned} \tag{A.9}$$

We now show how the above system given by Equations (A.7), (A.8) and (A.9), can be expressed in row vector / matrix form.

We first examine the term

$$F^{i,j+1}_{m+1} - F^{i,j-1}_{m+1}. \tag{A.10}$$

For $j = 1$, Term (A.10) equals

$$F^{i,2}_{m+1} - F^{i,0}_{m+1} = 2F^{i,2}_{m+1} - 2F^{i,1}_{m+1}$$

by using the Neumann boundary condition at $u = 0$.

For $j = \frac{\bar{u}}{\delta_u} - 1$, Term (A.10) equals

$$F^{i,\frac{\bar{u}}{\delta_u}}_{m+1} - F^{i,\frac{\bar{u}}{\delta_u}-2}_{m+1} = 2F^{i,\frac{\bar{u}}{\delta_u}-1}_{m+1} - 2F^{i,\frac{\bar{u}}{\delta_u}-2}_{m+1}$$

by using the Neumann boundary condition at $u = \frac{\bar{u}}{\delta_u}$. Hence, Term (A.10) in row vector form is

$$\mathbf{S}_2 F^{i,\cdot}_{m+1},$$

where

$$\mathbf{S}_2 = \begin{bmatrix}
-2 & 2 & 0 & 0 & \cdots & 0 \\
-1 & 0 & 1 & \ddots & \ddots & \vdots \\
0 & -1 & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & 1 & 0 \\
\vdots & \ddots & \ddots & -1 & 0 & 1 \\
0 & \cdots & 0 & 0 & -2 & 2
\end{bmatrix}.$$

We now examine the second-order forward / backward difference approximation at $j = 1$ and $j = \frac{\bar{u}}{\delta_u} - 1$, respectively.

For $j = 1$, we have

$$\begin{aligned}&- F_{m+1}^{i,3} + 4F_{m+1}^{i,2} - 5F_{m+1}^{i,1} + 2F_{m+1}^{i,0}\\ &= -F_{m+1}^{i,3} + 2F_{m+1}^{i,2} - F_{m+1}^{i,1}.\end{aligned}$$

For $j = \frac{\bar{u}}{\delta_u} - 1$, we have

$$\begin{aligned}&- F_{m+1}^{i,\frac{\bar{u}}{\delta_u}-3} + 4F_{m+1}^{i,\frac{\bar{u}}{\delta_u}-2} - 5F_{m+1}^{i,\frac{\bar{u}}{\delta_u}-1} + 2F_{m+1}^{i,\frac{\bar{u}}{\delta_u}}\\ &= -F_{m+1}^{i,\frac{\bar{u}}{\delta_u}-3} + 2F_{m+1}^{i,\frac{\bar{u}}{\delta_u}-2} - F_{m+1}^{i,\frac{\bar{u}}{\delta_u}-1}.\end{aligned}$$

The approximation for the second-order derivative with respect to the unspanned volatility component in row vector form is

$$\mathbf{T}_2 F_{m+1}^{i,\cdot},$$

where

$$\mathbf{T}_2 = \begin{bmatrix} -1 & 2 & -1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & -1 & 2 & -1 \end{bmatrix}.$$

The LHS of the system can now be expressed in row vector / matrix form as

$$F_{m+1}^{i,\cdot}\left(\mathbf{I} - \gamma_3\kappa\mathbf{S}_2'\mathbf{D}_5 - \gamma_4\sigma^2\mathbf{T}_2'\mathbf{D}_6\right) \tag{A.11}$$

where $\mathbf{I}$, $\mathbf{D}_5$, $\mathbf{D}_6$, $\mathbf{S}_2$ and $\mathbf{T}_2$ are all of size $(\bar{u}/\delta_u - 1)$ x $(\bar{u}/\delta_u - 1)$. $\mathbf{I}$ is again an identity matrix,

$$\mathbf{D}_6 = \begin{bmatrix} \delta_u & 0 & \cdots & 0 \\ 0 & 2\delta_u & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \bar{u} - \delta_u \end{bmatrix}, \text{ and}$$

$$\mathbf{D}_5 = \theta\mathbf{I} - \mathbf{D}_6.$$

We now define

$$\mathbf{H}_2 = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ \frac{\bar{u}}{\delta_u} - 1 \end{bmatrix},$$

such that the RHS of the system can be expressed in row vector / matrix form as

$$\begin{aligned}
&\gamma_1(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,\cdot}_{m+\frac{1}{2}} - F^{i-1,\cdot}_{m+\frac{1}{2}}) \\
&+ \gamma_2(i\delta_r(\bar{r} - i\delta_r)\delta_u)^2||\mathbf{H}_2||^2 \cdot (F^{i+1,\cdot}_{m+\frac{1}{2}} - 2F^{i,\cdot}_{m+\frac{1}{2}} + F^{i-1,\cdot}_{m+\frac{1}{2}}) \\
&+ (1 - i\delta_r\frac{\delta_\tau}{2})F^{i,\cdot}_{m+\frac{1}{2}},
\end{aligned} \tag{A.12}$$

where $||\cdot||$ is defined as the norm of a vector.

