QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 317 October 2012

Pricing Interest Rate Derivatives in a Multifactor HJM Model with Time Dependent Volatility

Ingo Benya, Carl Chiarella and Boda Kang

ISSN 1441-8010 www.qfrc.uts.edu.au

Pricing Interest Rate Derivatives in a Multifactor HJM Model with Time Dependent Volatility

Ingo Beyna
ingo.beyna@web.de
Frankfurt School of Finance & Management
Centre for Practical Quantitative Finance
Frankfurt, Germany

Carl Chiarella
carl.chiarella@uts.edu.au,
University of Technology Sydney,
Finance Discipline Group
Sydney, Australia

Boda Kang
boda.kang@uts.edu.au,
University of Technology Sydney,
Finance Discipline Group
Sydney, Australia

9th October 2012

Abstract

We investigate the partial differential equation (PDE) for pricing interest derivatives in the multi-factor Cheyette Model, which involves time-dependent volatility functions with a special structure. The high dimensional parabolic PDE that results is solved numerically via a modified sparse grid approach, that turns out to be accurate and efficient. In addition we study the corresponding Monte Carlo simulation, which is fast since the distribution of the state variables can be calculated explicitly. The results obtained from both methodologies are compared to the known analytical solutions for bonds and caplets. When there is no analytical solution, both European and Bermudan swaptions have been evaluated using the sparse grid PDE approach that is shown to outperform the Monte Carlo simulation.

Keywords: Cheyette Model, Gaussian HJM, Multi-Factor Model, PDE Valuation, Sparse Grid, Monte Carlo Simulation

1 Introduction

In 1992, Heath, Jarrow & Morton (1992) (henceforth HJM) developed a general framework to model the dynamics of the entire forward rate curve in an interest rate market. The associated valuation approach is based on two main assumptions: the first one postulates that it is not possible to gain riskless profit (The No arbitrage condition), and the second one assumes the completeness of the financial market. The HJM model, or strictly speaking the HJM framework, is a general model environment and incorporate many previously developed models like the model of Ho & Lee (1986), Vasicek (1977) and Hull & White (1990). The general setting mainly suffers from two disadvantages: first, the difficulty in applying the model in market practice and second, the extensive computational complexity caused by the high-dimensional stochastic process of the underlying. The first disadvantage was improved by the development of the LIBOR market model introduced by Brace, Gatarek & Musiela (1997), which combines the general risk-neutral yield curve model with market standard. The second disadvantage can be improved by restricting the general HJM model to a subset of models with a specific parametrization of the volatility function. The resulting system of stochastic differential equations (SDEs) describing the yield curve dynamics, reduces from a high-dimensional process into a low-dimensional structure of Markovian processes. This approach was developed by Cheyette (1992) and subsequently developed further by Bhar & Chiarella (1997), Chiarella & Kwon (2000) and Björk & Svensson (2001).

In practice, the Cheyette models usually incorporate several factors to achieve sufficient flexibility to represent the market state. The structure of the models supports a canonical construction of multifactor models as extensions of the one factor model. The model dynamics consider all factors and might become a high-dimensional SDE as each factor captures at least one dimension.

The class of Cheyette interest rate models is a specialization of the general HJM framework, so first we present the general set-up of HJM. We then limit the class of models to a specific process structure of the yield curve dynamics by imposing a parametrization of the forward rate volatility. This technique guides us to the representation of the class of Cheyette models.

The paper is organised as follows, the Cheyette model with the volatility

function is introduced in Section 2. The details of the 3-factor model are discussed in Section 3 after which a detailed analytical formula of both bond pricing and caplet pricing is derived in Section 4. The 3-factor model is calibrated to market data in Section 5. To demonstrate the performance of the numerical methods, the Quasi-Monte Carlo approach is studied in Section 6. We also formulate the price to be the solution of a four-dimensional partial differential equation (PDE) in Section 7. The comparison of the performance of different numerical approaches is demonstrated in Section 8 via a number of numerical examples before we draw the conclusion in Section 9. Lengthy derivations are given in the Appendix.

2 From the HJM to the Cheyette Model

Assume $B(t,T)$ to be the time t price of a zero-coupon bond maturing at time $T \geq t$. The usual continuously compounded forward rate at time t for instantaneous borrowing at time T is given by

$$f(t,T) = \frac{-\partial \ln B(t,T)}{\partial T},$$

or equivalently

$$B(t,T) = \exp\left(-\int_t^T f(t,u)du\right). \tag{1}$$

HJM have shown, that in any arbitrage-free term structure model with continuous evaluation of the yield curve, the forward rate has to satisfy

$$\mathrm{d}f(t,T) = \sum_{k=1}^{M}\left[\sigma_k(t,T)\int_t^T \sigma_k(t,s)\mathrm{d}s\right]\mathrm{d}t + \sum_{k=1}^{M}\sigma_k(t,T)\mathrm{d}W_k(t),$$

where $W(t)$ (with components $W_1(t),\cdots,W_M(t)$) is an M dimensional Brownian motion under the risk-neutral measure. The model is thus fully specified by a given volatility structure $\{\sigma(t,T)\}_{T\geq t}$ and the initial forward curve. The volatility function $\sigma(t,T)$ in an M factor model is an M dimensional vector

$$\sigma(t,T) = \begin{pmatrix} \sigma_1(t,T) \\ \vdots \\ \sigma_M(t,T) \end{pmatrix}.$$

The class of Cheyette interest rate models, first presented by Cheyette (1992), forms a subset of the general class of HJM models. As already suggested in the literature, one can choose a specific volatility structure that achieves an exogenous model of the yield curve with Markovian dynamics. We follow the approach of Cheyette (1996) and use a separable volatility term structure. Each component $\sigma_k(t,T)$ is assumed to be separable into time and maturity dependent factors and is parameterized by a finite sum of separable functions such that

$$\sigma_k(t,T) = \sum_{i=1}^{N_k} \frac{\alpha_i^{(k)}(T)}{\alpha_i^{(k)}(t)} \beta_i^{(k)}(t), \quad k = 1, ..., M, \tag{2}$$

where N_k denotes the number of volatility summands of factor k. The choice of the volatility structure affects the characteristic of the models as presented by Beyna & Wystup (2010). If we assume the volatility structure mentioned above, the forward rate can be reformulated as

$$\begin{aligned} f(t,T) = \quad & f(0,T) + \sum_{k=1}^{M} \left[\sum_{j=1}^{N_k} \frac{\alpha_j^{(k)}(T)}{\alpha_j^{(k)}(t)} \left(X_j^{(k)}(t) \right.\right. \\ & \left.\left. + \sum_{i=1}^{N_k} \frac{A_i^{(k)}(T) - A_i^{(k)}(t)}{\alpha_i^{(k)}(t)} V_{ij}^{(k)}(t) \right) \right], \end{aligned} \tag{3}$$

with deterministic variables

$$A_i^{(k)}(t) = \int_0^t \alpha_i^{(k)}(s) \mathrm{d}s, \tag{4}$$

$$V_{ij}^{(k)}(t) = V_{ji}^{(k)}(t) = \alpha_i^{(k)}(t) \alpha_j^{(k)}(t) \int_0^t \frac{\beta_i^{(k)}(s) \beta_j^{(k)}(s)}{\alpha_i^{(k)}(s) \alpha_j^{(k)}(s)} \mathrm{d}s, \tag{5}$$

and random variable

$$\begin{aligned} X_i^{(k)}(t) = & \alpha_i^{(k)}(t) \int_0^t \frac{\beta_i^{(k)}(s)}{\alpha_i^{(k)}(s)} \, \mathrm{d}W_k(s) + \\ & \alpha_i^{(k)}(t) \int_0^t \frac{\beta_i^{(k)}(s)}{\alpha_i^{(k)}(s)} \left[\sum_{j=1}^{N_k} \frac{A_j^{(k)}(t) - A_j^{(k)}(s)}{\alpha_j^{(k)}(s)} \beta_j^{(k)}(s) \right] \mathrm{d}s, \end{aligned} \tag{6}$$

for $k = 1, ..., M$ and $i, j = 1, ..., N_k$. The dynamics of the forward rate are determined by the state variables $X_i^{(k)}(t)$ for $k = 1, ..., M$ and $i = 1, ..., N_k$. Their dynamics are given by Markov processes as

$$\begin{aligned} \mathrm{d}X_i^{(k)}(t) = &\left(X_i^{(k)}(t) \frac{\partial(\log \alpha_i^{(k)}(t))}{\partial t} + \sum_{j=1}^{N_k} V_{ij}^{(k)}(t) \right) \mathrm{d}t \\ &+ \beta_i^{(k)}(t) \mathrm{d}W_k(t). \end{aligned} \tag{7}$$

Summarizing, the forward rate in an M-factor model is determined by $n = \sum_{k=1}^{M} N_k$ state variables.

The short rate $r(t)$ arises naturally as $r(t) = f(t, t)$ and applying the representation (3) leads directly to

$$r(t) = f(0, t) + \sum_{k=1}^{M} \sum_{j=1}^{N_k} X_j^{(k)}(t).$$

Thus, the short rate is given as the sum of the state variables.

The model set-up is quite general, however in practise it is popular to use parameterizations in the one factor case ($M = 1$) of the form $\sigma(t, T) = \mathbb{P}_m(t) \exp\Big(- \lambda(T - t)\Big)$, where $\mathbb{P}_m(t)$ denotes a polynomial of order m. Such volatility parameterizations can be obtained in the Cheyette Model by choosing the volatility function (2) for $M = 1$ and $N_1 = 1$ as $\alpha_1^{(1)}(t) = \exp(-\lambda t), \quad \beta_1^{(1)} = \mathbb{P}_m(t)$. These parameterizations cover well-known models such as the Ho-Lee Model, see Ho & Lee (1986), with

$$\sigma(t, T) = c$$

or the Hull-White Model, see Hull & White (1990), with

$$\sigma(t, T) = a \exp\Big(- \lambda(T - t)\Big).$$

A two factor Hull-White Model can be obtained easily by choosing

$$\sigma(t, T) = \begin{pmatrix} a_1 \exp(-\lambda_1(T - t)) \\ a_2 \exp(-\lambda_2(T - t)) \end{pmatrix}.$$

The representation of this two factor model in the Cheyette framework de-

termined by (2) is given by $M = 2, N_1 = 1, N_2 = 1$ and

$$\begin{aligned} \alpha_1^{(1)}(t) &= \exp(-\lambda_1 t), \quad \beta_1^{(1)}(t) = a_1, \\ \alpha_1^{(2)}(t) &= \exp(-\lambda_2 t), \quad \beta_1^{(2)}(t) = a_2. \end{aligned}$$

Remark 2.1.
Note that a necessary condition for the existence of a Markovian realization of the HJM Model provided in Björk & Svensson (2001) and Bhar & Chiarella (1997) is that the volatility $\sigma(T-t)$ as a function of time to maturity $T-t$ which only has the structure $\sigma(T-t) = \mathbb{P}_m(T-t)\exp(-\lambda(T-t))$, where $\mathbb{P}_m(x)$ denotes a polynomial of order m and $\lambda \in \mathbb{R}$ is a constant. In the most general case, the deterministic volatility functions have the form $\sigma(T-t) = \mathbb{P}_m(T-t)\exp(-\lambda(T-t))G(r(t))$, where $G(r(t))$ denotes a function of the instantaneous spot rate of interest. Nevertheless, the volatility structure is dependant on the time to maturity only. The work of Cheyette showed, one can achieve a Markovian realization of the HJM Model even if time and maturity are incorporated separately. Thus, it is possible to assume $\sigma(t,T) = \mathbb{P}_m(t)\exp\Big(-\lambda(T-t)\Big)$. Furthermore, the structure can be extended arbitrarily as long as it can be written in the form of (2). More details can be found in Appendix B.

3 The 3 Factor Exponential Model

To fill out the theory with some applications, we choose a specific model and use the PDE and the Monte Carlo Simulation to price interest rate derivatives. We choose a 3 factor model ($M = 3$), the 3 Factor Exponential Model, with volatility

$$\sigma(t,T) = \begin{pmatrix} \sigma_1(t,T) \\ \sigma_2(t,T) \\ \sigma_3(t,T) \end{pmatrix}, \tag{8}$$

with

$$\sigma_1(t,T) = c + \mathbb{P}_m^{(1)} \exp\Big(- \lambda^{(1)}(T-t)\Big), \tag{9}$$

$$\sigma_2(t,T) = \mathbb{P}_m^{(2)} \exp\Big(- \lambda^{(2)}(T-t)\Big), \tag{10}$$

$$\sigma_3(t,T) = \mathbb{P}_m^{(3)} \exp\Big(- \lambda^{(3)}(T-t)\Big), \tag{11}$$

where $c \in \mathbb{R}$ and $\lambda^{(k)} \in \mathbb{R}$ are constants and $\mathbb{P}_m^{(k)} = a_m^{(k)} t^m + ... + a_1^{(k)} t + a_0^{(k)}$ denotes a polynomial of order m. Using this parametrization, we have $N_1 = 2$, $N_2 = 1$ and $N_3 = 1$. Thus, the 3 Factor Exponential Model is driven by four state variables. The first factor is linked to the state variables $X_1^{(1)}(t)$ and $X_2^{(1)}(t)$ and the corresponding volatility parametrization is given by

$$\begin{aligned} &\alpha_1^{(1)}(t) = 1, \quad \beta_1^{(1)}(t) = c, \\ &\alpha_2^{(1)}(t) = \exp\Big(- \lambda^{(1)} t\Big), \quad \beta_2^{(1)}(t) = \mathbb{P}_m^{(1)}(t). \end{aligned}$$

The second and third factors are associated with the state variables $X_1^{(2)}(t)$ and $X_1^{(3)}(t)$ respectively and the corresponding volatility parameterizations are given by

$$\begin{aligned} &\alpha_1^{(2)}(t) = \exp\Big(- \lambda^{(2)} t\Big), \quad \beta_1^{(2)}(t) = \mathbb{P}_m^{(2)}(t), \\ &\alpha_1^{(3)}(t) = \exp\Big(- \lambda^{(3)} t\Big), \quad \beta_1^{(3)}(t) = \mathbb{P}_m^{(3)}(t). \end{aligned}$$

The parametrization of the volatility can be set up with arbitrary polynomial order. The order influences the flexibility of the model. Empirical tests have shown that linear polynomials ($m = 1$) already cover most of the cases of interest.

We incorporate a constant term in the volatility function of the first factor to increase the accuracy of the model calibration. Due to the additional constant term, the volatility of the first factor is shifted by a constant to a reasonable level. In particular, the calibration results improve for short maturities. The methodology and the results of the calibration are presented in Section 5.

The 3 Factor Exponential Model is represented by four state variables whose dynamics in integral form are given by (6). In the following we will show

the specific form in the 3 Factor Exponential Model. The first factor is determined by the first component of the volatility, $\sigma_1(t,T)$, as specified by (9). The state variables $X_1^{(1)}(t)$ and $X_2^{(1)}(t)$ are associated with the first factor and their dynamics are given by

$$\begin{aligned}
\mathrm{d}X_1^{(1)}(t) &= \left(\sum_{k=1}^{2} V_{1k}^{(1)}(t)\right)\mathrm{d}t + c\mathrm{d}W^{(1)}(t),\\
\mathrm{d}X_2^{(1)}(t) &= \left(-\lambda^{(1)}X_2(t) + \sum_{k=1}^{2} V_{2k}^{(1)}(t)\right)\mathrm{d}t + \left(a_1^{(1)}t + a_0^{(1)}\right)\mathrm{d}W^{(1)}(t).
\end{aligned}$$

The dynamics of the state variables and the representation of the forward rate (3) are based on time-dependent functions $V_{ij}^{(k)}(t)$ defined in (5). In the 3 Factor Exponential Model, these functions are linear combinations of exponential and quadratic functions of the volatility parameters and time. In particular, the functions are deterministic and can be calculated explicitly as is done in Appendix A.1.

The state variables of the second and third factor are associated with the second and third component of the volatility, $\sigma_2(t,T)$ and $\sigma_3(t,T)$, specified in (10) and (11). Their dynamics are given by

$$\begin{aligned}
\mathrm{d}X_1^{(2)}(t) &= \left[-\lambda^{(2)}X_1^{(2)}(t) + V_{11}^{(2)}(t)\right]\mathrm{d}t + \left(a_1^{(2)}t + a_0^{(2)}\right)\mathrm{d}W^{(2)}(t),\\
\mathrm{d}X_1^{(3)}(t) &= \left[-\lambda^{(3)}X_1^{(3)}(t) + V_{11}^{(3)}(t)\right]\mathrm{d}t + \left(a_1^{(3)}t + a_0^{(3)}\right)\mathrm{d}W^{(3)}(t).
\end{aligned}$$

Again, the time dependent functions $V_{ij}^{(k)}(t)$ are deterministic and given in Appendix A.1. For later use, we use the following representation of the state variable $X_1 = X_1^{(1)}$, $X_2 = X_2^{(1)}$, $X_3 = X_1^{(2)}$ and $X_4 = X_1^{(3)}$.

4 Analytical Pricing Formulas

We are able to derive analytical pricing formulas for bonds, caplets and floorlets in the class of Cheyette models. The bond pricing formula found from the definition of the bond price in (1) is combination with the representation of the forward rate in (3). The derivation of the caplet pricing formula is based on Black's formula, and develops further the ideas of Musiela &

Rutkowski (2005), who have worked in the Gaussian HJM model. Henrard (2003) has developed a pricing formula for European swaptions, which is limited to one factor models only. This semi-explicit formula incorporates the general volatility structure of the HJM framework and if we choose the volatility according to the assumptions of the Cheyette models given in (2), we derive a pricing formula for this type of model.

4.1 The Analytical Bond Pricing Formula

The analytical pricing formulas serve as a benchmark for the numerical pricing methods in the following discussion. Since we focus on the 3 Factor Exponential Model in the pricing of interest rate derivatives, we present the analytical pricing formulas for this model. We start by pricing bonds.

