Efficient Simulation and Valuation of Embedded Options using Monte Carlo Simulations

Stefan Singor

Insurance Risk Management

Ortec Finance

Supervisors:

Prof. Dr. Ir. C.W. Oosterlee

Drs. Dr. Ir. D.D.B. van Bragt RBA

Other Thesis Committee Members:

Dr. J.A.M. van der Weide

Dr. R.J. Fokkink

Faculty of Electrical Engineering, Mathematics and Computer Science

Delft University of Technology

Rotterdam, the Netherlands

July, 2009

Acknowledgment

There are many people I would like to thank, in particular Kees Oosterlee, who not only serves as my supervisor during the past nine months, but also encourages and challenges me throughout the period of my thesis project. The same thanks should also go to my second supervisor David van Bragt of Ortec Finance, who take efforts in guiding me trough the whole thesis project. I would also like to thank Dirk-Jan Kort of Ortec Finance, who helped me with some programming details. I am also grateful to Alexander Haastrecht for answering my questions about some topics discussed in this thesis. Furthermore, I want to thank the other committee members; Hans van der Weide and R.J. Fokkink. Last but not least, great love goes out to my parents and my girlfriend who have supported me during the thesis project.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Current situation

Market valuation of insurance liabilities is becoming more and more important in the insurance and pension industry. This development is mainly due to new accounting standards (IFRS phase II) and new regulatory frameworks (Solvency II). Until recent years, liabilities were often valued at book value, which means that future cash-flows are discounted with a fixed interest rate. Although this approach is conceptually simple and easy to apply, it is not consistent with the current market situation. Therefore, more and more pension and insurance companies want to value both their assets and liabilities at market value. For an insurance company this is a complex task since the liabilities contain embedded options. Embedded options are rights in insurance policies that can provide a profit to policy holders but never a loss. These options are characterized by an asymmetric pay-off structure (see for more information, for example, Steehouwer and Van Bragt (2007)).

Due to the presence of these options, the valuation at market value is complex, because time consuming and complex valuation techniques have to be applied. In practice, Monte Carlo (MC) simulations are often used for valuation. The advantage of this method is that different kinds of options can be valued and that it gives a uniform simulation framework wherein several options can be priced at once. Despite the flexibility of the MC method, there are some practical drawbacks. The main drawback is the computation time which is required to achieve an acceptable error level, i.e. the convergence rate is low. For risk management or reporting purposes many valuations have to be made which result in even higher computation times.

For the generation of risk-neutral scenarios, which are needed in a MC framework, basically two methods can be used. On the one hand one can use a so-called deflator approach, which is a stochastic discount factor that can directly be applied to real world scenarios and, which is beneficial in an Asset and Liability Management (ALM) framework (see Hoevenaars (2008)). The main drawback of this method is that the calibration of the stochastic discount factor is difficult. On the other hand, risk-neutral (arbitrage-free) models can be used. The advantage of these models is that the generation of risk-neutral scenarios and the discounting of future cash flows can be done relatively easy. Therefore, we only consider risk-neutral models in this thesis.

Nowadays, extensions of the (standard) Black-Scholes (BS) model become very important in several important models in finance, such as the Heston model (see Heston (1993)). This is an extension of the BS model by modeling the variance process by a Cox-Ingersoll-Ross (CIR) process. Using this model the computation of European option prices can be done in a numerically stable and rather efficient way (see for example Fang and Oosterlee (2008)). This allows for fast calibration to market or historical data, which is one of the reasons why there is much interest in embedding the Heston model in derivative pricing models, as a means to capture the volatility smiles and skews, which are present in option market data. Because of the importance of the Heston model in option pricing theory, this model is of interest in this thesis.