By defining

$$\mathbf{G}_2 = \mathbf{I} - \gamma_3\kappa\mathbf{S}_2^{'}\mathbf{D}_5 - \gamma_4\sigma^2\mathbf{T}_2^{'}\mathbf{D}_6,$$

and setting the LHS (A.11) equal to the RHS (A.12), the second-half time step of the system in row vector / matrix form is given as

$$\begin{aligned}
F^{i,\cdot}_{m+1} = [&\gamma_1(i\delta_r - \lambda_1)(i\delta_r - \lambda_2)(F^{i+1,\cdot}_{m+\frac{1}{2}} - F^{i-1,\cdot}_{m+\frac{1}{2}}) \\
&+ \gamma_2(i\delta_r(\bar{r} - i\delta_r)\delta_u)^2||\mathbf{H}_2||^2 \cdot (F^{i+1,\cdot}_{m+\frac{1}{2}} - 2F^{i,\cdot}_{m+\frac{1}{2}} + F^{i-1,\cdot}_{m+\frac{1}{2}}) \\
&+ (1 - i\delta_r\frac{\delta_\tau}{2})F^{i,\cdot}_{m+\frac{1}{2}}]\mathbf{G}_2^{-1},
\end{aligned}$$

for $0 < i < \frac{\bar{r}}{\delta_u}$.

Appendix B

Fourier-based pricing method

B.1 Caplet price based on Fourier-transform methods

We use the following identity from Fourier analysis within the proof, valid for any $\mu > 0$ and $s \in \mathbb{R}$:

$$s^+ = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{(\mu+i\lambda)s} \frac{1}{(\mu+i\lambda)^2}. \tag{B.1}$$

Let $q(ds)$ denote the conditional distribution of the random variable $p_{\text{caplet}}(Z_{T_0})$ so that $\hat{q}(x) = \int_{\mathbb{R}} \mathrm{e}^{xs} q(ds)$ for $x \in \mathbb{C}$. Let $\mu > 0$ be such that $\hat{q}(\mu) < \infty$. Then

$$\int_{\mathbb{R}^2} \left| \mathrm{e}^{(\mu+i\lambda)s} \frac{1}{(\mu+i\lambda)^2} \right| d\lambda q(ds) = \int_{\mathbb{R}^2} \frac{\mathrm{e}^{\mu s}}{\mu^2+\lambda^2} d\lambda q(ds) = \int_{\mathbb{R}} \mathrm{e}^{\mu s} q(ds) \int_{\mathbb{R}} \frac{1}{\mu^2+\lambda^2} d\lambda < \infty,$$

where the second equality follows from Tonelli's theorem. Hence, this justifies applying Fubini's theorem in the following calculation:

$$\begin{aligned}
\mathbb{E}[p_{\text{caplet}}(Z_{T_0})^+] &= \int_{\mathbb{R}} s^+ q(ds) \\
&= \int_{\mathbb{R}} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{(\mu+i\lambda)s} \frac{1}{(\mu+i\lambda)^2} d\lambda \right) q(ds) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\hat{q}(\mu+i\lambda)}{(\mu+i\lambda)^2} d\lambda \\
&= \frac{1}{\pi} \int_0^{\infty} \mathrm{Re} \left[\frac{\hat{q}(\mu+i\lambda)}{(\mu+i\lambda)^2} \right] d\lambda
\end{aligned}$$

Here, the identity (B.1) is used in the second equality and Fubini's theorem is applied in the third equality. The final equality follows using the fact that the left, and hence, right side is real, and the integrand $(\mu+i\lambda)^{-2}\hat{q}(\mu+i\lambda)$ is an even function of λ. The resulting expression for the caplet price follows by using the above expectation and Equation (3.7).

B.2 Exponential affine transform formula

Let X_t be the square root process (2.14). For any $0 \leq t \leq T$, and $u \in \mathbb{C}$, $v \in \mathbb{C}^2$ such that $\mathbb{E}[|\exp(v^\top X_T)|] < \infty$ we have

$$\mathbb{E}_t \left[\mathrm{e}^{u+v^\top X_T} \right] = \mathrm{e}^{\Phi(T-t)+\Psi(T-t)^\top X_t},$$

where $\Phi : \mathbb{R}_+ \to \mathbb{C}$, $\Psi : \mathbb{R}_+ \to \mathbb{C}^2$ solve the system

$$\begin{aligned} \Phi'(\tau) &= b^\top \Psi(\tau), \\ \Psi_i'(\tau) &= \beta_i^\top \Psi(\tau) + \frac{1}{2}\sigma_i^2 \Psi_i(\tau)^2, \quad i = 1, 2, \end{aligned}$$

with initial condition $\Phi(0) = u$ and $\Psi(0) = v$. The solution to this system is unique.