4.1.1 Multi-Factor Cheyette Model

Theorem 4.1.
Given $X_t = x = \{X_j^{(k)}\}_{k=1,\dots,M;j=1,\dots,N_k}$ at time t then the price of a zero-coupon bond in the M-factor Cheyette Model is given by

$$B(t,T) = \frac{B(0,T)}{B(0,t)} \exp\Big[-\sum_{k=1}^{M}\sum_{j=1}^{N_k} G_j^{(k)}(t,T) X_j^{(k)}(t) - H(t,T)\Big], \tag{12}$$

with

$$G_j^{(k)}(t,T) = \frac{A_j^{(k)}(T) - A_j^{(k)}(t)}{\alpha_j^{(k)}(t)},$$

$$H(t,T) = \sum_{k=1}^{M}\sum_{i,j=1}^{N_k} \frac{\Big(A_i^{(k)}(T) - A_i^{(k)}(t)\Big)\Big(A_j^{(k)}(T) - A_j^{(k)}(t)\Big)}{2\alpha_i^{(k)}(t)\alpha_j^{(k)}(t)} V_{ij}^{(k)}(t).$$

Proof. Please refer to Appendix D.1 for details. □

4.1.2 The 3 Factor Exponential Model

Lemma 4.2 (Analytical Bond Price in the 3 Factor Exponential Model)**.**
Given $X_t = x = \{X_j^{(k)}\}_{k=1,\dots,M;j=1,\dots,N_k}$ at time t the price of a zero-coupon

bond in the 3 *Factor Exponential Model is given by*

$$B(t,T) = \frac{B(0,T)}{B(0,t)} \exp\Big[-\sum_{k=1}^{3}\sum_{j=1}^{N_k} G_j^{(k)}(t,T) X_j^{(k)}(t) - H(t,T)\Big], \qquad (13)$$

with $N_1 = 2$, $N_2 = 1$, $N_3 = 1$ *and*

$$G_1^{(1)}(t,T) = T - t,$$
$$G_2^{(1)}(t,T) = \frac{1}{\lambda^{(1)}}\Big(1 - \exp(-\lambda^{(1)}(T-t))\Big),$$
$$G_1^{(j)}(t,T) = \frac{1}{\lambda^{(j)}}\Big(1 - \exp(-\lambda^{(j)}(T-t))\Big),\ j = 2,3,$$
$$H(t,T) = \frac{1}{2}(T-t)^2 V_{11}^{(1)}(t) + \frac{(T-t)\Big(\exp(-\lambda^{(1)}t) - \exp(-\lambda^{(1)}T)\Big)}{\lambda^{(1)}\exp(-\lambda^{(1)}t)} V_{12}^{(1)}(t)$$
$$+ \frac{\Big(\exp(-\lambda^{(1)}t) - \exp(-\lambda^{(1)}T)\Big)^2}{2(\lambda^{(1)})^2\exp(-2\lambda^{(1)}t)} V_{22}^{(1)}(t) + \sum_{j=2}^{3} \frac{\Big(\exp(-\lambda^{(j)}t) - \exp(-\lambda^{(j)}T)\Big)^2}{2(\lambda^{(j)})^2\exp(-2\lambda^{(j)}t)} V_{11}^{(j)}(t)$$

Proof.
We use equations (9) - (11) and the bond price formula (12). Also note in this 3 factor case, $M = 3$ and $N_1 = 2$, $N_2 = 1$ and $N_3 = 1$. □

4.2 An Analytical Caplet Pricing Formula

An interest rate cap is an insurance against increasing interest rates and consists of several individual European call options on the variable interest rate. The options are called caplets and the most popular are caplets on the LIBOR rate. The underlying LIBOR rate $R(t, T_C, T_B)$ at time T_C for the period $[T_C, T_B]$ observed at time t can be expressed by bond prices as (see Brace et al. (1997))

$$R(t,T_C,T_B) = \frac{1}{\Delta}\left(\frac{B(t,T_C)}{B(t,T_B)} - 1\right), \qquad (14)$$

with $\Delta = T_B - T_C$ in the corresponding day count convention. The underlying rate of the caplets is fixed at time T_C and the payoff occurs at time T_B. The price at time t of a caplet starting at T_C and ending at T_B ($t \le T_C \le T_B$)

with strike K and nominal amount 1 is given by

$$C(t,T_C,T_B) = \Delta\mathbb{E}\Bigg[\exp\Big(-\int_t^{T_B} r(s)\mathrm{d}s\Big)\Big(R(t,T_C,T_B)-K\Big)^+\Big|\mathfrak{F}_t\Bigg], \quad (15)$$

where $\mathbb{E}[.]$ denotes the expected value under the risk-neutral measure Q. Changing the measure to the T_B-forward measure Q^{T_B} implies that the discount factor can be excluded from the expected value. Applying the change of measure theorem and expressing the LIBOR rate by bond prices, one ends up with

$$\begin{aligned} C(t,T_C,T_B) =& \Delta B(t,T_B)\mathbb{E}^{T_B}\Bigg[\Big(R(t,T_C,T_B)-K\Big)^+\Big|\mathfrak{F}_t\Bigg] \\ =& \Delta B(t,T_B)\mathbb{E}^{T_B}\Bigg[\max\Big(\frac{1}{\Delta}\Big(\frac{B(t,T_C)}{B(t,T_B)}-1\Big)-K,0\Big)\Big|\mathfrak{F}_t\Bigg] \\ =& B(t,T_B)\mathbb{E}^{T_B}\Bigg[\max\Big(\frac{B(t,T_C)}{B(t.T_B)}-(1+K\Delta),0\Big)\Big|\mathfrak{F}_t\Bigg], \end{aligned} \quad (16)$$

where $\mathbb{E}^{T_B}[.]$ denotes the expectation under the T_B forward measure Q^{T_B}. The computation of the expected values requires knowledge of the dynamics of the forward bond price $\frac{B(t,T_C)}{B(t,T_B)}$ under the Q^{T_B} measure.

A first step to this result we recall that Björk, Kabanov & Runggaldier (1997) have given the dynamics of the bond price $B(t,T)$ in the HJM methodology as determined by the SDE

$$\mathrm{d}B(t,T) = B(t,T)\Big(a(t,T)\mathrm{d}t + b(t,T)\cdot\mathrm{d}W_t\Big). \quad (17)$$

The drift $a(t,T)$ and the volatility $b(t,T)$ are determined by the forward rate volatility $\sigma(t,T)$. In the following, we only use the bond price volatility given by

$$b(t,T) = -\int_t^T \sigma(t,u)\mathrm{d}u. \quad (18)$$

The forward rate volatility $\sigma(t,T)$ is an $\mathbb{R}^M$ valued process and determines the bond price dynamics. Musiela & Rutkowski (2005) showed the following result for the class of Markovian HJM Models.

Lemma 4.3.

For any fixed $T > 0$, the process W^T given by the formula

$$W^T(t) = W(t) - \int_0^t b(u,T)\mathrm{d}u, \quad \forall t \in [0,T]$$

follows a standard d-dimensional Brownian Motion under the T-forward measure Q^T, in which W denotes a standard M-dimensional Brownian Motion under the risk-neutral measure Q and $b(t,T)$ denotes a $\mathbb{R}^M$-valued and $\mathcal{F}_t$ measurable function given by the volatility of a zero-coupon bond with maturity T.

Applying this general result, the connection between W^{T_B} and W_t can be reformulated in terms of the dynamics as

$$\mathrm{d}W^{T_B}(t) = \mathrm{d}W(t) - b(t,T)\mathrm{d}t.$$

Thus the dynamics of the forward bond price in the Markovian HJM methodology under the T_B forward measure are given by

$$\mathrm{d}\left(\frac{B(t,T_C)}{B(t,T_B)}\right) = \frac{B(t,T_C)}{B(t,T_B)}\Big(b(t,T_B) - b(t,T_C)\Big) \cdot \mathrm{d}W^{T_B}(t). \tag{19}$$

Musiela and Rutkowski use the fact, that the forward bond price is a martingale under the T_B forward measure. Thus, the volatility of the bond price $b(t,T)$ determines the change of measure from the risk-neutral one to the forward measure.

Notice that $W^{T_B}(t)$ is an M-dimensional vector of independent Brownian Motions with respect to the Q^{T_B} measure and the bond price volatility is thus also a vector. To clarify the notation, we rewrite (19) as

$$\mathrm{d}\left(\frac{B(t,T_C)}{B(t,T_B)}\right) = \frac{B(t,T_C)}{B(t,T_B)} \sum_{k=1}^{M} \Big(b_k(t,T_B) - b_k(t,T_C)\Big)\mathrm{d}W_k^{T_B}(t),$$

where the index k indicates the k-th component of each vector. The forward bond price is a martingale and its instantaneous volatility is deterministic, thus the expected value in (16) can be computed by Black's formula as presented by Hull (2005). Consequently, the pricing formula for a caplet is

given by

$$C(t,T_C,T_B) = B(t,T_C)\mathcal{N}(d_1) - B(t,T_B)(1+K\Delta)\mathcal{N}(d_2), \qquad (20)$$

with

$$\begin{aligned} d_1 =& \frac{\ln\left(\frac{B(t,T_C)}{B(t,T_B)(1+K\Delta)}\right) + \frac{1}{2}v^2(t,T_C,T_B)}{v(t,T_C,T_B)}, \\ d_2 =& \frac{\ln\left(\frac{B(t,T_C)}{B(t,T_B)(1+K\Delta)}\right) - \frac{1}{2}v^2(t,T_C,T_B)}{v(t,T_C,T_B)}, \end{aligned}$$

where

$$v^2(t,T_C,T_B) = \sum_{k=1}^{M} \int_t^{T_C} |b_k(u,T_B) - b_k(u,T_C)|^2 \mathrm{d}u$$

and $\mathcal{N}(\cdot)$ denotes the the cumulative distribution function of the normal distribution.

4.2.1 The 3 Factor Exponential Model

The analytical caplet price in a specific model can be obtained easily by determining the bond price volatility $b(t,T)$ on the basis of the forward rate volatility $\sigma(t,T)$. Consider the forward rate volatility in the 3 Factor Exponential Model defined in equations (9) - (11). The bond price volatilities according to (18) turn out to be

$$\begin{aligned} b_1(t,T) =& \frac{-1}{\lambda^{(1)}}\left[a_0^{(1)} + a_1^{(1)}t - c\lambda^{(1)}t - \exp\left(\lambda^{(1)}(t-T)\right)(a_0^{(1)} + a_1^{(1)}t) + c\lambda^{(1)}T\right] \\ b_2(t,T) =& \frac{a_0^{(2)} + a_1^{(2)}t}{\lambda^{(2)}}\left(\exp\left(\lambda^{(2)}(t-T)\right) - 1\right) \\ b_3(t,T) =& \frac{a_0^{(3)} + a_1^{(3)}t}{\lambda^{(3)}}\left(\exp\left(\lambda^{(3)}(t-T)\right) - 1\right). \end{aligned}$$

The caplet pricing formula is determined by the quantity $v^2(t, T_C, T_B)$ given by

$$v^2(t, T_C, T_B) = \sum_{k=1}^{3} \int_t^{T_C} (b_k(u, T_B) - b_k(u, T_C))^2 \mathrm{d}u.$$

This deterministic function can be computed explicitly and is given in Appendix A.2.

5 Calibration of the 3 Factor Exponential Model

We now turn to the 3 Factor Exponential Model, which incorporates 10 parameters in total. In order to apply the model to pricing derivatives we first require a calibration to the current state of the market. Therefore, we look for values of the parameters that best represent the market. We have calibrated the model to (at-the-money) caps with maturities up to 10 years based on market data provided by Moosmueller and Knauf [1] observed on the 30/4/2010, which can be found in Appendix C. The calibration is based on the pricing formulas by characteristic functions presented by Beyna & Wystup (2011). As shown, the characteristic functions in the general Cheyette Model can be calculated explicitly for an arbitrary number of factors. Thus, the evaluation of a cap turns out to be fast and accurate. The calibration of the model is equivalent to a (global) minimization problem, that minimizes the differences between the Black-Scholes Model (market data) and the 3-factor Cheyette Model. The differences are measured in terms of implied Black-Scholes volatilities. Therefore, we have to transform the prices in the Cheyette Model to implied Black-Scholes volatilities by inverting the pricing formula in the Black-Scholes Model. Beyna & Wystup (2010) have analyzed the calibration of the Cheyette Model to traded products and that the most stable and accurate results are obtained by using the Simulated Annealing algorithm to solve the global minimization problem. This optimization method finds the global minimum with high probability as it supports the feature of leaving local minima with a predefined probability as for example presented by Kirkpatrick, Gelatt & Vecchi (1983). We applied the algorithm to the calibration problem with

[1] Moosmueller and Knauf is a German software and consulting house.

the parameters given in Table 1.

We incorporated 8 caps and the resulting parameters of the minimum are shown in Table 2. The minimal value is given as the mean squared error in implied volatility and has the value $1.77E-06$. The resulting volatility functions are plotted in Figure 1.

Parameter	Value
α	1.1
β	2.2
γ	0.5
'Initial Value'	5.0
Iterations	1000

Table 1: Parameters of the Simulated Annealing Algorithm. 'Initial Value' is the starting point of the Simulated Annealing Algorithm.

	constant	$a_1^{(k)}$	$a_0^{(k)}$	$\lambda^{(k)}$
1. Factor	0.0097	-0.0005	-0.000165	-0.004
2. Factor	N/A	0.000021	-0.000742	-0.430
3. Factor	N/A	0.0000193	0.000701	-0.51

Table 2: Results of the calibration of the 3 Factor Exponential Model to caplets.

6 Quasi Monte Carlo Simulation

In Quasi-Monte Carlo methods the construction of random numbers is in a special way. The algorithm tries to cover the unit hypercube with lowest discrepancy as shown example for Sobol numbers in Figure 2. Dick & Pillichshammer (2010) and Niederreiter (1992) present several methods to create low-discrepancy numbers. Applying these random numbers, the accuracy of the Monte Carlo simulation will increase and the convergence can accelerate up to $O(\frac{1}{n})$ in contrast to $O(\frac{1}{\sqrt{n}})$ in the case of the ordinary Monte Carlo method, where n denotes the number of sample paths.

The implementation of the Quasi-Monte Carlo method that we use is based on Sobol random numbers and follows the ideas of Joe & Kuo (2008) to

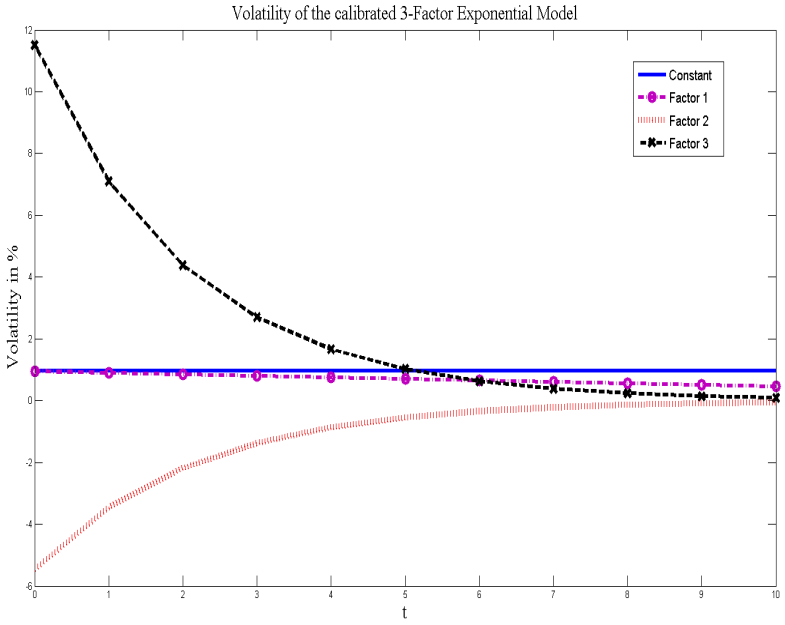


Figure 1: Volatility of the calibrated 3 Factor Exponential Model

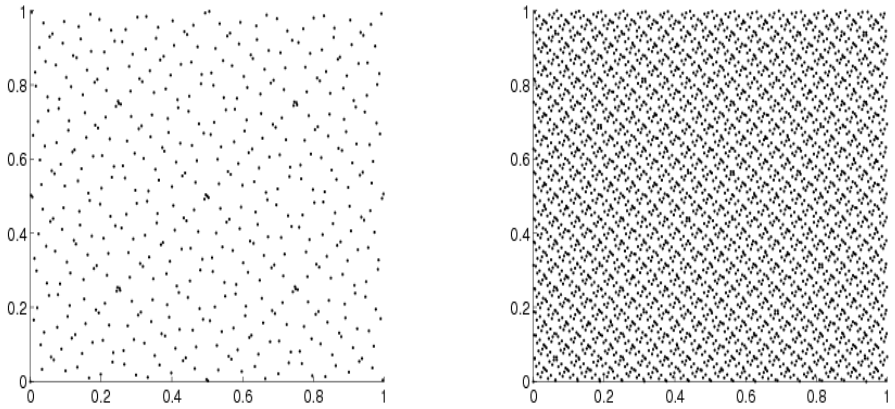
construct Sobol sequences. As in their work, we are able to construct Sobol sequences of up to $21,201$ dimensions. In the case of pricing derivatives, the dimension is given by the number of state variables multiplied by the number of fixing times per path. In the case of a path-independent derivative, the number of fixing times equals one. The dimension increases for path-dependent products.

6.1 Pricing under the forward measure

As we know from Lemma 4.3, the change of measure to the T-forward measure causes a change of the drift of each state variable $X_i(t)$ only. In the following we will apply this result to the 3 Factor Exponential Model. The elements of the bond price volatility $b(t,T)$ are given in Section 4.2.1.

6.1.1 Distribution of the state variables

The construction of sample paths can be done in several different ways as presented in Glasserman (2003). The best results are obtained by exact methods, that are based on the joint distribution of the underlying stochas-



(a) First 512 Points (b) First 4096 Points

Figure 2: A two-dimensional Sobol sequence with first 512 points in the left panel and first 4096 points in the right panel.

tic processes. In contrast to discretization methods, such as the Euler discretization, this method does not cause any discretization error in time. In the setup of the Cheyette models, since the volatility functions are not stochastic the state variables are normally distributed with time-dependent mean and variance.

The following Theorem provides the distribution of the state variables in the 3 Factor Exponential Model under the T-forward measure.

Theorem 6.1. *$X_1(t)$ is normally distributed,*

$$X_1(t) \sim \mathcal{N}\Big(M_1(t), S_1^2(t)\Big),$$

with mean $M_1(t) = X_1(0) + \int_0^t U_1(s,T)\mathrm{d}s$ *and variance* $S_1^2(t) = \int_0^t c^2\mathrm{d}s$*;*

$X_2(t)$ is normally distributed,

$$X_2(t) \sim \mathcal{N}\Big(M_2(t), S_2^2(t)\Big),$$

with mean $M_2(t) = X_2(0)\exp(-\lambda^{(1)}t) + \int_0^t U_2(s,T)\exp(\lambda^{(1)}(s-t))\mathrm{d}s$ *and variance* $S_2^2(t) = \int_0^t \exp\Big(-2\lambda^{(1)}(t-s)\Big)(a_0^{(1)} + a_1^{(1)}s)^2\mathrm{d}s.$

State variables $X_3(t)$ and $X_4(t)$ are also normal distributed, with

$$X_3(t) \sim \mathcal{N}\Big(M_3(t), S_3^2(t)\Big)$$

and

$$X_4(t) \sim \mathcal{N}\Big(M_4(t), S_4^2(t)\Big).$$

The means $M_i(t)$ and the variances $S_i^2(t)$ associated to the state variables $X_i(t)$ for $i = 1, ..., 4$ are presented in equations (28) - (35) in the Appendix D.2.