In recent time several papers have emerged dealing with the efficient simulation of the Heston model in a MC framework. See for example Broady and Kaya (2006), Lord and Koekkoek (2008), Anderson (2007), Haastrecht and Pelsser (2008), Zhu (2008) and Glasserman and Kim (2008). These papers show that derivative pricing using the (standard) Euler-Maclaurin (Euler) discretization method can be improved significantly, i.e. obtaining the same accuracy in

significantly fewer time steps (only sample at important times). Although much attention is paid to stochastic variance by using the Heston model, no attention has been paid yet to the extension of stochastic interest rates, which is of particular interest in this thesis too.

1.2 Objectives of the project

As an introduction, we would like to remark that the use of risk-neutral MC simulations for valuing embedded options, involves two errors:

- A discretization error due to the discretization of the underlying continuous model[1]. The discretization error is controlled by the number of time steps;
- A statistical error due to the finite samples in the MC estimator. The statistical error is controlled by the sample length.

In order to obtain more accurate results in less computation time, one has to make use of state-of-the-art discretization schemes to minimize the discretization error and variance reduction techniques to minimize the statistical error. More information on discretization schemes can be found in Kloeden and Platen (1999) and Kloeden and Platen *et al.* (1994). A classical textbook about MC methods is Glasserman (2003).

As a starting point we assume that the valuation of (embedded) options is done by a two-factor Hull-White Black-Scholes (HWBS) model in combination with the MC method. The HWBS model is an extension of the (standard) BS model with stochastic interest rate driven by a Hull-White (HW) model (see Brigo and Mercurio (2006) for more information). Once the parameters of the model are calibrated, scenarios are generated using the (standard) Euler discretization scheme.

Since we wish to value (embedded) options as fast as possible on the basis of a high-dimensional underlying model, the current valuation method is too time consuming. Computation time can be saved in two ways. On the one hand we can improve the discretization method so that the discretization error is (much) less for a given computational budget. On the other hand we can construct an appropriate combination of variance reduction techniques so that the statistical error of the MC is reduced and, by construction, more accurate results are obtained in less computation time.

The objectives of this thesis are as follows:

- Extend the two-factor HWBS risk-neutral model by a stochastic variance model driven by a CIR process so that the asset price is generated by a Heston model extended with stochastic interest rates generated by a two-factor HW model. Throughout the thesis we call this particular model the extended Heston model;
- Investigate the distributional properties of the underlying variables, i.e. of the interest rate, the variance and the asset price process;
- Make an inventory of existing simulation schemes;
- Propose a discretization scheme for the extended Heston model, i.e. propose a discretization scheme for the interest rate process, the variance process and the asset price process;

[1]This is only the case when a discretization scheme is used. There is no discretization error when an unbiased estimator of the underlying model is used.

- Compare the proposed discretization schemes with the Euler discretization scheme by comparing statistics and by valuing (embedded) options;
- Investigate which variance reduction methods are most applicable to value embedded options using the extended Heston model;

1.3 Organization of the rapport

We will first explain in Chapter 2 in which context the thesis is written, i.e. we give some general background information about ALM and the importance of risk-neutral valuation for pension and insurance companies.

Chapter 3 deals with the (underlying) model description, which we call, as already mentioned, the extended Heston model. We explain why we use this particular underlying model for valuation and simulation, and we outline the basic distributional properties of the model.

Once the model specifications and distributional properties have been presented, we discuss in Chapter 4 several simulation schemes, which are used for efficient path-simulation. We briefly introduce existing discretization schemes from the literature; some of them are used in later comparative experiments. A simulation scheme of the two-factor HW model is constructed and we consider a new scheme to discretize the variance process, where we pay special attention to computational issues. We discuss how to discretize the asset price process and, finally, we also discuss a martingale correction method, empirical martingale simulation (EMS) and some convergence properties.

In Chapter 5 we briefly discuss a number of variance reduction techniques, which are of interest for valuing embedded options. Some of these techniques or combinations of these techniques are used in numerical experiments.

In Chapter 6 the calibration procedure is briefly discussed. Due to the complexity of the considered underlying model, we use MC simulations for calibration in combination with a control variate technique. In particular, a control variate is constructed and used to improve the MC simulations, so that more reliable estimates can be obtained.