Further, the time-dependent covariance matrix of $X_1(t)$ to $X_4(t)$, denoted $\Sigma = \Sigma(t)$ is given by

$$\Sigma(t) = \begin{pmatrix} S_1^2(t) & c(\frac{1}{2}a_1^{(1)}t^2 + a_0^{(1)}t) & 0 & 0 \\ c(\frac{1}{2}a_1^{(1)}t^2 + a_0^{(1)}t) & S_2^2(t) & 0 & 0 \\ 0 & 0 & S_3^2(t) & 0 \\ 0 & 0 & 0 & S_4^2(t) \end{pmatrix}.$$

Proof. Please refer to Appendix D.2 for details. □

Remark 6.2.
The covariance is essential to determine the joint distribution of the state variables, and it is needed to generate sample paths. The 3 Factor Exponential Model contains four state variables and is driven by three independent Brownian Motions. The state variables X_1 and X_2 are driven by the same Brownian Motion, thus they are not independent and the covariance does not vanish. In contrast, the other covariance entries vanish, because the state variables are pairwise independent.

7 PDE Valuation

In the following section we present the methodology to price interest rate derivatives in a multi-factor Cheyette Model by using the PDEs that result from the application of the Feynman-Kac Theorem. First, we derive the associated valuation PDE and develop terminal conditions corresponding to the bond and caplet examples. Due to the high dimensionality of the PDE and the resulting high computational effort, we cannot apply ordinary numerical methods like Finite Elements or Finite Differences. Instead we will make

use of sparse grid methods based on Finite Differences. This approach has been applied successfully to high dimensional PDEs by Bungartz & Griebel (2004) and Reisinger (2004). Based on standard techniques, we are able to increase the accuracy by adding additional points to the grid as shown by Chiarella & Kang (2012). Applying this technique, we obtain reliable results with small error.

7.1 Derivation of the Valuation PDE

The dynamics of the forward rate are determined by the state variables $X_i^{(k)}(t)$ which are defined in (6). A multi-factor model with M factors each incorporating N_k summands in the volatility parametrization (2) is driven by $n = \sum_{k=1}^{M} N_k$ state variables. The dynamics of the state variables are specified in (7) and we have in particular that the dynamics form Markov processes. Thus the model dynamics can be expressed by a PDE. The theorem of Feynman-Kac implies the following PDE:

Theorem 7.1 (The Valuation PDE)**.**
Under the volatility structure (2) the expected value $g(t,x)$ of any forward rate dependant contingent claim $\phi : \mathbb{R}^n \to \mathbb{R}$ is given by

$$g(t,x) = \mathbb{E}\left[\exp\Big(-\int_t^T r(\tau)\mathrm{d}\tau \Big)\phi(X_T)\Big| X_t = x \right]$$

and solves the PDE

$$\begin{aligned}\frac{\partial g}{\partial t} &+ \sum_{i=1}^{n}\left(x_i(t)(\partial_t \log \alpha_i(t)) + \sum_{j=1}^{N_k} V_{ij}(t)\right)\frac{\partial g}{\partial x_i} \\ &+ \frac{1}{2}\sum_{i,j=1}^{n}\left[\sigma\sigma^T\right]_{ij}(x)\frac{\partial^2 g}{\partial x_i \partial x_j} - r(t)g \\ &= 0\end{aligned}$$

with terminal condition

$$g(T,x) = \phi(x),$$

where $\sigma = \sigma(t) \in \mathbb{R}^{n\times M}$ denotes the diffusion of the multi-dimensional state

variable X and is defined by

$$\sigma_{ij}(t) = \begin{cases} \beta^{(j)}_{i-Z(j-1)}(t), & Z(j-1) < i \leq Z(j) \\ 0, & otherwise. \end{cases} \tag{21}$$

with $Z(j) = \sum_{k=1}^{j-1} N_k$ and where N_k denotes the number of summands of the volatility of the k-th factor.

Example 7.2.
The definition of the matrix σ is very abstract and to clarify the notation we present as an example the matrix of the 3-Factor Exponential Model defined in Section 3. The diffusion terms of the first factor are given by $\beta^{(1)}_1(t) = c$, $\beta^{(1)}_2(t) = a^{(1)}_1 t + a^{(1)}_0$ and the diffusion terms of the second and third factor by $\beta^{(2)}_1(t) = a^{(2)}_1 t + a^{(2)}_0$ and $\beta^{(3)}_1(t) = a^{(3)}_1 t + a^{(3)}_0$ respectively. The numbers of summands per factor N_k are $N_1 = 2$, $N_2 = 1$ and $N_3 = 1$. Thus the matrix σ is given by

$$\sigma = \begin{pmatrix} c & 0 & 0 \\ a^{(1)}_1 t + a^{(1)}_0 & 0 & 0 \\ 0 & a^{(2)}_1 t + a^{(2)}_0 & 0 \\ 0 & 0 & a^{(3)}_1 t + a^{(3)}_0 \end{pmatrix} \tag{22}$$

and

$$\left[\sigma\sigma^T\right] = \begin{pmatrix} c^2 & c\left(a^{(1)}_1 t + a^{(1)}_0\right) & 0 & 0 \\ c\left(a^{(1)}_1 t + a^{(1)}_0\right) & \left(a^{(1)}_1 t + a^{(1)}_0\right)^2 & 0 & 0 \\ 0 & 0 & \left(a^{(2)}_1 t + a^{(2)}_0\right)^2 & 0 \\ 0 & 0 & 0 & \left(a^{(3)}_1 t + a^{(3)}_0\right)^2 \end{pmatrix}$$

Thus, the matrix σ is a $\left(\sum_{k=1}^{M} N_k\right) \times M$ matrix and the matrix $\left[\sigma\sigma^T\right]$ is a $\left(\sum_{k=1}^{M} N_k\right) \times \left(\sum_{k=1}^{M} N_k\right)$ matrix, where M denotes the number of factors and N_k the number of volatility summands of factor k.

7.2 The 3 Factor Exponential Model

In the case of the 3 Factor Exponential Model, the volatility is specified in (9) - (11). The model is driven by four state variables and the valuation PDE is given in the following Theorem.

Theorem 7.3 (Valuation PDE in case of the 3 Factor Exponential Model)**.** *Under the volatility structure (8) the expected value $g(t,x)$ of any forward rate dependant contingent claim $\phi : \mathbb{R}^4 \to \mathbb{R}$ given by*

$$g(t,x) = \mathbb{E}\left[\exp\Big(-\int_t^T r(\tau)\mathrm{d}\tau\Big)\phi(X_T)\Big| X_t = x\right]$$

satisfies the PDE

$$\begin{aligned}\frac{\partial g(t,x)}{\partial t} &+ \sum_{i=1}^{4} b_i(t,x)\frac{\partial g(t,x)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{4}\Big[\sigma\sigma^T\Big]_{ij}(x)\frac{\partial^2 g(t,x)}{\partial x_i \partial x_j} \\ &- \left(f(0,t) + \sum_{i=1}^{4} x_i(t)\right) g(t,x) \\ &= 0 \end{aligned} \tag{23}$$

with terminal condition $g(T,x) = \phi(x)$. The functions $b_i(t,x)$ are defined as

$$\begin{aligned} b_1(t,x) =& V_{11}^{(1)}(t) + V_{12}^{(1)}(t) \\ b_2(t,x) =& -\lambda^{(1)}x_2(t) + V_{21}^{(1)}(t) + V_{22}^{(1)}(t) \\ b_3(t,x) =& -\lambda^{(2)}x_3(t) + V_{11}^{(2)}(t) \\ b_4(t,x) =& -\lambda^{(3)}x_4(t) + V_{11}^{(3)}(t) \end{aligned}$$

with the definitions of $V_{ij}^{(k)}(t)$ as given in Section 3. The matrix $\sigma(t) \in \mathbb{R}^{4\times 3}$ defined in (21) consists of the diffusion terms of X_i and is given by Equ.(22)*.*

7.3 Terminal Condition

The valuation PDE depends on the dynamics of the model and is the same for any derivative one wishes to price. The characteristic of the specific derivative is incorporated in the terminal condition, which is given by the

payoff of the derivative at maturity as well as the boundary condition.

7.3.1 Bonds

The terminal condition at time T for a zero-coupon bond maturing at time T is given by

$$\phi(X_T) = 1. \tag{24}$$

7.3.2 Caplets

The underlying rate of a caplet $R(T_C, T_B)$ with option lifetime T_C and maturity of the bond T_B is fixed at time T_C ($T_B > T_C$). Thus, the terminal condition is given at time T_C by the terminal payoff. In the case of a caplet, the payoff occurs at time T_B, thus we have to incorporate discounting from T_B to T_C. This factor is not covered by the PDE, because the PDE ends at time T_C, when the underlying rate is fixed. The additional discount factor is given by the bond price $B(T_C, T_B)$, that can be expressed in closed-form by (12). Thus, the terminal condition at time T_C of a caplet with strike rate K is given by

$$\phi(X_{T_C}) = B(T_C, T_B) \max \Big[R(T_C, T_B) - K, 0 \Big] \tag{25}$$

The underlying rate $R(T_C, T_B)$ is defined by

$$R(T_C, T_B) = \frac{1}{T_B - T_C} \left[\exp\Big(\int_{T_C}^{T_B} f(T_C, s) \mathrm{d}s \Big) - 1 \right]. \tag{26}$$

Theorem 7.4 (Terminal Condition for Caplets in the 3 Factor Exponential Model)**.**
We assume the initial forward rate is constant, namely $f(0,T) = f_0, \forall T$. The terminal condition at time T_C of the valuation PDE specified in Theorem 7.3 for pricing caplets with option lifetime T_C, bond maturity T_B and strike rate K is given by

$$\Phi(X_{T_C}) = \exp(-\Delta f_0)\exp\Big[-\sum_{k=1}^{3}\sum_{j=1}^{N_k} G_j^{(k)}(T_C,T_B)X_j^{(k)}(T_C) - H(T_C,T_B)\Big],$$
$$\max\left(\frac{1}{\Delta}\Big(\exp(\Delta f_0)\exp(\sum_{i=1}^{4} K_i(X_i,T_C,T_B))\Big) - K, 0\right),$$

with $G_j^{(k)}(t,T)$ *and* $H(t,T)$ *as defined in Lemma 4.2 and* $K_i(X_i,T_C,T_B)$ *defined in (36) - (39) in the proof.*

Proof. Please refer to Appendix D.3 for details. □

7.4 Numerical Method for PDE Valuation

The valuation PDE derived in Section 7.1 will be solved with the sparse grid combination technique together with a standard Crank Nicolson Finite Difference method with projected successive over-relaxation (PSOR). We present the details of the Finite Difference scheme with PSOR and the Sparse Grid combination technique in Appendix E. We also refer the reader to Chiarella & Kang (2012) for the detailed implementation of a different example.

8 Numerical Results

In Sections 6 and 7, we have derived the theory for pricing interest rate derivatives in a multi-factor Cheyette model using Monte Carlo Simulation and PDEs. In the following we apply these methodologies to the 3 Factor Exponential Model specified in Section 3. In Section 4 we derived analytical pricing formulas for bonds and caplets, which will serve as a benchmark when applicable.

8.1 Bonds

The first test of the methodology is the pricing zero coupon bonds. In the first step we compute the prices of zero coupon bonds today ($t = 0$) with maturities up to 10 years. In the second step, we investigate the evolution of a zero bond with a maturity of 10 years, $B(t,10)$ for $t = 1,3,...,9$. In both cases, we use the analytical solution (12) to verify the results.

T	Analytical Price	MC Price	Error	PDE Price	Error
2	0.904837	0.904839	$1.9330E-6$	0.904837	$3.5848E-8$
4	0.818730	0.818694	$3.6701E-5$	0.818730	$6.3151E-8$
6	0.740818	0.740852	$3.4115E-5$	0.740820	$2.0729E-6$
8	0.670320	0.670335	$1.5462E-5$	0.670336	$1.6586E-5$
10	0.606530	0.603177	$2.1287E-4$	0.606540	$1.0071E-5$

Table 3: Results of pricing bonds $B(0,T)$ with sparse grids (PDE) and Monte Carlo simulation. Both prices are compared to the analytical solution.

Time t	Analytical Price	MC Price	Error	PDE Price	Error
1	0.622177	0.622808	$6.3143E-4$	0.622803	$6.2643E-4$
3	0.679364	0.679345	$1.8854E-5$	0.679284	$7.9854E-5$
5	0.753310	0.753300	$1.0152E-5$	0.753346	$3.5848E-5$
7	0.841417	0.841415	$2.4080E-6$	0.841462	$4.4592E-5$
9	0.945997	0.945996	$1.1581E-6$	0.945975	$2.2159E-5$

Table 4: Results of pricing bonds $B(t,10)$ by Monte Carlo Simulation and PDEs. Both prices are compared to the analytical solution.

The PDE solution is based on the PDE derived in Theorem 7.3 supplemented by the terminal condition (24). We use the standard sparse grid technique with time step size of $\frac{1}{360}$. The Monte Carlo simulation is based on the distribution of the state variables derived in Theorem 6.1.

In the the first step, we price bonds $B(0,T)$ starting today ($t=0$) with different maturities T up to 10 years. The results of the Monte Carlo simulation and the PDE approach are given in Table 3. Furthermore, the results are compared to the analytical solution and the error is quoted. Summarizing, both methods deliver results close to the analytical solution and thus, we can state that both methods perform well.

In the second step, we computed the prices for bonds $B(t,10)$ maturing in ten years and change the starting points t of the bond. The results of the Monte Carlo simulation and of the PDE approach are presented in Table 4. As one can see, the results of both methodologies match well and thus we conclude that both deliver reliable and accurate results.

8.2 Caplets

A further application of the numerical methods is the pricing of caplets, which are more complicated than bonds due to the fact that they are options. Furthermore, the initial condition for the PDE is non-differentiable, which might cause numerical problems in the solution.

The numerical test incorporates caplets with different strikes and maturities up to 6 years. The strike rates vary between 3% (in-the-money, ITM), 5% (at-the-money, ATM) and 7% (out-of-the money, OTM). The corresponding prices are computed by using a Monte Carlo simulation and by using the PDE approach based on sparse grids. Possible errors are identified by comparing the results to the analytical pricing formula presented in Section 4.2.1. The error can be measured in terms of the price differences, but this measure depends on the product specification, e.g. strike and maturity. Therefore, we transform the prices to implied volatilities in the Black-Scholes Model. This transform is one-to-one and generates a standardized measure.

The Monte Carlo simulation computes the prices of the caplets under the forward measure, see Section 6.1, and makes use of the distribution of the state variables derived in Section 6.1.1. Due to the change of measure, the pricing problem is path independent and the valuation is very fast. The state variables are simulated up to the option lifetime T_C when the underlying cap rate is fixed. As the payoff occurs at maturity of the bond T_B ($> T_C$), we have to discount from T_B to zero. The discount factor $B(0, T_B)$ is computed by the analytical bond price (12). The ordinary Monte Carlo simulation contains $5,000,000$ paths and the results are accurate as we will show later.

Furthermore, we have implemented a Quasi-Monte Carlo simulation using Sobol sequences as explained in Section 6. The convergence order increases and thus we do not need to use as many paths as in the ordinary Monte Carlo simulation. In order to get comparable results we use $2^{18} = 262,144$ paths. The CPU time for Quasi-Monte Carlo decreases by a factor 10 to obtain comparable accuracy. The computation of a caplet takes about 2.3 seconds of CPU time. This effect is even intensified, due to the deterministic character of the generator allowing us to reuse the values for all caplets. Thus, the CPU time becomes smaller if we compute several derivatives with the same Sobol sequence.

Table 5 and Table 6 show the results for ATM caplets (strike rate of 5%). The results of the Monte Carlo simulation are sufficiently accurate as

T_C	T_B	Analytical Price	MC Price	Error in Price	Std. Error	Error in implied vol
1.0	2.0	0.004183	0.004183	$1.29E-07$	$2.53E-06$	0.0007%
2.0	3.0	0.005318	0.005316	$1.26E-06$	$3.29E-06$	0.0053%
3.0	4.0	0.006077	0.006078	$1.21E-07$	$3.81E-06$	0.0004%
4.0	5.0	0.006792	0.006796	$4.31E-06$	$4.31E-06$	0.0159%
5.0	6.0	0.007788	0.007786	$1.53E-07$	$4.99E-06$	0.0048%

Table 5: Results of pricing caplets with strike 5% and varying maturities by using an ordinary Monte Carlo simulation with 5.000.000 paths. One can see, that the error in price is always below the standard error. The computation of each caplet takes 24 seconds of CPU time.

T_C	T_B	Analytical Price	QMC Price	Error in Price	Error in implied vol
1.0	2.0	0.004183	0.004181	$1.9722E-06$	0.0110%
2.0	3.0	0.005318	0.005312	$5.3584E-06$	0.0223%
3.0	4.0	0.006077	0.006077	$4.8872E-08$	0.0002%
4.0	5.0	0.006792	0.006777	$1.5391E-05$	0.0177%
5.0	6.0	0.007788	0.007766	$2.1846E-05$	0.0285%

Table 6: Results of pricing caplets with strike 5% and varying maturities by using a Quasi-Monte Carlo simulation with 262144 paths. One can see, that prices are in a comparable range as the corresponding results of an ordinary Monte Carlo simulation as presented in Table 6. The computation of each caplet takes 2.3 seconds of CPU time.

the error in implied volatility is less then 0.0159%. Furthermore, one sees that the pricing error is always below the standard error of the Monte Carlo simulation. This indicates that any desired accuracy could be obtained by increasing the number of paths.

The PDE valuation is based on the PDE shown in Theorem 7.3 combined with the initial condition derived in Section 7.3. We have implemented the sparse grid technique presented in Section 7 and compute solutions for level 1, 2 and 3. We have used a time step of $\frac{1}{360}$ corresponding to a daily fixing of the state variables. The results based on the standard sparse grid methodology were not sufficiently accurate and therefore we implemented the modified sparse grid technique presented in Section E.2.2, that adds points to the sparse grid. The poor results are due to the mixed derivatives in the PDE, that take very high values, for instance of the order of 10^6, in the case of pricing caplets. The mixed derivatives only occur with respect to the first and second state variable, because they are associated to the same Brownian Motion. For this reason, we have to add points just in the first and second dimension of the PDE. This adjustment improves the results dramatically.

Alternatively, we have tried to get rid of the mixed derivatives by transforming the PDE. We have rotated the basis of the spatial domain by the eigenvectors of the matrix $\left[\sigma\sigma^T\right]$ in (23) so that the mixed derivatives vanish. The numerical results were not as good as the results of the modified sparse grid method. A possible explanation might be the time dependence of the rotation matrix, which has to be adjusted on every time step.

However, the modified sparse grid gives accurate results and we present the results for ATM caplets in Table 7. The structure of the sparse grid technique implies that higher accuracy and reliability is obtained by the higher levels. A higher level implies a finer discretization of the spatial domain. Table 7 shows that the results of the standard method (Added Points = 0) is bad as the error in implied volatility takes values between 0.8% and 1.9%. Adding points to the first and second dimension improves the results as shown in Figure 3.

If we add 3 points, the error in implied volatility decreases and the maximal error is 0.0563%. In addition to the results for pricing caplets with strike 5%, we investigated caplets with strike 3% and 7% as well. The results presented in Table 8 are comparable to the results for caplets with strike 5%.

T_C	T_B	Analytical Price	PDE Price Level 1	PDE Price Level 2	PDE Price Level 3	No. of Added Points	Error in Price	Error in implied volatility
1	2	0.004182	0.003694	0.004222	0.004328	0	$1.4503E-04$	0.8092%
2	3	0.005317	0.005131	0.005519	0.005623	0	$3.0542E-04$	1.2741%
3	4	0.006077	0.006094	0.006403	0.006592	0	$5.1420E-04$	1.8528%
4	5	0.006791	0.006832	0.007116	0.007248	0	$4.5571E-04$	1.5053%
5	6	0.007787	0.007473	0.007779	0.008412	0	$6.2383E-04$	1.9621%
1	2	0.004182	0.004170	0.004224	0.004265	1	$8.1906E-04$	0.4569%
2	3	0.005317	0.005388	0.005372	0.005487	1	$1.6922E-04$	0.7057%
3	4	0.006077	0.006160	0.006144	0.006364	1	$2.8591E-04$	1.0295%
4	5	0.006791	0.006713	0.006751	0.007121	1	$3.2946E-04$	1.0878%
5	6	0.007787	0.007154	0.007298	0.007628	1	$1.6007E-04$	0.5137%
1	2	0.004182	0.004179	0.004238	0.004222	2	$3.9508E-05$	0.2204%
2	3	0.005317	0.005269	0.005303	0.005434	2	$1.1632E-04$	0.4850%
3	4	0.006077	0.005946	0.006027	0.006246	2	$1.6820E-04$	0.5154%
4	5	0.006791	0.006415	0.006565	0.006917	2	$1.2509E-04$	0.4127%
5	6	0.007787	0.006779	0.007059	0.007922	2	$1.3419E-04$	0.4211%
1	2	0.004182	0.004151	0.004184	0.004181	3	$2.1948E-06$	0.0123%
2	3	0.005317	0.005236	0.005382	0.005304	3	$1.3513E-05$	0.0563%
3	4	0.006077	0.005889	0.005994	0.006083	3	$5.1603E-06$	0.0186%
4	5	0.006791	0.006323	0.006497	0.006787	3	$5.0917E-06$	0.0168%
5	6	0.007787	0.007206	0.007368	0.007791	3	$2.9307E-06$	0.0092%

Table 7: Results of pricing caplets with strike 5% and varying maturities. The different levels of the PDE price corresponding to the sparse grid approach indicate the refinement of the discretized spatial domain. The accuracy of the results increases as the number of added points in the first and second dimension increases. The errors are computed with respect to the PDE price of level 3.

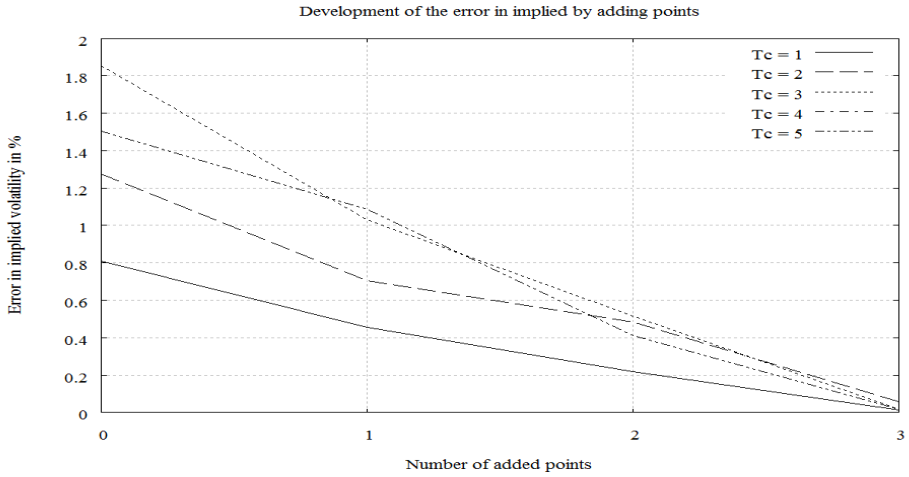


Figure 3: Plot of the error in implied volatility for caplets with strike 5% and varying maturities. The option lifetime T_C changes between 1 and 5 years and each bond matures one year later. The data corresponds to the data in Table 7. One can see, that the error decreases for all caplets if the number of added points is increased. A reasonable level is reached by adding 3 points in the first and second dimension.

Summarizing the results, one can state that both methodologies, Monte Carlo simulation and PDE approach, deliver reliable and sufficiently accurate results. For European derivatives, the Monte Carlo simulation is significantly faster. Table 9 summarizes the results of pricing caplets and compares both methods directly.

An interesting aspect for numerical methods apart from accuracy is the computational effort. The Monte Carlo simulation is very fast, because we know the distribution of the state variables explicitly. Thus, pricing a caplet with $5,000,000$ paths takes 24 seconds of CPU time.[2] In comparison, the computational effort in solving the PDE is significantly higher, in particular if we increase the number of points in the first and second dimension. For example the valuation of the caplet with option lifetime T_C and bond maturity T_B takes about $28,707$ seconds of CPU time [3] for one grid. Although we can compute all grids in parallel [4], one cannot neglect the computational

[2]We used a Windows based PC with Intel Core TM 2 Quad CPU @ 2.66 GHz and 3.49 GB RAM.

[3]We used a Linux based PC with CPU Xeon X5680 @ 3.33 GHz and 24 GB RAM

[4]We used the FEIT and Business High Performance Computing Linux Cluster of the University of Technology, Sydney.

T_C	T_B	Strike	Analytical Price	PDE Price Level 3	No. of Added Points	Error in impl. vol.
1	2	3%	0.019295	0.019255	0	0.5166%
2	3	3%	0.018603	0.018733	0	1.1591%
3	4	3%	0.018062	0.017950	0	0.8175%
4	5	3%	0.017720	0.017508	0	1.3026%
5	6	3%	0.017687	0.017309	0	2.0462%
1	2	3%	0.019295	0.019276	1	0.0280%
2	3	3%	0.018603	0.018550	1	0.4833%
3	4	3%	0.018062	0.017986	1	0.5502%
4	5	3%	0.017720	0.017494	1	1.3939%
5	6	3%	0.017687	0.017274	1	2.2395%
1	2	3%	0.019295	0.019292	3	0.0186%
2	3	3%	0.018603	0.018601	3	0.0172%
3	4	3%	0.018062	0.018060	3	0.0161%
4	5	3%	0.017720	0, 017721	3	0.0105%
5	6	3%	0.017687	0, 017689	3	0.0090%
1	2	7%	0.000108	0.000069	0	1.2263%
2	3	7%	0.000501	0.000416	0	0.7897%
3	4	7%	0.000975	0.000830	0	0.8339%
4	5	7%	0.001547	0.001297	0	1.0991%
5	6	7%	0.002424	0.001900	0	1.8953%
1	2	7%	0.000108	0.000106	1	0.0469%
2	3	7%	0.000501	0.000477	1	0.2192%
3	4	7%	0.000975	0.000894	1	0.4586%
4	5	7%	0.001547	0.001347	1	0.8752%
5	6	7%	0.002424	0.001927	1	0.9923%
1	2	7%	0.000108	0.000108	3	0.0199%
2	3	7%	0.000501	0.000492	3	0.0778%
3	4	7%	0.000975	0.000966	3	0.0474%
4	5	7%	0.001547	0.001545	3	0.0075%
5	6	7%	0.002424	0.002433	3	0.0343%

Table 8: Results of pricing caplets with strike 3% and 7% and varying maturities. The accuracy of the results increases as the number of added points in the first and second dimension increases.

effort. But if we apply the numerical method to exotic products, that are path-dependent, the computational effort does not change. In contrast, the computation time for Monte Carlo simulations will increase dramatically.

8.3 European Swaptions

In addition to pricing bonds and caplets, we incorporate the valuation of swaptions in the numerical test. First we focus on European swaptions and apply the modified sparse grid technique. Thereby we use two additional points and compute prices up to level three according to Section E.2.2. The valuation is based on the 3 Factor Exponential Model with parameters specified in Section 5 quoted in Table 2. Since there are no analytical pricing formulas existing, we benchmark the results of the PDE method to the outcomes of the Monte Carlo simulation. The results of the PDE method are presented in Table 10, which includes a comparison to the corresponding values of the Monte Carlo simulation.

The numerical test incorporates 18 European swaptions with varying option lifetimes T_0 up to three years and swap lifetimes T_N up to five years. Furthermore, we modify the strike rate between 3% (ITM), 5% (ATM) and 7% (OTM). The nominal amount is one and the initial forward rate is assumed to be 5%. The results of the PDE valuation are close to the corresponding outcomes of the Monte Carlo simulation as the price differences are limited by $9.81E-4$. Thus, we conclude that the pricing method is reliable and delivers accurate results. Since we only use two additional points in the sparse grid discretization, the results can even be improved by adding more points as demonstrated for caplets in Section 8.2.

8.4 Bermudan Swaptions

So far, we have applied the PDE valuation technique to plain-vanilla interest rate derivatives only, but the method can be used to price Exotics as well. Therefore we price Bermudan swaptions in the 3 Factor Exponential Model. In contrast to European swaptions, Bermudan swaptions have the right to enter a predefined swap at several fixed dates. All in all, we have valued nine Bermudan swaptions with varying option lifetimes T_0 and strike rates. The additional exercise date is given at half of the option lifetime

T_C	T_B	Strike	Analytical Price	impl. vol. A.P.	MC Price	impl. vol. MC Price	Error in impl. vol	CPU Time	PDE Price	impl. vol. PDE	Error in imp. vol.	CPU Time
1	2	3%	0.019295	47.77%	0.019291	47.78%	0.0142%	24	0.019292	47.75%	0.0186%	9130
2	3	3%	0,018603	36.11%	0.018604	36.11%	0.0074%	24	0.018601	36.10%	0.0172%	17900
3	4	3%	0,018062	32.33%	0.018067	32.22%	0.0039%	24	0.018060	32.32%	0.0161%	26000
4	5	3%	0,017720	31.31%	0.017725	31.32%	0.0173%	24	0.017722	31.32%	0.0105%	38615
5	6	3%	0,017687	32.58%	0.017682	32.57%	0.0099%	24	0.017688	32.59%	0.0090%	60729
1	2	5%	0.004183	23.23%	0.004183	23.23%	0.0007%	24	0.004181	23.21%	0.0123%	9460
2	3	5%	0.005318	21.99%	0.005316	21.98%	0.0053%	24	0.005304	21.93%	0.0563%	18667
3	4	5%	0.006078	21.61%	0.006078	21.61%	0.0004%	24	0.006082	21.63%	0.0186%	28707
4	5	5%	0.006792	22.04%	0.006797	22.05%	0.0159%	24	0.006787	22.02%	0.0168%	39181
5	6	5%	0.007788	23.85%	0.007786	23.84%	0.0048%	24	0.007790	23.86%	0.0092%	62832
1	2	7%	0.000108	17.80%	0.000109	17.81%	0.0067%	24	0.000108	17.82%	0.0199%	9,211
2	3	7%	0.000501	17.49%	0.000501	17.49%	0.0025%	24	0.000501	17.41%	0.0778%	19,236
3	4	7%	0.000975	17.47%	0.000974	17.46%	0.0067%	24	0.000975	17.42%	0.0474%	28,056
4	5	7%	0.001547	18.00%	0.001546	18.00%	0.0045%	24	0.001547	17.99%	0.0075%	39,829
5	6	7%	0.002424	19.65%	0.002424	19.65%	0.0012%	24	0.002424	19.65%	0.0343%	49,895

Table 9: Results of pricing caplets with varying strikes (ITM, ATM, OTM) with the PDE approach and the Monte Carlo Simulation. Both prices are compared to the analytical solution. The error is quoted in terms of price differences and in terms of differences in implied volatility. The CPU time is quoted in seconds.

T_0	T_N	Strike	PDE Price	CPU Time	MC Price	CPU Time	Pricing Differences
1	3	3%	0.054195	71	0.054157	59	$3.82E-5$
1	3	5%	0.011791	71	0.011237	59	$5.54E-4$
1	3	7%	0.000351	71	0.000262	59	$8.83E-5$
2	3	3%	0.052727	128	0.052563	59	$1.64E-4$
2	3	5%	0.015154	119	0.014953	59	$2.00E-4$
2	3	7%	0.001629	127	0.001438	59	$1.91E-4$
3	3	3%	0.051839	213	0.051666	59	$1.73E-4$
3	3	5%	0.018482	205	0.018030	59	$4.52E-4$
3	3	7%	0.003232	204	0.003241	59	$9.25E-6$
1	5	3%	0.086456	275	0.086246	72	$2.10E-4$
1	5	5%	0.019440	274	0.019403	72	$3.67E-5$
1	5	7%	0.000755	269	0.000686	72	$6.92E-5$
2	5	3%	0.084957	378	0.085011	72	$5.46E-5$
2	5	5%	0.026822	347	0.027416	72	$5.94E-4$
2	5	7%	0.003811	351	0.003876	72	$6.54E-4$
3	5	3%	0.085963	391	0.086944	72	$9.81E-4$
3	5	5%	0.035456	386	0.036260	72	$8.04E-4$
3	5	7%	0.009112	395	0.009850	72	$7.38E-4$

Table 10: Results of pricing European swaptions with strike rate of 3%, 5% and 7% and varying option and swap lifetimes. The modified sparse grid technique uses two additional point and a computational level of three. The results of the Monte Carlo simulation are used as a benchmark.

$T_{Ex} = \frac{1}{2}T_0$. The results are presented in Table 11 and benchmarked to outcomes of the Monte Carlo simulation. The pricing difference is $6.45E-4$ on average and is limited by $1.65E-3$. The results are based on the modified sparse grid technique with two additional points and a computation level of three. The results appear rather accurate and can even be improved by adding more points or increasing the number of computation levels. The pricing difference between European and Bermudan swaptions is the value of the additional exercise right. In Table 12 we compare both prices and highlight the price differences.

As already indicated, the computation time is more or less constant for increase in the number of exercise dates. Since the computational effort in the case of the Monte Carlo simulation increases exponentially in the number of exercise dates the PDE approach should be preferred for Bermudan swap-

T_0	T_N	T_{Ex}	Strike	PDE Price	CPU Time	MC Price	CPU Time	Price Difference
1	3	0.5	3%	0.058531	531	0.059184	2,364	$6.54E-4$
1	3	0.5	5%	0.012930	524	0.014581	2,355	$1.65E-3$
1	3	0.5	7%	0.001308	517	0.001360	2,363	$5.20E-5$
2	3	0.5	3%	0.056269	447	0.057178	2,377	$9.09E-4$
2	3	0.5	5%	0.018467	421	0.018484	2,366	$1.72E-5$
2	3	0.5	7%	0.002633	433	0.003981	2,371	$1.35E-3$
3	3	0.5	3%	0.055458	359	0.055185	2,366	$2.74E-4$
3	3	0.5	5%	0.022159	356	0.021665	2,376	$4.94E-4$
3	3	0.5	7%	0.006125	371	0.006530	2,367	$4.05E-4$

Table 11: Results of pricing Bermudan swaptions with strike rate of 3%, 5% and 7% and varying option and swap lifetimes. The additional exercise date T_{Ex} occurs at half of the option lifetime. The modified sparse grid technique uses two additional point and a computational level of three. The results of the Monte Carlo simulation are used as a benchmark.

tions. Increasing the number of exercise dates delivers an approximation of American swaptions as the price converges. To demonstrate this fact numerically, we price a Bermudan swaption and add exercise dates continuously. The exercise dates are given by

$$T_{Ex}^{(j)} = \frac{1}{12} j, \quad \text{for } j = 1, \ldots, N_{Ex}. \tag{27}$$

The number of exercise dates N_{Ex} rises up to 23, which corresponds to a monthly exercise right. The pricing results are illustrated in Figure 4 and one sees, that the price seems to converge at a price level of 0.01667699 for increasing number of exercise dates.

9 Conclusion

We have investigated the pricing of interest rate derivatives in the Cheyette Model by solving the corresponding PDE using a sparse grid approach and by using Monte Carlo simulations. It turns out, that the standard sparse grid technique does not deliver accurate results in case of the 3 Factor Exponential Model for caplets since the mixed derivatives in the first and second dimension take high values. Applying the modified sparse grid technique by adding a suitable number of points to a grid in the first and second dimen-

T_0	T_N	Strike	European Price	T_{Ex}	Bermudan Price	Difference
1.0	3.0	3%	0.054195	0.5	0.058531	0.004336
1.0	3.0	5%	0.011791	0.5	0.012930	0.001139
1.0	3.0	7%	0.000351	0.5	0.001308	0.000958
2.0	3.0	3%	0.052727	1.0	0.056269	0.003542
2.0	3.0	5%	0.015154	1.0	0.018467	0.003313
2.0	3.0	7%	0.001629	1.0	0.002633	0.001004
3.0	3.0	3%	0.051839	1.5	0.055458	0.003620
3.0	3.0	5%	0.018482	1.5	0.022159	0.003677
3.0	3.0	7%	0.003232	1.5	0.006125	0.002893

Table 12: Comparison of the prices between European and Bermudan swaption based on PDE valuation using the modified sparse grid technique. The price difference is the value of the additional exercise right at time T_{Ex}.