Chapter 7 contains the numerical results. On the one hand we consider the Heston model and test our new discretization scheme to existing discretization schemes. On the other hand we consider the extended Heston model and perform numerical experiments on comparing the proposed discretization scheme with the (standard) Euler scheme. We also show the effect of variance reduction techniques on the accuracy of pricing options. We present these results by means of illustrations and tables. In the end, we measure the gain in computation time over the current approach to value embedded options.

Finally, the last chapter contains the overall conclusions and directions for future research. This includes a discussion of the objectives and, based on the numerical results, recommendations are given.

2 Importance of risk-neutral valuation for insurance and pension companies

This chapter contains general information about the field of study for which the research for this MSc thesis, is done. We especially include more information about the areas for which risk-neutral valuation plays an important role. As we already mentioned, the liabilities of a life-insurer contain embedded options, which need to be valued at market value. Thus, life-insurers intensively use risk-neutral valuation techniques in order to obtain the market value of their liabilities. A pension fund also has to make decisions about indexation and the height of the premium. A relatively new way of doing this is by using 'Value Based ALM'. This is a market-consistent valuation of premiums and indexations for which risk-neutral scenarios are also needed. For more information about Value based ALM we refer to Hoevenaars (2008) and Kocken (2006).

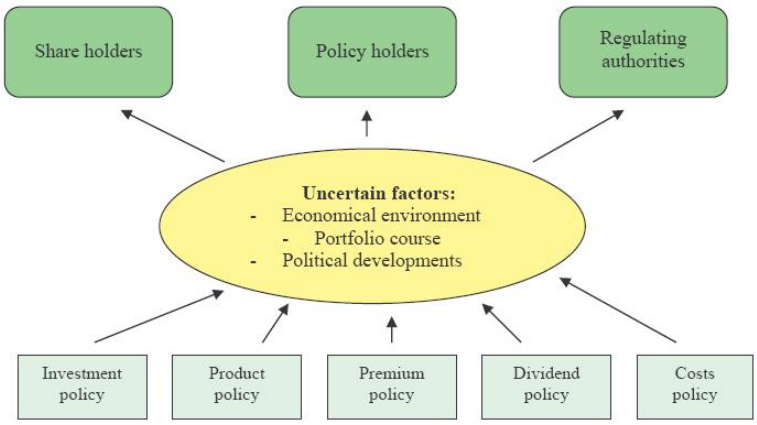
2.1 Asset and liability management

ALM is a risk management tool, which is intensively used by pension funds, banks, insurers, wealth planners and housing corporations. ALM models analyze the balance sheet and optimize the trade off between risk and return given the objectives and constraints of the stakeholders.

In the first place, ALM models are used for advisory and support to determine several policies, such as premium, investments and indexation. Besides, ALM models provide insight, transparency and responsibility. ALM models often consist of a scenario approach. This includes the generation of scenarios of several (economic) variables over time such that decisions can be made based on these scenarios.

Figure 1 shows an ALM decision diagram for an life-insurance company.

Figure 1: ALM decision diagram



First of all, the objectives of the stakeholders of a life-insurer are important in a ALM framework. For example, the shareholders wish to have stable returns on their investments, the policyholders wish to have high profit sharing and low premiums and the supervisor of a life-insurer wishes a high and stable solvency ratio[2]. We remark that the stakeholders have different objectives. For example, investing in risky assets provides for higher expected returns, which is preferable

[2]The solvency ratio of a life-insurer is defined as the ratio between the market value of the assets minus the market value of the liabilities and the required capital.

for the shareholders, however because the solvency ratio risks are high this is unacceptable for the supervisor.

Secondly, the policy instruments, which consists of (among others), investment policy, dividend policy, costs policy, premium policy and product policy, are used by the insurer to optimize the trade off between risk and return given the objectives and constraints of the stakeholders.