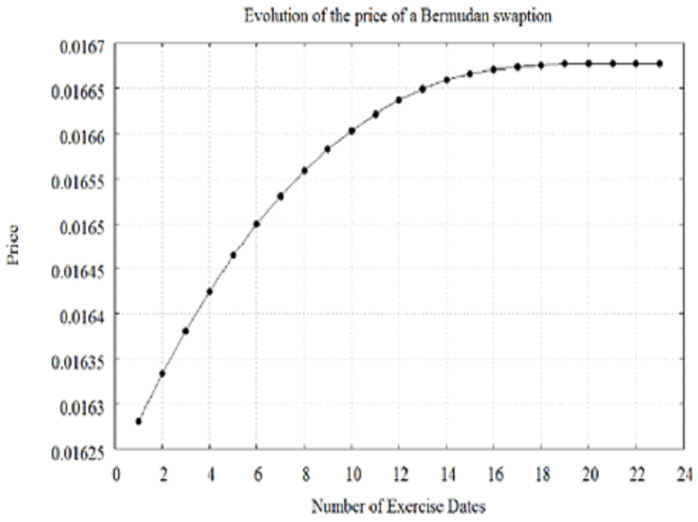


Figure 4: Evolution of the price of a Bermudan swaption for increasing number of exercise dates. The Bermudan swaption has option maturity $T_0 = 2$, swap lifetime $T_n = 3$ and strike $K = 3\%$. The exercise dates are given by (27) for increasing number N_{Ex}.

sion, the PDE can be solved with sufficient accuracy. The mixed derivatives occur just in the first and second dimension, because both state variables are driven by the same Brownian Motion. Thus it is sufficient to add points in these dimensions only. The combination technique of the sparse grid approach supports the computation on several CPUs parallel and consequently, the PDE can be solved in a efficient manner. In addition to the PDE ap-

proach we have studied the pricing by Monte Carlo simulation. In the case of the Cheyette Models, the distribution of the state variables turns out to be normal with time dependent mean and variance. However, as we know the distribution explicitly we can apply an exact simulation technique to generate the sample paths. Furthermore, we can modify the pricing formula by changing the numéraire and use a forward measure. Thus, pricing for example a caplet is no longer path dependant. Combining both improvements, the Monte Carlo simulation turns out to be fast and accurate.

At this point, we have used analytical pricing formulas as a benchmark and we can state, that both methodologies work reliably and accurately for bonds and caplets.

In future research, we will focus on swaptions, European and Bermudan style, and further exotic interest rate derivatives. In these cases, no analytical pricing formula in the Cheyette Model is known so far and thus, we can apply the Monte Carlo simulation as a benchmark for the sparse grid approach.

References

Beyna, I. & Wystup, U. (2010), 'On the calibration of the Cheyette Interest Rate Model', *Frankfurt School of Finance and Management- Centre for Practical Quantitative Finance Working Paper Series* **25**.

Beyna, I. & Wystup, U. (2011), 'Characteristic Functions in the Cheyette Interest Rate Model', *Frankfurt School of Finance and Management-Centre for Practical Quantitative Finance Working Paper Series* **28**.

Bhar, R. & Chiarella, C. (1997), 'Transformation of Heath-Jarrow-Morton Models to Markovian Systems', *European Journal of Finance* **3**, 1–26.

Björk, T., Kabanov, Y. & Runggaldier, W. (1997), 'Bond Market Structure in the Presence of Marked Point Processes', *Mathematical Finance* **7**, 211–239.

Björk, T. & Svensson, L. (2001), 'On the Existence of Finite Dimensional Realizations for Nonlinear Forward Rate Models', *Mathematical Finance* **11**(2), 205–243.

Brace, A., Gatarek, D. & Musiela, M. (1997), 'The market model of interest rate dynamics', *Mathematical Finance* **7**(2), 127–154.

Bungartz, H.-J. & Griebel, M. (2004), 'Sparse grids', *Acta Numerica* **13**, 1–123.

Cheyette, O. (1992), 'Term structure dynamics and mortgage valuation', *The Journal of Fixed Income* **1**(4), 28–41.

Cheyette, O. (1996), 'Markov Representation of the Heath-Jarrow-Morton Model', *SSRN eLibrary* .

Chiarella, C. & Kang, B. (2012), 'The Evaluation of American Compound Option Prices under Stochastic Volatility and Stochastic Interest Rates', *To Appear in Journal of Computational Finance* .

Chiarella, C. & Kwon, O. (2000), 'A Complete Markovian Stochastic Volatility Model in the HJM Framework', *Asia-Pacific Financial Markets* **7**(4), 293–304.

Dick, J. & Pillichshammer, F. (2010), *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press.

Glasserman, P. (2003), *Monte Carlo Methods in Financial Engineering*, 1st edn, Springer.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation', *Econometrica* **60**(1), 77–105.

Henrard, M. (2003), 'Explicit Bond Option and Swaption Formula in Heath-Jarrow-Morton One Factor Model', *International Journal of Theoretical and Applied Finance* **6**, 57–72.

Ho, T. & Lee, S.-B. (1986), 'Term Structure Movements and Pricing Interest Rate Contingent Claims', *Journal of Finance* **41**, 1011–1029.

Hull, J. C. (2005), *Options, Futures, and Other Derivatives*, 5 edn, Financial Times.

Hull, J. & White, A. (1990), 'Pricing interest-rate derivative securities', *The Review of Financial Studies* **3**(4), 573–592.

Joe, S. & Kuo, F. Y. (2008), 'Constructing Sobol sequences with better two-dimensional projections', *SIAM Journal on Scientific Computing* **30**, 2635–2654.

Kirkpatrick, S., Gelatt, C. & Vecchi, M. (1983), 'Optimization by Simulated Annealing', *Science* **220**(4598).

Musiela, M. & Rutkowski, M. (2005), *Martingale Methods in Financial Modelling*, Vol. 2nd, Springer Berlin Heidelberg.

Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Society for Industrial and Applied Mathematics.

Reisinger, C. (2004), Numerische Methoden f¨ur hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben, PhD thesis, Universität Heidelberg.

Reisinger, C. (2008), Analysis of Linear Difference Schemes in the Sparse Grid Combination Technique, Technical report, Mathematical Institute Oxford, UK.

Reisinger, C. & Wittum, G. (2007), 'Efficient Hierarchical Approximation of High-Dimensional Option Pricing Problems', *SIAM Journal of Scientific Computing* **29**(1), 440–458.

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

A Additional Calculus in the 3 Factor Exponential Model

A.1 State Variables

The 3 Factor Exponential Model is represented by four state variables. According to (6) they are given by

$$
\begin{aligned}
X_1^{(1)}(t) &= \int_0^t c\mathrm{d}W^{(1)}(s) + \int_0^t c\left(\sum_{k=1}^{2} \frac{A_k^{(1)} - A_k^{(1)}(s)}{\alpha_k(s)}\beta_k(s)\right)\mathrm{d}s \\
&= \int_0^t c\mathrm{d}W^{(1)}(s) \\
&\quad + c\int_0^t \left[c(t-s) + \frac{(a_1^{(1)}s + a_0^{(1)})\Big[\exp(-\lambda^{(1)}s) - \exp(-\lambda^{(1)}t)\Big]}{\lambda^{(1)}\exp(-\lambda^{(1)}s)}\right]\mathrm{d}s \\
X_2^{(1)}(t) &= \int_0^t \exp[-\lambda^{(1)}(t-s)]\Big(a_1^{(1)}s + a_0^{(1)}\Big)\mathrm{d}W^{(1)}(s) \\
&\quad + \int_0^t \exp[-\lambda^{(1)}(t-s)]\Big(a_1^{(1)}s + a_0^{(1)}\Big) \\
&\quad \left[c(t-s) + \frac{(a_1^{(1)}s + a_0^{(1)})\Big[\exp(-\lambda^{(1)}s) - \exp(-\lambda^{(1)}t)\Big]}{\lambda^{(1)}\exp(-\lambda^{(1)}s)}\right]\mathrm{d}s \\
X_1^{(2)}(t) &= \int_0^t \exp\Big(-\lambda^{(2)}(t-s)\Big)\Big(a_1^{(2)}s + a_0^{(2)}\Big)\mathrm{d}W^{(2)}(s) \\
&\quad + \frac{1}{\lambda^{(2)}}\int_0^t \left[\exp\Big(-\lambda^{(2)}(t-2s)\Big)\Big(a_1^{(2)}s + a_0^{(2)}\Big)^2\right. \\
&\quad \left.\Big(\exp(-\lambda^{(2)}s) - \exp(-\lambda^{(2)}t)\Big)\right]\mathrm{d}s
\end{aligned}
$$

$$X_1^{(3)}(t) = \int_0^t \exp\Big(-\lambda^{(3)}(t-s)\Big)\Big(a_1^{(3)}s + a_0^{(3)}\Big)\mathrm{d}W^{(3)}(s)$$
$$+ \frac{1}{\lambda^{(3)}} \int_0^t \Bigg[\exp\Big(-\lambda^{(3)}(t-2s)\Big)\Big(a_1^{(3)}s + a_0^{(3)}\Big)^2$$
$$\Big(\exp(-\lambda^{(3)}s) - \exp(-\lambda^{(3)}t)\Big)\Bigg]\mathrm{d}s$$

The dynamics of the state variables and the representation of the forward rate is based on time-dependent function $V_{ij}^{(k)}$ as defined in (5). In the 3 Factor Exponential Model, these functions are combinations of exponential and linear functions in time and the volatility parameters:

$$V_{11}^{(1)}(t) = c^2 t,$$
$$V_{12}^{(1)}(t) = V_{21}^{(1)}(t) = \frac{c}{(\lambda^{(1)})^2}\Bigg[-a_1^{(1)} + a_0^{(1)}\lambda^{(1)}$$
$$+ \exp(-\lambda^{(1)}t)(a_1^{(1)} - a_0^{(1)}\lambda^{(1)}) + a_1^{(1)}\lambda^{(1)}t\Bigg],$$
$$V_{22}^{(1)}(t) = \frac{1}{4(\lambda^{(1)})^3}\Bigg[2(a_0^{(1)})^2(\lambda^{(1)})^2 + 2a_0^{(1)}a_1^{(1)}\lambda^{(1)}(-1 + 2\lambda^{(1)}t)$$
$$- \exp(-2\lambda^{(1)}t)\Bigg((a_1^{(1)})^2 - 2a_0^{(1)}a_1^{(1)}\lambda^{(1)} + 2(a_0^{(1)})^2(\lambda^{(1)})^2\Bigg)$$
$$+ (a_1^{(1)})^2\Bigg(1 + 2\lambda^{(1)}t(-1 + \lambda^{(1)}t)\Bigg)\Bigg]$$
$$V_{11}^{(2)}(t) = \frac{1}{4(\lambda^{(2)})^3}\Bigg[2(a_0^{(2)})^2(\lambda^{(2)})^2 + 2a_0^{(2)}a_1^{(2)}\lambda^{(2)}(-1 + 2\lambda^{(2)}t)$$
$$- \exp(-2\lambda^{(2)}t)\Bigg((a_1^{(2)})^2 - 2a_0^{(2)}a_1^{(2)}\lambda^{(2)} + 2(a_0^{(2)})^2(\lambda^{(2)})^2\Bigg)$$
$$+ (a_1^{(2)})^2\Bigg(1 + 2\lambda^{(2)}t(-1 + \lambda^{(2)}t)\Bigg)\Bigg]$$
$$V_{11}^{(3)}(t) = \frac{1}{4(\lambda^{(3)})^3}\Bigg[2(a_0^{(3)})^2(\lambda^{(3)})^2 + 2a_0^{(3)}a_1^{(3)}\lambda^{(3)}(-1 + 2\lambda^{(3)}t)$$

$$- \exp(-2\lambda^{(3)}t)\Bigg((a_1^{(3)})^2 - 2a_0^{(3)}a_1^{(3)}\lambda^{(3)} + 2(a_0^{(3)})^2(\lambda^{(3)})^2\Bigg)$$
$$+ (a_1^{(3)})^2\Bigg(1 + 2\lambda^{(3)}t(-1 + \lambda^{(3)}t)\Bigg)\Bigg]$$

A.2 Caplet Pricing

The analytical pricing formula for caplets (20) incorporates mainly bond prices given by (12) and the bond price volatility $b(t,T)$. The caplet pricing formula is determined by the quantity

$$v^2(t,T_C,T_B) = \sum_{k=1}^{M} \int_t^{T_C} |b_k(u,T_B) - b_k(u,T_C)|^2 \mathrm{d}u.$$

The bond price volatility is specified in Section 4.2.1. Using the short-hand notation $U_k(t,T_C,T_B) := \int_t^{T_C} [b_k(u,T_B) - b_k(u,T_C)]^2 \mathrm{d}u$, $v^2(t,T_C,T_B)$ can be written as

$$v^2(t,T_C,T_B) = \sum_{k=1}^{M} U_k(t,T_C,T_B)$$

with

$$U_1(t,T_C,T_B) = \frac{-\exp(\lambda^{(1)}t)\Big(\exp(-\lambda^{(1)}T_B) - \exp(-\lambda^{(1)}T_C)\Big)^2}{4(\lambda^{(1)})^5}$$
$$\Bigg[2(a_0^{(1)})^2(\lambda^{(1)})^2 + 2a_0^{(1)}a_1^{(1)}\lambda^{(1)}(-1 + 2\lambda^{(1)}t) + (a_1^{(1)})^2\Big(1 + 2\lambda^{(1)}t(-1$$
$$+ \lambda^{(1)}t)\Big)\Bigg] - \frac{2c}{(\lambda^{(1)})^3}\exp\Big(-\lambda^{(1)}(T_B + T_C - t)\Big)\Big(\exp(\lambda^{(1)}T_B)$$
$$- \exp(\lambda^{(1)}TC)\Big)\Big(a_0^{(1)}\lambda^{(1)} + a_1^{(1)}(-1 + \lambda^{(1)}t)\Big)(T_B - T_C) - c^2t(T_B - T_C)^2$$
$$+ c^2(T_B - T_C)^2T_C + \frac{2c}{(\lambda^{(1)})^3}\exp(-\lambda^{(1)}T_B)\Big(\exp(\lambda^{(1)}T_B) - \exp(\lambda^{(1)}T_C)\Big)$$
$$(T_B - T_C)\Big(a_0^{(1)}\lambda^{(1)} + a_1^{(1)}(-1 + \lambda^{(1)}T_C)\Big) + \frac{1}{4(\lambda^{(1)})^5}\exp(-2\lambda^{(1)}T_B)$$

$$\Big(\exp(\lambda^{(1)}T_B) - \exp(\lambda^{(1)}T_C)\Big)^2 \Bigg[2(a_0^{(1)})^2(\lambda^{(1)})^2 + 2a_0^{(1)}a_1^{(1)}\lambda^{(1)}(-1 + 2\lambda^{(1)}T_C) + (a_1^{(1)})^2\Big(1 + 2\lambda^{(1)}T_C(-1 + \lambda^{(1)}T_C)\Big)\Bigg]$$

$$U_2(t,T_C,T_B) = \frac{\Big(\exp(-\lambda^{(2)}T_B) - \exp(-\lambda^{(2)}T_C)\Big)^2}{4(\lambda^{(2)})^5}\Bigg[-\exp(2\lambda^{(2)}t)\Bigg(2(a_0^{(2)})^2(\lambda^{(2)})^2 + 2a_0^{(2)}a_1^{(2)}\lambda^{(2)}(-1 + 2\lambda^{(2)}t) + (a_1^{(2)})^2(1 + 2\lambda^{(2)}t(-1 + \lambda^{(2)}t))\Bigg) + \exp(2\lambda^{(2)}T_C)\Bigg(2(a_0^{(2)})^2(\lambda^{(2)})^2 + 2a_0^{(2)}a_1^{(2)}\lambda^{(2)}(-1 + 2\lambda^{(2)}T_C) + (a_1^{(2)})^2(1 + 2\lambda^{(2)}T_C(-1 + \lambda^{(2)}T_C))\Bigg)\Bigg]$$

$$U_3(t,T_C,T_B) = \frac{\Big(\exp(-\lambda^{(3)}T_B) - \exp(-\lambda^{(3)}T_C)\Big)^2}{4(\lambda^{(3)})^5}\Bigg[-\exp(2\lambda^{(3)}t)\Bigg(2(a_0^{(3)})^2(\lambda^{(3)})^2 + 2a_0^{(3)}a_1^{(3)}\lambda^{(3)}(-1 + 2\lambda^{(3)}t) + (a_1^{(3)})^2(1 + 2\lambda^{(3)}t(-1 + \lambda^{(3)}t))\Bigg) + \exp(2\lambda^{(3)}T_C)\Bigg(2(a_0^{(3)})^2(\lambda^{(3)})^2 + 2a_0^{(3)}a_1^{(3)}\lambda^{(3)}(-1 + 2\lambda^{(3)}T_C) + (a_1^{(3)})^2(1 + 2\lambda^{(3)}T_C(-1 + \lambda^{(3)}T_C))\Bigg)\Bigg]$$

B Remarks on the Model

The framework of Cheyette (1996) is not equivalent to the framework used by Björk & Svensson (2001) and Chiarella & Kwon (2000). These authors assume a volatility parametrization (1 Factor Model)

$$\sigma(t,T) = \sigma_0 + \mathbb{P}_m(T-t)\exp(-\lambda(T-t))$$

where $\mathbb{P}_m(\tau) = a_0 + a_1\tau + ... + a_m\tau^m$. In contrast, the corresponding volatility parametrization of Cheyette and used throughout this work is assumed to be

$$\overline{\sigma}(t,T) = \sigma_0 + \mathbb{P}_m(t)\exp(-\lambda(T-t)).$$

The only difference is that the polynomial depends on time t only instead on the time to maturity $T-t$ in case of Björk & Svensson (2001) and Chiarella & Kwon (2000). In the general HJM framework the forward rate is given by

$$f(t,T) = f(0,t) + \int_0^t \sigma(u,t)\left(\int_u^t \sigma(u,y)\mathrm{d}y\right)\mathrm{d}u + \int_0^t \sigma(u,t)\mathrm{d}W(u).$$

Using the short-hand notation

$$L(t,T) = \int_0^t \sigma(u,t)\left(\int_u^t \sigma(u,y)\mathrm{d}y\right)\mathrm{d}u$$

the dynamics are given by

$$\mathrm{d}f(t,T) = \left[\frac{\partial}{\partial t}f(0,t) + \frac{\partial}{\partial t}L(t,T) + \int_0^t \frac{\partial}{\partial t}\sigma(u,t)\mathrm{d}W(u)\right]\mathrm{d}t + \sigma(t,t)\mathrm{d}W(t).$$

At this point one can see the differences between the two approaches occurs in the stochastic integral $\int_0^t \frac{\partial}{\partial t}\sigma(u,t)\mathrm{d}W(u)$. In the case of Cheyette the derivation with respect to the second argument is given by

$$\begin{aligned}\frac{\partial}{\partial t}\overline{\sigma}(u,t) =& \frac{\partial}{\partial t}\left(\sigma_0 + \mathbb{P}_m(u)\exp(-\lambda(t-u))\right)\\ =& -\mathbb{P}_m(u)\lambda\exp(-\lambda(t-u)).\end{aligned}$$

In contrast, this expression becomes more complicated in the framework of Björk & Svensson (2001) and Chiarella & Kwon (2000):

$$\begin{aligned}\frac{\partial}{\partial t}\sigma(u,t) =& \frac{\partial}{\partial t}\left(\sigma_0 + \mathbb{P}_m(t-u)\exp(-\lambda(t-u))\right)\\ =& \left(\frac{\partial}{\partial t}\mathbb{P}_m(t-u)\right)\exp(-\lambda(t-u)) - \mathbb{P}_m(t-u)\lambda\exp(-\lambda(t-u))\\ =& \left(a_1 + 2a_2(t-u) + ... + ma_m(t-u)^{n-1}\right)\exp(-\lambda(t-u))\\ & - \mathbb{P}_m(t-u)\lambda\exp(-\lambda(t-u)).\end{aligned}$$

Consequently, the dynamics of both frameworks differ and this leads to different state variables representing the model. In particular, the number of state

variables in the framework of Björk & Svensson (2001) and Chiarella & Kwon (2000) depends on the polynomial order m in the volatility parametrization. In contrast, the approach by Cheyette includes an arbitrary polynomial without requiring more state variables.

C The Market Data Used

The calibration of the 3 Factor Exponential Model is based on market data observed on 30/4/2010 and provided by Moosmüller & Knauf. We have calibrated the model to ATM caplets with maturities up to 10 years. The caplet volatilities have been calculated by volatility stripping based on the following volatilities for caps with strike 5% and a tenor of 6 months.

Maturity	Volatility in %	Maturity	Volatility in %
1Y	66.9	6Y	23.8
18M	63.7	7Y	21.7
2Y	56.3	8Y	20.2
3Y	38.6	9Y	19.1
4Y	31.7	10Y	18.3
5Y	27.0		

Table 13: Cap volatilities with strike 5% and a tenor of 6 months observed on 30/4/2010.

D Proofs

D.1 Proof of Theorem 4.1

Proof.
The forward rate in the multi-factor model (M factors) is given by (3). Using this definition, the zero coupon bond can be written as

$$\begin{aligned} B(t,T) &= \exp\left[-\int_t^T f(t,\tau)\mathrm{d}\tau\right] \\ &= \exp\left[-\int_t^T f(0,\tau)\mathrm{d}\tau\right]\exp\left[-\int_t^T \sum_{k=1}^{M}\left(\sum_{j=1}^{N_k}\frac{\alpha_j^{(k)}(\tau)}{\alpha_j^{(k)}(t)}\left[X_j^{(k)}(t)\right.\right.\right. \end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=1}^{N_k}\frac{A_i^{(k)}(\tau)-A_i^{(k)}(t)}{\alpha_i^{(k)}(t)}V_{ij}^{(k)}(t)\Big]\Bigg)\mathrm{d}\tau\Bigg]\\
=&\frac{B(0,T)}{B(0,t)}\exp\Bigg[-\int_t^T\Bigg(\sum_{k=1}^{M}\sum_{j=1}^{N_k}\frac{\alpha_j^{(k)}(\tau)}{\alpha_j^{(k)}(t)}X_j^{(k)}(t)\Bigg)\mathrm{d}\tau\\
&-\int_t^T\Bigg(\sum_{k=1}^{M}\sum_{i,j=1}^{N_k}\frac{\alpha_j^{(k)}(\tau)}{\alpha_j^{(k)}(t)}\frac{A_i^{(k)}(\tau)-A_i^{(k)}(t)}{\alpha_i^{(k)}(t)}V_{ij}^{(k)}(t)\Bigg)\mathrm{d}\tau\Bigg]
\end{aligned}
$$

Using $A_i^{(k)}(t)=\int_0^t\alpha_i^{(k)}(s)\mathrm{d}s$ and $A_i^{(k)}(\tau)-A_i^{(k)}(t)=\int_t^\tau\alpha_i^{(k)}(s)\mathrm{d}s$ one obtains

$$
\begin{aligned}
B(t,T)=&\frac{B(0,T)}{B(0,t)}\exp\Bigg[-\int_t^T\Bigg(\sum_{k=1}^{M}\sum_{j=1}^{N_k}\frac{\alpha_j^{(k)}(\tau)}{\alpha_j^{(k)}(t)}X_j^{(k)}(t)\Bigg)\mathrm{d}\tau\\
&-\int_t^T\Bigg(\sum_{k=1}^{M}\sum_{i,j=1}^{N_k}\frac{\alpha_j^{(k)}(\tau)}{\alpha_j^{(k)}(t)}\Big(\int_t^\tau\frac{\alpha_i^{(k)}(s)}{\alpha_i^{(k)}(t)}\mathrm{d}s\Big)V_{ij}^{(k)}(t)\Bigg)\mathrm{d}\tau\Bigg]
\end{aligned}
$$

Using integration by parts:

$$
\begin{aligned}
&\sum_{k=1}^{M}\sum_{i,j=1}^{N_k}\int_t^T\frac{\alpha_j^{(k)}(\tau)}{\alpha_j^{(k)}(t)}\Big(\int_t^\tau\frac{\alpha_i^{(k)}(s)}{\alpha_i^{(k)}(t)}\mathrm{d}s\Big)\mathrm{d}\tau\\
&\qquad=\frac{1}{2}\sum_{k=1}^{M}\sum_{i,j=1}^{N_k}\frac{\Big(A_j^{(k)}(T)-A_j^{(k)}(t)\Big)\Big(A_i^{(k)}(T)-A_i^{(k)}(t)\Big)}{\alpha_i^{(k)}(t)\alpha_j^{(k)}(t)}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
B(t,T)=&\frac{B(0,T)}{B(0,t)}\exp\Bigg[-\int_t^T\Bigg(\sum_{k=1}^{M}\sum_{j=1}^{N_k}\frac{\alpha_j^{(k)}(\tau)}{\alpha_j^{(k)}(t)}X_j^{(k)}(t)\Bigg)\mathrm{d}\tau\\
&-\sum_{k=1}^{M}\sum_{i,j=1}^{N_k}\frac{\Big(A_j^{(k)}(T)-A_j^{(k)}(t)\Big)\Big(A_i^{(k)}(T)-A_i^{(k)}(t)\Big)}{2\alpha_i^{(k)}(t)\alpha_j^{(k)}(t)}V_{ij}(t)\Bigg]
\end{aligned}
$$

$$
=\frac{B(0,T)}{B(0,t)}\exp\Bigg[-\sum_{k=1}^{M}\sum_{j=1}^{N_k}\frac{A_j^{(k)}(T)-A_j^{(k)}(t)}{\alpha_j^{(k)}(t)}X_j^{(k)}(t)
$$

$$-\sum_{k=1}^{M}\sum_{i,j=1}^{N_k}\frac{(A_j^{(k)}(T)-A_j^{(k)}(t))(A_i^{(k)}(T)-A_i^{(k)}(t))}{2\alpha_i^{(k)}(t)\alpha_j^{(k)}(t)}V_{ij}(t)\Bigg]$$

Applying the definitions of $G_j^{(k)}(t,T)$ and $H(t,T)$ leads directly to formula (12).

□

D.2 Proof of Theorem 6.1

Proof.
The dynamics of the state variable $X_1(t)$ under the T-forward measure are given by

$$\begin{aligned}\mathrm{d}X_1(t) = \Bigg[&V_{11}^{(1)}(t)+V_{12}^{(1)}(t)-c\Big[(T-t)c-\\&\frac{a_0^{(1)}+a_1^{(1)}t}{\lambda^{(1)}}\Big(\exp(\lambda^{(1)}(t-T))-1\Big)\Big]\Bigg]\mathrm{d}t+c\mathrm{d}W^{T,(1)}(t).\end{aligned}$$

Using the short-hand notation

$$U_1(t,T) := V_{11}^{(1)}(t)+V_{12}^{(1)}(t)-c\Big[(T-t)c-\frac{a_0^{(1)}+a_1^{(1)}t}{\lambda^{(1)}}\Big(\exp(\lambda^{(1)}(t-T))-1\Big)\Big],$$

this can be written as

$$X_1(t)-X_1(0)=\int_0^t U_1(s,T)\mathrm{d}s+\int_0^t c\mathrm{d}W^{T,(1)}(s).$$

Thus, $X_1(t)$ is normally distributed with mean $M_1(t)=X_1(0)+\int_0^t U_1(s,T)\mathrm{d}s$ and variance $S_1^2(t)=\int_0^t c^2\mathrm{d}s$,

$$X_1(t)\sim\mathcal{N}\Big(M_1(t),S_1^2(t)\Big).$$

The dynamics of the state variable $X_2(t)$ under the T-forward measure is given by

$$\begin{aligned}\mathrm{d}X_2(t) =& \Bigg[-\lambda^{(1)} X_2(t) + V_{21}^{(1)}(t) + V_{22}^{(1)}(t) - (a_1^{(1)} t + a_0^{(1)})\Big((T-t)c \\ & - \frac{a_0^{(1)} + a_1^{(1)} t}{\lambda^{(1)}}\Big(\exp(\lambda^{(1)}(t-T)) - 1\Big)\Big)\Bigg]\mathrm{d}t \\ & + (a_1^{(1)} t + a_0^{(1)})\mathrm{d}W^{T,(1)}(t)\end{aligned}$$

Introducing the short-hand notation

$$\begin{aligned}U_2(t,T) =& V_{21}^{(1)}(t) + V_{22}^{(1)}(t) - (a_1^{(1)} t + a_0^{(1)})\Big[(T-t)c \\ & - \frac{a_0^{(1)} + a_1^{(1)} t}{\lambda^{(1)}}\Big(\exp(\lambda^{(1)}(t-T)) - 1\Big)\Big],\end{aligned}$$

the dynamic can be written as

$$\mathrm{d}X_2(t) + \lambda^{(1)} X_2(t) = U_2(t,T)\mathrm{d}t + (a_0^{(1)} + a_1^{(1)} t)\mathrm{d}W^{T,(1)}(t)$$

which maybe rewritten as

$$\begin{aligned}\mathrm{d}\Big(X_2(t)\exp(\lambda^{(1)} t)\Big) =& \, U_2(t,T)\exp(\lambda^{(1)} t)\mathrm{d}t \\ & + \exp(\lambda^{(1)} t)(a_0^{(1)} + a_1^{(1)} t)\mathrm{d}W^{T,(1)}(t)\end{aligned}$$

which implies that

$$\begin{aligned}X_2(t) = X_2(0)\exp(-\lambda^{(1)} t) &+ \int_0^t U_2(s,T)\exp(\lambda^{(1)}(s-t))\mathrm{d}s \\ &+ \int_0^t \exp(\lambda^{(1)}(s-t))(a_0^{(1)} + a_1^{(1)} s)\mathrm{d}W^{T,(1)}(s).\end{aligned}$$

Thus, $X_2(t)$ is normally distributed,

$$X_2(t) \sim \mathcal{N}\Big(M_2(t), S_2^2(t)\Big),$$

with mean

$$M_2(t) = X_2(0)\exp(-\lambda^{(1)}t) + \int_0^t U_2(s,T)\exp(\lambda^{(1)}(s-t))\mathrm{d}s$$

and variance

$$S_2^2(t) = \int_0^t \exp\Big(-2\lambda^{(1)}(t-s)\Big)(a_0^{(1)} + a_1^{(1)}s)^2\mathrm{d}s.$$

The distribution of the state variable $X_3(t)$ and $X_4(t)$ can be computed analogous to $X_2(t)$ and one ends with

$$X_3(t) \sim \mathcal{N}\Big(M_3(t), S_3^2(t)\Big)$$

and

$$X_4(t) \sim \mathcal{N}\Big(M_4(t), S_4^2(t)\Big).$$

The means $M_i(t)$ and the variances $S_i^2(t)$ associated to the state variables $X_i(t)$ for $i = 1, ..., 4$ are presented in equations (28) - (35) below.

$$\begin{aligned}
M_1(t) =& X_1(0) + \int_0^t U_1(s,T)\mathrm{d}s \\
=& X_1(0) + \int_0^t V_{11}^{(1)}(t) + V_{12}^{(1)}(t) - c^2(T-s) \\
&+ \frac{c(a_0^{(1)} + a_1^{(1)}s)}{\lambda^{(1)}}\Big(\exp(\lambda^{(1)}(s-T)) - 1\Big)\mathrm{d}s \\
=& X_1(0) + \frac{\exp(-\lambda^{(1)}t)}{2(\lambda^{(1)})^3}\Bigg[-2c\Big(a_1^{(1)} - a_0^{(1)}\lambda^{(1)}\Big) + c\exp(\lambda^{(1)}t)\Big(-2a_0^{(1)}\lambda^{(1)} \\
&+ (\lambda^{(1)})^2 t(2a_0^{(1)} + c\lambda^{(1)}t) + a_1^{(1)}(2 + \lambda^{(1)}t(-2 + \lambda^{(1)}t))\Big)\Bigg] \\
&+ \frac{\exp(-\lambda^{(1)}T)}{2(\lambda^{(1)})^3}\Bigg[2c\Big(a_1^{(1)} - a_0^{(1)}\lambda^{(1)} + \exp(\lambda^{(1)}t)\Big(a_0^{(1)}\lambda^{(1)} \\
&+ a_1^{(1)}(-1 + \lambda^{(1)}t)\Big)\Big) + c\exp(\lambda^{(1)}T)(\lambda^{(1)})^2 t\Big(-2a_0^{(1)} - a_1^{(1)}t
\end{aligned}$$

$$+ c\lambda^{(1)}(t-2T)\Big)\Bigg] \tag{28}$$

$$\begin{aligned}
M_2(t) =& X_2(0)\exp(-\lambda^{(1)}t) + \int_0^t U_2(s,T)\exp(\lambda^{(1)}(s-t))\mathrm{d}s \\
=& X_2(0)\exp(-\lambda^{(1)}t) + \frac{\exp(-2\lambda^{(1)}t)}{4(\lambda^{(1)})^4}\Bigg[(a_1^{(1)})^2 - 2a_0^{(1)}a_1^{(1)}\lambda^{(1)} + 2(a_0^{(1)})^2 \\
& (\lambda^{(1)})^2 - 4\exp(\lambda^{(1)}t)\Big(2(a_1^{(1)})^2 + a_0^{(1)}(\lambda^{(1)})^2(a_0^{(1)} + c + c\lambda^{(1)}t) \\
& - a_1^{(1)}\lambda^{(1)}\Big(2(a_0^{(1)} + c) + c\lambda^{(1)}t\Big)\Big) \\
& + \exp(2\lambda^{(1)}t)\Big(2a_0^{(1)}(a_0^{(1)} + 2c)(\lambda^{(1)})^2 + 2a_1^{(1)}\lambda^{(1)}(-3a_0^{(1)} - 4c \\
& + 2(a_0^{(1)} + c)\lambda^{(1)}t) + (a_1^{(1)})^2(7 + 2\lambda^{(1)}t(-3 + \lambda^{(1)}t))\Big)\Bigg] \\
& - \frac{\exp(-\lambda^{(1)}(t+T))}{4(\lambda^{(1)})^4}\Bigg[(a_1^{(1)})^2\Big(1 - 8\exp(\lambda^{(1)}T) + 4\exp(\lambda^{(1)}(t+T)) \\
& \Big(2 + \lambda^{(1)}t(-2 + \lambda^{(1)}t)\Big) + \exp(2\lambda^{(1)}t)(-1 - 2\lambda^{(1)}t(-1 + \lambda^{(1)}t))\Big) \\
& + 2a_0^{(1)}(\lambda^{(1)})^2\Big(- a_0^{(1)}(-1 + \exp(\lambda^{(1)}t))(1 + \exp(\lambda^{(1)}t) - 2\exp(\lambda^{(1)}T)) \\
& - 2c\exp(\lambda^{(1)}T)(1 + \lambda^{(1)}T) + 2c\exp(\lambda^{(1)}(t+T))(1 - \lambda^{(1)}t + \lambda^{(1)}T)\Big) \\
& + 2a_1^{(1)}\lambda^{(1)}\Big(a_0^{(1)}\Big(- 1 + 4\exp(\lambda^{(1)}T) + \exp(2\lambda^{(1)}t)(1 - 2\lambda^{(1)}t) \\
& + 4\exp(\lambda^{(1)}(t+T))(-1 + \lambda^{(1)}t)\Big) + 2c\exp(\lambda^{(1)}T)\Big(2 + \lambda^{(1)}T \\
& - \exp(\lambda^{(1)}t)\Big(2 + \lambda^{(1)}(T + t(-2 + \lambda^{(1)}t - \lambda^{(1)}T))\Big)\Big)\Big)\Bigg]
\end{aligned} \tag{29}$$

$$\begin{aligned}
M_3(t) =& X_3(0)\exp(-\lambda^{(2)}t) + \frac{\exp(-2\lambda^{(2)}t)}{24(\lambda^{(2)})^4}\Bigg[6(a_0^{(2)})^2(\lambda^{(2)})^2(1 + \exp(2\lambda^{(2)}t) \\
& (-1 + 2\lambda^{(2)}t)) + 6a_0^{(2)}a_1^{(2)}\lambda^{(2)}\Big(- 1 + \exp(2\lambda^{(2)}t)(1 + 2\lambda^{(2)}t(-1 \\
& + \lambda^{(2)}t))\Big) + (a_1^{(2)})^2\Big(3 + \exp(2\lambda^{(2)}t)(-3 + 2\lambda^{(2)}t(3 + \lambda^{(2)}t \\
& (-3 + 2\lambda^{(2)}t)))\Big)\Bigg] + \frac{\exp(-\lambda^{(2)}(t+T))}{4(\lambda^{(2)})^4}\Bigg[2(a_0^{(2)})^2(-1 + \exp(\lambda^{(2)}t)) \\
& (1 + \exp(\lambda^{(2)}t) - 2\exp(\lambda^{(2)}T))(\lambda^{(2)})^2 + 2a_0^{(2)}a_1^{(2)}\lambda^{(2)}
\end{aligned}$$

$$\Big(1-4\exp(\lambda^{(2)}T)+\exp(\lambda^{(2)}(t+T))(4-4\lambda^{(2)}t)+ \exp(2\lambda^{(2)}t)(-1+2\lambda^{(2)}t)\Big) + (a_1^{(2)})^2\Big(-1+8\exp(\lambda^{(2)}T)-4\exp(\lambda^{(2)}(t+T))(2+\lambda^{(2)}t (-2+\lambda^{(2)}t))+\exp(2\lambda^{(2)}t)\Big(1+2\lambda^{(2)}t(-1+\lambda^{(2)}t)\Big)\Big)\Bigg] \tag{30}$$

$$\begin{aligned} M_4(t) =& X_4(0)\exp(-\lambda^{(3)}t)+\frac{\exp(-2\lambda^{(3)}t)}{24(\lambda^{(3)})^4}\Bigg[6(a_0^{(3)})^2(\lambda^{(3)})^2(1+\exp(2\lambda^{(3)}t) \\ & (-1+2\lambda^{(3)}t))+6a_0^{(3)}a_1^{(3)}\lambda^{(3)}\Big(-1+\exp(2\lambda^{(3)}t)(1+2\lambda^{(3)}t(-1 \\ & +\lambda^{(3)}t))\Big)+(a_1^{(3)})^2\Big(3+\exp(2\lambda^{(3)}t)(-3+2\lambda^{(3)}t(3+\lambda^{(3)}t \\ & (-3+2\lambda^{(3)}t)))\Big)\Bigg]+\frac{\exp(-\lambda^{(3)}(t+T))}{4(\lambda^{(3)})^4}\Bigg[2(a_0^{(3)})^2(-1+\exp(\lambda^{(3)}t)) \\ & (1+\exp(\lambda^{(3)}t)-2\exp(\lambda^{(3)}T))(\lambda^{(3)})^2+2a_0^{(3)}a_1^{(3)}\lambda^{(3)}\Big(1-4\exp(\lambda^{(3)}T) \\ & +\exp(\lambda^{(3)}(t+T))(4-4\lambda^{(3)}t)+\exp(2\lambda^{(3)}t)(-1+2\lambda^{(3)}t)\Big) \\ & +(a_1^{(3)})^2\Big(-1+8\exp(\lambda^{(3)}T)-4\exp(\lambda^{(3)}(t+T))(2+\lambda^{(3)}t \\ & (-2+\lambda^{(3)}t))+\exp(2\lambda^{(3)}t)\Big(1+2\lambda^{(3)}t(-1+\lambda^{(3)}t)\Big)\Big)\Bigg] \end{aligned} \tag{31}$$

$$S_1^2(t)=\int_0^t c^2\mathrm{d}s=tc^2 \tag{32}$$

$$\begin{aligned} S_2^2(t) =& \int_0^t \exp\Big(-2\lambda^{(1)}(t-s)\Big)(a_0^{(1)}+a_1^{(1)}s)^2\mathrm{d}s \\ =& \frac{1}{4(\lambda^{(1)})^3}\Bigg[2(a_0^{(1)})^2(\lambda^{(1)})^2-\exp(-2\lambda^{(1)}t)\Big((a_1^{(1)})^2-2a_0^{(1)}a_1^{(1)}\lambda^{(1)} \\ & +2(a_0^{(1)})^2(\lambda^{(1)})^2\Big)+2a_0^{(1)}a_1^{(1)}\lambda^{(1)}(-1+2\lambda^{(1)}t) \\ & +(a_1^{(1)})^2(1+2\lambda^{(1)}t(-1+\lambda^{(1)}t))\Bigg] \end{aligned} \tag{33}$$

$$
\begin{aligned}
S_3^2(t) =& \int_0^t \exp(-2\lambda^{(2)}(t-s))\Big(a_1^{(2)}s + a_0^{(2)}\Big)^2 \mathrm{d}s \\
=& \frac{1}{4(\lambda^{(2)})^3}\Bigg[2(a_0^{(2)})^2(\lambda^{(2)})^2 - \exp(-2\lambda^{(2)}t)\Big((a_1^{(2)})^2 - 2a_0^{(2)}a_1^{(2)}\lambda^{(2)} \\
&+ 2(a_0^{(2)})^2(\lambda^{(2)})^2\Big) + 2a_0^{(2)}a_1^{(2)}\lambda^{(2)}(-1+2\lambda^{(2)}t) \\
&+ (a_1^{(2)})^2\Big(1 + 2\lambda^{(2)}t(-1+\lambda^{(2)}t)\Big)\Bigg] \qquad (34)
\end{aligned}
$$

$$
\begin{aligned}
S_4^2(t) =& \int_0^t \exp(-2\lambda^{(3)}(t-s))\Big(a_1^{(3)}s + a_0^{(3)}\Big)^2 \mathrm{d}s \\
=& \frac{1}{4(\lambda^{(3)})^3}\Bigg[2(a_0^{(3)})^2(\lambda^{(3)})^2 - \exp(-2\lambda^{(3)}t)\Big((a_1^{(3)})^2 - 2a_0^{(3)}a_1^{(3)}\lambda^{(3)} \\
&+ 2(a_0^{(3)})^2(\lambda^{(3)})^2\Big) + 2a_0^{(3)}a_1^{(3)}\lambda^{(3)}(-1+2\lambda^{(3)}t) \\
&+ (a_1^{(3)})^2\Big(1 + 2\lambda^{(3)}t(-1+\lambda^{(3)}t)\Big)\Bigg] \qquad (35)
\end{aligned}
$$

Next, the covariance is essential to determine the joint distribution of the state variables, and it is needed to generate sample paths. The 3 Factor Exponential Model contains four state variables and is driven by three independent Brownian Motions. The state variables X_1 and X_2 are driven by the same Brownian Motion, thus they are not independent and the covariance does not vanish. In contrast, the other covariance entries vanish, because the state variables are pairwise independent. Hence, we just have to incorporate the covariance of X_1 and X_2. The state variables X_1 and X_2 are given by

$$
\begin{aligned}
X_1(t) =& M_1(t) + \int_0^t c\mathrm{d}W^{(1)}(s), \\
X_2(t) =& M_2(t) + \int_0^t (a_1^{(1)}s + a_0^{(1)})\mathrm{d}W^{(1)}(s),
\end{aligned}
$$

from which we calculate the covariance as

$$\begin{aligned}
\operatorname{cov}(X_1, X_2) =& \mathbb{E}[X_1(t)X_2(t)] - \underbrace{\mathbb{E}[X_1(t)]\mathbb{E}[X_2(t)]}_{=M_1(t)M_2(t)} \\
=& \mathbb{E}[M_1(t)M_2(t)] + \underbrace{\mathbb{E}[M_1(t)\int_0^t (a_1^{(1)}s + a_0^{(1)})\mathrm{d}W^{(1)}(s)]}_{=0} \\
&+ \underbrace{\mathbb{E}[M_2(t)\int_0^t c\mathrm{d}W^{(1)}(s)]}_{=0} \\
&+ \mathbb{E}\Big[\int_0^t (a_1^{(1)}s + a_0^{(1)})\mathrm{d}W^{(1)}(s)\int_0^t c\mathrm{d}W^{(1)}(s)\Big] - M_1(t)M_2(t) \\
=& \mathbb{E}\Big[\int_0^t (a_1^{(1)}s + a_0^{(1)})\mathrm{d}W^{(1)}(s)\int_0^t c\mathrm{d}W^{(1)}(s)\Big]
\end{aligned}$$

Calculating the expression $\int_0^t (a_1^{(1)}s + a_0^{(1)})\mathrm{d}W^{(1)}(s)\int_0^t c\mathrm{d}W^{(1)}(s)$ by Itô's Formula and taking expectation leads to

$$\operatorname{cov}(X_1, X_2) = \int_0^t c(a_1^{(1)}s + a_0^{(1)})\mathrm{d}s = c\Big(\frac{1}{2}a_1^{(1)}t^2 + a_0^{(1)}t\Big).$$

Thus, the time-dependent covariance matrix $\Sigma = \Sigma(t)$ in the 3 Factor Exponential Model is given by

$$\Sigma(t) = \begin{pmatrix} S_1^2(t) & c(\frac{1}{2}a_1^{(1)}t^2 + a_0^{(1)}t) & 0 & 0 \\ c(\frac{1}{2}a_1^{(1)}t^2 + a_0^{(1)}t) & S_2^2(t) & 0 & 0 \\ 0 & 0 & S_3^2(t) & 0 \\ 0 & 0 & 0 & S_4^2(t) \end{pmatrix}.$$

□

D.3 Proof of Theorem 7.4

Proof.
Applying the 3 Factor Exponential Model, the forward rate $f(t,T)$ is given by

$$\begin{aligned} f(t,T) =& f(0,T) + \sum_{k=1}^{3} \left(\sum_{j=1}^{N_k} \frac{\alpha_j^{(k)}(T)}{\alpha_j^{(k)}(t)} \left[X_j^{(k)}(t) \right.\right. \\ & \left.\left. + \sum_{i=1}^{N_k} \frac{A_i^{(k)}(T) - A_i^{(k)}(t)}{\alpha_i^{(k)}(t)} V_{ij}^{(k)}(t) \right] \right). \end{aligned}$$

Using the notation $X_1(t) = X_1^{(1)}(t)$, $X_2(t) = X_2^{(1)}(t)$, $X_3(t) = X_1^{(2)}(t)$ and $X_4(t) = X_1^{(3)}(t)$ and the functions α, β of the volatility parametrization (9) - (11), one ends up with

$$\begin{aligned} f(t,T) =& f(0,T) + X_1(t) + (T-t)V_{11}^{(1)}(t) \\ & + \Big(\exp(-\lambda^{(1)}t) - \exp(-\lambda^{(1)}T) \Big) \frac{V_{12}^{(1)}(t)}{\lambda^{(1)} \exp(-\lambda^{(1)}t)} \\ & + \exp(-\lambda^{(1)}(T-t)) \left[X_2(t) + (T-t)V_{21}^{(1)}(t) + \Big(\exp(-\lambda(1)t) \right. \\ & \left. - \exp(-\lambda^{(1)}T) \Big) \frac{V_{22}^{(1)}(t)}{\lambda^{(1)} \exp(-\lambda^{(1)}t)} \right] + \exp(-\lambda^{(2)}(T-t)) \left[X_3(t) \right. \\ & \left. + \frac{V_{11}^{(2)}(t)}{\lambda^{(2)} \exp(-\lambda^{(2)}t)} \Big(\exp(-\lambda^{(2)}t) - \exp(-\lambda^{(2)}T) \Big) \right] \\ & + \exp(-\lambda^{(3)}(T-t)) \left[X_4(t) + \frac{V_{11}^{(3)}(t)}{\lambda^{(3)} \exp(-\lambda^{(3)}t)} \right. \\ & \left. \Big(\exp(-\lambda^{(3)}t) - \exp(-\lambda^{(3)}T) \Big) \right] \end{aligned}$$

We have to express the underlying cap rate $R(T_C, T_B)$ in terms of the state variables X_i. Therefore we first focus on the integral $\int_{T_C}^{T_B} f(T_C, s)\mathrm{d}s$ in (26):

$$\int_{T_C}^{T_B} f(T_C,s)\mathrm{d}s = \int_{T_C}^{T_B} f(0,s)\mathrm{d}s + \int_{T_C}^{T_B} \Bigg(X_1(T_C) + (s-T_C)V_{11}^{(1)}(T_C)$$

$$
\begin{aligned}
&+\Big(\exp(-\lambda^{(1)}T_C)-\exp(-\lambda^{(1)}s)\Big)\frac{V_{12}^{(1)}(T_C)}{\lambda^{(1)}\exp(-\lambda^{(1)}s)}\Bigg)\mathrm{d}s\\
&+\int_{T_C}^{T_B}\Bigg(\exp(-\lambda^{(1)}(s-T_C))\Bigg[X_2(T_C)+(s-T_C)V_{21}^{(1)}(T_C)\\
&+\Big(\exp(-\lambda^{(1)}T_C)-\exp(-\lambda^{(1)}s)\Big)\frac{V_{22}^{(1)}(T_C)}{\lambda^{(1)}\exp(-\lambda^{(1)}s)}\Bigg]\Bigg)\mathrm{d}s\\
&+\int_{T_C}^{T_B}\Bigg(\exp(-\lambda^{(2)}(s-T_C))\Bigg[X_3(T_C)\\
&+\frac{V_{11}^{(2)}(T_C)}{\lambda^{(2)}\exp(-\lambda^{(2)}T_C)}\Big(\exp(-\lambda^{(2)}T_C)-\exp(-\lambda^{(2)}s)\Big)\Bigg]\Bigg)\mathrm{d}s\\
&+\int_{T_C}^{T_B}\Bigg(\exp(-\lambda^{(3)}(s-T_C))\Bigg[X_4(T_C)\\
&+\frac{V_{11}^{(3)}(T_C)}{\lambda^{(3)}\exp(-\lambda^{(3)}T_C)}\Big(\exp(-\lambda^{(3)}T_C)-\exp(-\lambda^{(2)}s)\Big)\Bigg]\Bigg)\mathrm{d}s
\end{aligned}
$$

Each integral can be computed separately and we introduce the short-hand notation $K_i(X_i, T_C, T_B)$ for $i = 1, ...4$:

$$
\begin{aligned}
K_1(X_1,T_C,T_B)=&\int_{T_C}^{T_B}\Bigg(X_1(T_C)+(s-T_C)V_{11}^{(1)}(T_C)\\
&+\Big(\exp(-\lambda^{(1)}T_C)-\exp(-\lambda^{(1)}s)\Big)\frac{V_{12}^{(1)}(T_C)}{\lambda^{(1)}\exp(-\lambda^{(1)}s)}\Bigg)\mathrm{d}s\\
=&\frac{1}{2(\lambda^{(1)})^2}\Bigg[2(-1+\exp(\lambda^{(1)}(T_C-T_B)))V_{12}^{(1)}(T_C)\\
&+\lambda^{(1)}(T_C-T_B)\Big(-2V_{12}^{(1)}(T_C)+\\
&+\lambda^{(1)}\Big(T_CV_{11}^{(1)}(T_C)-T_BV_{11}^{(1)}(T_C)-2x_1(T_C)\Big)\Big)\Bigg] \qquad (36)\\
K_2(X_2,T_C,T_B)=&\int_{T_C}^{T_B}\Bigg(\exp(-\lambda^{(1)}(s-T_C))\Bigg[X_2(T_C)+(s-T_C)V_{21}^{(1)}(T_C)
\end{aligned}
$$

$$
\begin{aligned}
&+\Big(\exp(-\lambda(1)T_C)-\exp(-\lambda^{(1)}s)\Big)\frac{V_{22}^{(1)}(T_C)}{\lambda^{(1)}\exp(-\lambda^{(1)}s)}\Bigg]\Bigg)\mathrm{d}s\\
=&\frac{\exp(-2\lambda^{(1)}T_B)}{2(\lambda^{(1)})^2}\Bigg[\exp(2\lambda^{(1)}T_C)V_{22}^{(1)}(T_C)\\
&+2\exp(\lambda^{(1)}(T_C+T_B))\Big((-1+\lambda^{(1)}T_C-\lambda^{(1)}T_B)V_{21}^{(1)}(T_C)\\
&-V_{22}^{(1)}(T_C)-\lambda^{(1)}x_2(T_C)\Big)\\
&+\exp(2\lambda^{(1)}T_B)\Big(2V_{21}^{(1)}(T_C)+V_{22}^{(1)}(T_C)+2\lambda^{(1)}X_2(T_C)\Big)\Bigg]
\end{aligned}
\tag{37}
$$

$$
\begin{aligned}
K_3(X_3,T_C,T_B)=&\int_{T_C}^{T_B}\Bigg(\exp(-\lambda^{(2)}(s-T_C))\Bigg[X_3(T_C)\\
&+\frac{V_{11}^{(2)}(T_C)}{\lambda^{(2)}\exp(-\lambda^{(2)}T_C)}\Big(\exp(-\lambda^{(2)}T_C)-\exp(-\lambda^{(2)}s)\Big)\Bigg]\Bigg)\mathrm{d}s\\
=&\frac{1-\exp(\lambda^{(2)}(T_C-T_B))}{2(\lambda^{(2)})^2}\Bigg[V_{11}^{(2)}(T_C)-\exp(\lambda^{(2)}(T_C-T_B))\\
&V_{11}^{(2)}(T_C)+2\lambda^{(2)}X_3(T_C)\Bigg]
\end{aligned}
\tag{38}
$$

$$
\begin{aligned}
K_4(X_4,T_C,T_B)=&\int_{T_C}^{T_B}\Bigg(\exp(-\lambda^{(3)}(s-T_C))\Bigg[X_4(T_C)\\
&+\frac{V_{11}^{(3)}(T_C)}{\lambda^{(3)}\exp(-\lambda^{(3)}T_C)}\Big(\exp(-\lambda^{(3)}T_C)-\exp(-\lambda^{(3)}s)\Big)\Bigg]\Bigg)\mathrm{d}s\\
=&\frac{1-\exp(\lambda^{(3)}(T_C-T_B))}{2(\lambda^{(3)})^2}\Bigg[V_{11}^{(3)}(T_C)-\exp(\lambda^{(3)}(T_C-T_B))\\
&V_{11}^{(3)}(T_C)+2\lambda^{(3)}X_4(T_C)\Bigg]
\end{aligned}
\tag{39}
$$

Using the functions $K_i(X_i,T_C,T_B)$ for $i=1,...,4$, the cap rate $R(T_C,T_B)$ can be written as

$$
R(T_C,T_B)=\frac{1}{\Delta}\left[\exp\left(\int_{T_C}^{T_B}f(0,s)\mathrm{d}s+\sum_{i=1}^{4}K_i(X_i,T_C,T_B)\right)-1\right]
$$

$$= \frac{1}{\Delta}\left[\exp\left(\int_{T_C}^{T_B} f(0,s)\mathrm{d}s \right) \exp\left(\sum_{i=1}^{4} K_i(X_i, T_C, T_B) \right) - 1 \right]$$

If we assume a constant initial forward rate f_0 the expression reduces to

$$R(T_C, T_B) = \frac{1}{\Delta}\left[\exp\left(\Delta f_0 \right) \exp\left(\sum_{i=1}^{4} K_i(X_i, T_C, T_B) \right) - 1 \right] \tag{40}$$

Equation (25) in combination with the bond price formula in Lemma (4.2) leads to the terminal condition of the PDE for pricing a caplet in the 3 Factor Exponential Model, namely

$$\phi(X_{T_C}) = \exp(-\Delta f_0) \exp\Big[- \sum_{k=1}^{3} \sum_{j=1}^{N_k} G_j^{(k)}(T_C, T_B) X_j^{(k)}(T_C) - H(T_C, T_B) \Big]$$
$$\max\left[\frac{1}{\Delta}\left(\exp\Big(\Delta f_0\Big) \exp\Big(\sum_{i=1}^{4} K_i(X_i, T_C, T_B) \Big) - 1 \right) - K, 0 \right]$$

with $G_j^{(k)}(t,T)$ and $H(t,T)$ defined in Lemma 4.2. □

E Numerical Method for PDE Valuation

E.1 Finite Difference with PSOR

E.1.1 Boundary conditions

We use the linearity condition on the boundaries, for $i = 1, 2, \ldots, n$,

$$\frac{\partial^2 g}{\partial x_i^2} = 0, \ \forall x \in \partial\Omega.$$

This boundary condition is basically a linear extrapolation at the particular boundary.

E.1.2 Discretisation

The discretization of the PDE is based on central differences to approximate the derivatives of the state variables, as is usual for Finite Differences schemes. The derivatives in the time direction are obtained by changing the variables to $\tau = T - t$ and implementing a θ-scheme (for each one of n

directions, $i = 1, 2, \ldots, n$). Thus, we end up with the discretization of the PDE for $m = 0, 1, \ldots, N_\tau - 1$ given by

$$\begin{aligned}\frac{g_i^{m+1} - g_i^m}{\Delta\tau} =&(1-\theta)\left[\sigma_i^m \frac{g_{i+1}^m - 2g_i^m + g_{i-1}^m}{\Delta x_i^2} + b_i^m \frac{g_{i+1}^m - g_{i-1}^m}{2\Delta x_i}\right.\\ &\left. - \left(f^m + \sum_{i=1}^{n} x_j\right) g_i^m\right] + \theta\left[\sigma_i^{m+1} \frac{g_{i+1}^{m+1} - 2g_i^{m+1} + g_{i-1}^{m+1}}{\Delta x_i^2}\right.\\ &\left. + b_i^{m+1} \frac{g_{i+1}^{m+1} - g_{i-1}^{m+1}}{2\Delta x_i} - \left(f^{m+1} + \sum_{j=1}^{n} x_j\right) g_i^{m+1}\right]\end{aligned}$$

Depending on the choice of θ, we have a fully implicit scheme ($\theta = 1$) or the Crank-Nicolson scheme ($\theta = \frac{1}{2}$). We use the implicit scheme for the first few time steps and then switch to the Crank Nicolson Scheme for the remaining time steps. We start with the implicit scheme to reduce possible oscillations due to the non-differentiable initial condition in the case of a caplet. Later, we use the Crank Nicolson Scheme as it has the highest convergence rate, which is second-order in time.

E.1.3 PSOR

The spatial discretization above leads to a semi-discrete equation which has the matrix representation

$$\frac{\partial \boldsymbol{g}}{\partial \tau} = \boldsymbol{A}\boldsymbol{g} \tag{41}$$

where $\boldsymbol{A}$ is a block tridiagonal $(N_1 + 1) \cdots (N_d + 1) \times (N_1 + 1) \cdots (N_d + 1)$ matrix and $\boldsymbol{g}$ is a vector of length $(N_1+1) \cdots (N_d+1)$. Using the $\theta-$ scheme to discretise the semi-discrete problem (41), it can be written as

$$(\boldsymbol{I} + \theta\Delta\tau\boldsymbol{A})\, \boldsymbol{g}^{(m+1)} = (\boldsymbol{I} - (1-\theta)\Delta\tau\boldsymbol{A})\, \boldsymbol{g}^{(m)},\ m = 0, \ldots, N_\tau - 1, \tag{42}$$

where N_τ is the number of time steps and $\boldsymbol{I}$ is the identity matrix of the same dimension as $\boldsymbol{A}$.

After the discretisation of the underlying PDE with n spatial variables an approximate price of an American option can be obtained by solving a

sequence of linear complementarity problems (LCPs)

$$\begin{cases} \boldsymbol{B}\boldsymbol{g}^{(m+1)} \geq \boldsymbol{E}\boldsymbol{g}^{(m)}, \;\; \boldsymbol{g}^{(m+1)} \geq \boldsymbol{\phi}, \\ \left(\boldsymbol{B}\boldsymbol{g}^{(m+1)} - \boldsymbol{E}\boldsymbol{g}^{(m)}\right)^T \left(\boldsymbol{g}^{(m+1)} - \boldsymbol{g}\right) = 0, \end{cases} \tag{43}$$

for $m = 0, \ldots, N_\tau - 1$. The general methodology is suitable for European, Bermudan and American style derivatives. The matrices $\boldsymbol{B}$ and $\boldsymbol{E}$ in (43) are defined by the left hand side and right hand side of Eqn. (42) respectively. The initial value $\boldsymbol{g}^{(0)}$ is given by the discrete form $\boldsymbol{\phi}$ of the payoff function ϕ.

The system (43) is then solved by the projected successive over-relaxation method (PSOR). This method is an element-wise iterative solution method. Consider the computation of the vector $\boldsymbol{g}^{(m+1)}$ and a desired tolerance ε. Let $\boldsymbol{y}^k$ be the kth iterated solution of $\boldsymbol{g}^{(m)}$, then by processing each element and, defining $h^{(m)} = \boldsymbol{E}\boldsymbol{g}^{(m)}$ and also defining

$$z_i^k = h_i^{(m)} - \sum_{j=0}^{i-1} B_{ij} y_j^k - \sum_{j=i}^{N} B_{ij} y_j^{k-1}, \tag{44}$$

with B_{ij} entries of matrix $\boldsymbol{B}$ and y_j^k the $j-$th position of vector $\boldsymbol{y}^k$. Then for a European style derivative without early exercise features, the new iterated solution for position i reads

$$y_i^k = y_i^{k-1} + \frac{z_i^k}{B_{ii}}; \tag{45}$$

while for a American or Bermudan style derivatives with early exercise features, the new iterated solution for position i reads

$$y_i^k = \max\left(\phi_i, y_i^{k-1} + \frac{z_i^k}{B_{ii}}\right). \tag{46}$$

We observe that in equation (44) the iterated solution at step $k+1$ already occurs, the reason being the element-wise processing. For element j, the known values of $\boldsymbol{y}^k$ are used. The iteration stops when the solution of the original problem is lower than the desired tolerance, that is

$$||\boldsymbol{g}^{(m+1)} - \boldsymbol{g}^{(m)}||_\infty < \varepsilon,$$

where $||\boldsymbol{x}||_\infty$ is the L^∞ norm of the vector $\boldsymbol{x}$ defined by $||\boldsymbol{x}||_\infty = \max_i |x_i|$.

E.2 Sparse Grid Implementation

In order to tackle the computationally demanding task of solving the 4-dimensional parabolic PDE as given in Theorem 7.1 we apply the sparse grid approach that turns out to be quite fast and accurate. The sparse grid combination technique for solving PDEs was introduced by Reisinger (2004) and Bungartz & Griebel (2004). This general technique can be applied to many different problems and one reasonable application in finance is presented by Chiarella & Kang (2012). The combination technique requires the solution of the original equation only on a set of conventional subspaces defined on Cartesian grids specified in a certain way and a subsequent extrapolation step, but still retains a certain convergence order.

E.2.1 The Sparse Grid Combination Technique

The spatial domain of the PDE in the 3 Factor Exponential Model is four dimensional and this number changes due to changes in the model. The sparse grid technique is very general and can be applied to high dimensional problems. In order not to complicate the introduction of the method, we reduce the following discussion of the sparse grid technique to two dimensions only, as we can illustrate the theory by pictures. The methodology can be extended easily to multi-dimensions.

In the following, consider the $2-$dimensional cube $\Omega := [0,1]\times[0,1]$ and a Cartesian grid with mesh size $h_j = 2^{-l_j}$ (corresponding to a level $l_j \in \mathbb{N}_0$) in the directions $j = 1, 2$. The indices $j = 1$ and $j = 2$ represent the directions of the first state variable X_1 and second state variable X_2 respectively.

For a vector $\mathbf{h} = (h_1, h_2)$ we denote by $c_{\mathbf{h}}$ the representation of a function on such a grid with points

$$\mathbf{x}_h = (i_1 \cdot h_1,\ i_2 \cdot h_2) \text{ for } j = 1, 2,\ 0 \leq i_j \leq N_j,\ N_j = 1/h_j = 2^{l_j}.$$

For a given level l, the above grid consists all possible combinations of (l_1, l_2) with $0 \leq l_1, l_2 \leq l$. In total, there are $2^{2(l+1)}$ points in the grid. The curse of dimensionality comes into effect as the level l is increased.

However, with the same level l, the *sparse* grid, will consist of the points

$$\mathbf{x}_h = (i_1 \cdot h_1,\ i_2 \cdot h_2) \text{ for } j = 1, 2,\ 0 \leq i_j \leq N_j,\ N_j = 1/h_j = 2^{l_j},$$

satisfying $l_1 + l_2 = l$. It is clear that there are $l+1$ choices of such combinations of (l_1, l_2) these $(0, l), (1, l-1), \cdots, (l-1, 1), (l, 0)$. Figure 5 provides an example of a standard sparse grid hierarchy with level $l = 5$ with respect to the six different combinations.

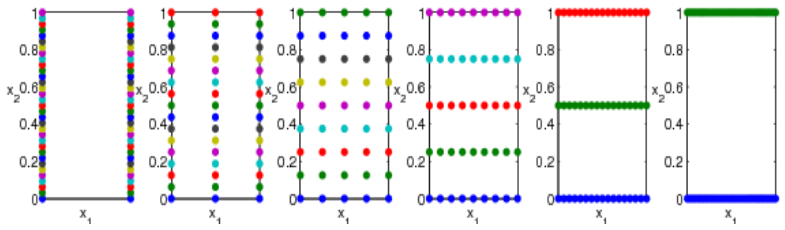


Figure 5: A sparse grid hierarchy of level 5 with respect to each combination. From the left to right, these are $(0, 5), (1, 4), (2, 3), (3, 2), (4, 1), (5, 0)$ respectively

Obviously, the above grids share the common properties that they are dense in one direction but sparse in the other direction. If we put all of the above grids together, we will obtain the standard sparse grid shown in Figure 6.

Let $\mathbf{c}_h$ be the discrete vector of function values at the grid points of the standard sparse grid. In general, $\mathbf{c}_h$ is the finite difference solution to the PDE of interest on the corresponding grid $\mathbf{h}$. The solution can be extended to Ω by a suitable multi-linear interpolation operator $\mathcal{I}$[5] in the point wise sense according to

$$c_{\mathbf{h}}(X_1, X_2, \tau) = \mathcal{I}\mathbf{c}_h, \forall (X_1, X_2) \in \Omega.$$

Next, we define the family C of solutions corresponding to the different sparse grids (as in Figure 5 for instance) by $C = (C(\mathbf{i}))_{\mathbf{i} \in \mathbb{N}^2}$ with

$$C(\mathbf{i}) := c_{2^{-\mathbf{i}}},$$

[5]A thorough error analysis of the multi-linear interpolation operator can be found in Reisinger (2008) who gives a generic derivation for linear difference schemes through an error correction technique employing semi-discretisations and obtains error formulae as well.

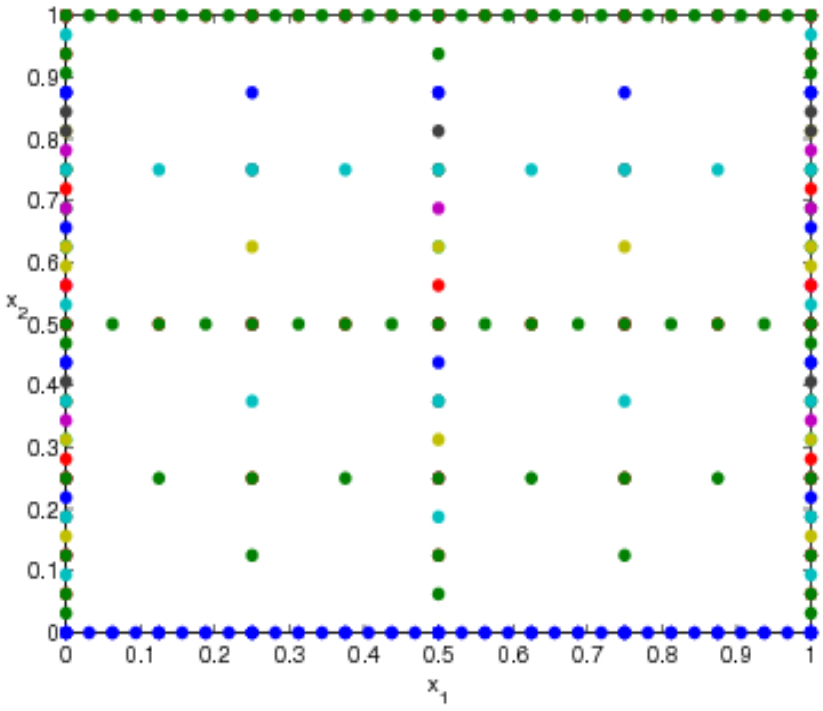


Figure 6: A standard sparse grid of level 5.

that is the family of numerical approximations (after proper interpolation) $c_{\mathbf{h}}$ on tensor product grids with $h_k = 2^{-i_k}$. For example, the solution on the first grid in Figure 5 would be $C(0,5)$ etc. The combination technique presented by Reisinger & Wittum (2007) tells us that the solution c_l (l is the level of the sparse grid) of the corresponding PDE is

$$c_l = \sum_{n=0}^{l} C(n, l-n) - \sum_{n=0}^{l-1} C(n, l-1-n). \tag{47}$$

The procedure involves solving the PDE in parallel on each of the sparse grids of level l and level $l-1$ respectively. See Figure 5 for $l = 5$ as an example. Thus there are $(2l+1)$ PDE solvers running simultaneously. The theory developed by Reisinger & Wittum (2007) shows that Equation (47) combines all solutions together to yield a more accurate solution to the PDE. The essential principle of the extrapolation is that all lower order error terms

cancel out in the combination formula (47) and only the highest order terms

$$h_1^2 \cdot h_2^2 = (2^{-l_1} \cdot 2^{-l_2})^2 = 4^{-l}$$

remain. Taking advantage of this cancelation mechanism, Eq.(47) is able to produce quite accurate results fairly quickly. The details of the error analysis can be found in Reisinger (2004) and Reisinger & Wittum (2007).

E.2.2 Modified Sparse Grid

We are able to apply the technique described above to the four dimensional PDE discussed in Section 7.2. However, for our model, the PDE contains mixed derivatives whose values become very large and we have found that it is difficult to implement the above described sparse grid combination techniques using the standard sparse grid shown in Figure 5 as it produces rather bad results. Furthermore, the standard sparse grid provides the difficulty that it is difficult to approximate accurately the boundary conditions for some 'extreme' grid points, e.g. the first $(0, 5)$ and the last $(5, 0)$ grid in Figure 5 which only have two points in one of the two directions, e.g. X_1 or X_2.

In order to overcome these problems, we have modified the above approach slightly, namely by adding a fixed number of points to both the X_1 and X_2 directions in each of the grids, which results in a relative 'balance' in both directions, and, indeed the numerical results indicate that this modification produces accurate and efficient prices. The technique has already been presented by Chiarella & Kang (2012).

More specifically, we let both levels l_1, l_2 in different directions start from a small but non zero value l_s which means that the new modified levels $\hat{l}_1, \hat{l}_2$ are defined as

$$\hat{l}_1 = l_1 + l_s,\ \hat{l}_2 = l_2 + l_s,\ l_1 + l_2 = l,\ l_1, l_2 \in \mathbb{N}_0.$$

Then the total level of the modified sparse grid becomes $2l_s + l$ in terms of the level originally defined. Hence, the modified sparse grid will consist of the points

$$\hat{\mathbf{x}}_h = (i_1 \cdot \hat{h}_1, i_2 \cdot \hat{h}_2),\ \text{for } j = 1, 2,\ \hat{N}_j = 1/\hat{h}_j = 2^{\hat{l}_j},$$

$$\hat{l}_1 = l_1 + l_s,\ \hat{l}_2 = l_2 + l_s,\ l_1 + l_2 = l.$$

Similarly, there are still $l+1$ choices of combinations of (l_1, l_2) with $(0, l), (1, l-1), \cdots, (l-1, 1), (l, 0)$, consequently, we will have the combinations $(l_s, l + l_s), (l_s + 1, l + l_s - 1), \cdots, (l_s + l - 1, l_s - 1), (l_s + l, l_s)$ for the modified sparse grid. Figure 7 shows an example of a modified sparse grid hierarchy with $l_s = 2, l = 2$ $(2l_s + l = 6)$, namely $(2, 4), (3, 3), (4, 2)$. Figure 8 shows the corresponding modified sparse grid.

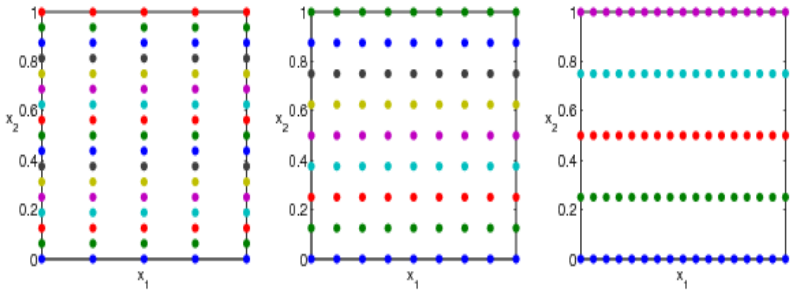


Figure 7: A modified sparse grid hierarchy with a initial level 2 and total level 6 with respect to each combination.

Analogously to Equation (47), a solution $\hat{c}_{l_s,l}$ (l_s, l are the initial level and the level of the sparse grid respectively) of the corresponding PDE in two dimensions with the modified sparse grid is given as

$$\hat{c}_{(l_s,l)} = \sum_{n=0}^{l} C(l_s + n, l_s + l - n) - \sum_{n=0}^{l-1} C(l_s + n, l_s + l - 1 - n). \qquad (48)$$

We implement the modified sparse grid combination technique to solve the PDE in Section 7.2 in order to obtain the prices of bonds and caplets given by the terminal conditions specified in Section 7.3). In the implementation, a standard Crank Nicolson finite difference method with projected successive over-relaxation (PSOR), as described in Section E.1, has been applied to solve the four dimensional PDE.

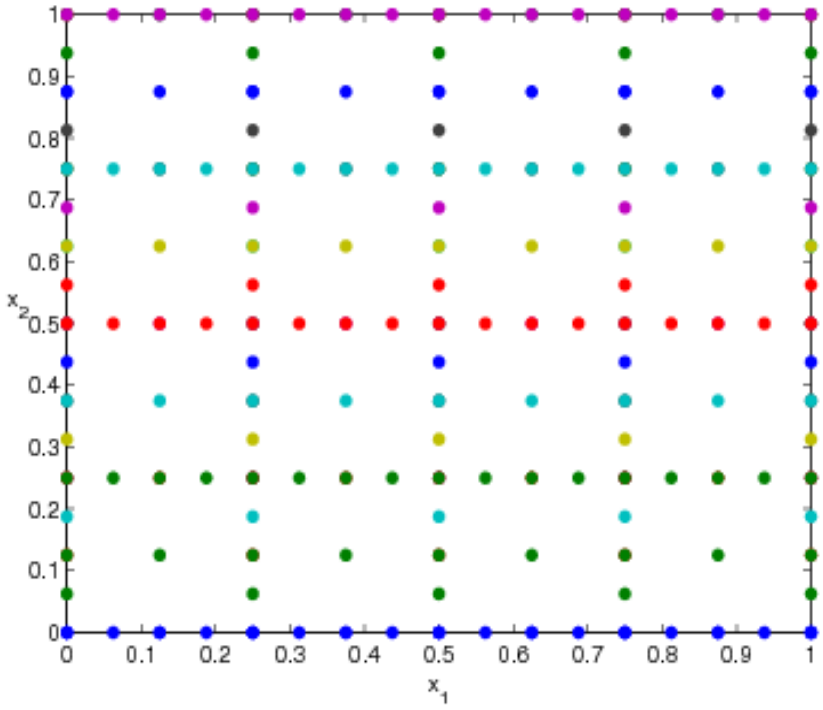


Figure 8: A modified sparse grid with a initial level 2 and total level 6.