Thirdly, an insurer deals with external uncertainties in which decisions have to be made. These uncertainties have an influence on both the future development of economic variables, such as inflation, interest rates and returns on different asset categories, and on the future development of the new business of an (life) insurer.

More information about ALM can be found in Zenios and Ziemba (2006), Zenios and Ziemba (2007) and Steehouwer and Van Bragt (2007).

2.2 Embedded options

In this chapter we discuss the embedded option considered in this thesis in more detail. We especially focus on the pay-off structure and other characteristics such as time to maturity.

In general, embedded options can be categorized in profit sharing, guarantees and behavioral options. In the profit sharing category, the policy holder benefits from high interest rates, but does not pay extra in case of low interest rates. In the guarantees category, an insurance company pays in case of low investment returns but receives nothing in case of high investment returns. In case of behavioral options, the policy holder has the right to perform actions that an insurance company cannot decline, for example, the policy holder can exercise the option at an early time.

Embedded options are often complex because they often depend on several (economic) variables and have very long maturities (typically $50 - 100$ years). These (economic) variables are often total return categories, such as stocks, fixed income and commodities, inflation and real and nominal interest rates. While there are many different kinds of embedded options, we only discuss the profit sharing embedded option, because this embedded option is used in numerical experiments.

Profit sharing embedded option

Traditional insurance products guarantee a certain insured amount. This is often combined with profit sharing. In this case a reference return is paid out to the policy holder if this return exceeds a prespecified threshold interest rate. Various types of profit sharing options exist, but in most cases the profit sharing depends on the yield of a basket of representative bonds. Since the exact profit sharing is nontrivial, because it depends on baskets of government bonds, the valuation of profit sharing options is often done using an (average) forward swap rate as an approximation of the profit sharing rate. The main reason is that the market of swaps is much more complete than the market of government bonds and, as a consequence, more market data for calibration is available.

The time t price of a profit sharing embedded option is given by:

$$\Pi_{ORA}(t,T) = \mathbb{E}\left[e^{-\int_t^T r_N(u)du}\, TR(t) \max\left(c\left(PSR(t) - K(t)\right), 0\right)\right],$$

where $r_N(t)$ is the nominal short interest rate at time t, $TR(t)$ denotes the technical reserve, c is the percentage that is distributed to the policy holder, $K(t)$ is the strike level of the option and $PSR(t)$ is the profit sharing rate at time t. The latter is often a (weighted) average of historic and forward swap rates.

In this thesis we consider the most common form of profit sharing in Holland, which is based on a moving average of the so-called u-rate. The u-rate is a 3-months average of u-rate-parts, where the subsequent u-rate-parts are weighted averages of an effective return on a basket of government bonds. This leads to complicated expressions. For that reason, the u-rate is approximated by a swap-rate[3]. Thus, the profit sharing rate is approximated by a moving average of swap-rates, i.e.

$$PSR(t) = \frac{1}{n} \sum_{j=t-n+1}^{t} y_{j,j+m}(j),$$

where $y_{t,t+m}(t)$ is the m-year swap rate at time t. For more information on this topic we refer to Plat and Pelsser (2008).

[3]It turns out that the 7-year swap rate is a good approximation, see (Plat and Pelsser (2008))

3 Model description and distributional properties

In this chapter we discuss the risk-neutral model, which is of interest in this thesis. We pay special attention to distributional properties because those are of importance for efficient implementation of statistics and to set up an efficient simulation scheme, which is the topic of Chapter 4.

3.1 Model description

Traditionally, the asset price and inflation dynamics are modeled by the Black-Scholes (BS) model. However, the BS model does not represent reality well enough because the interest rate and variance are kept constant over future time. Many empirical studies have shown that the interest rate and variance are not constant over future time, but also follow a stochastic process. For that important reason we model the interest rate and variance of the asset price and the inflation as stochastic processes. We model the interest rate by a two-factor Hull-White (HW) model and the variance process of the asset price and the inflation is modeled by a Cox-Ingersoll-Ross (CIR) model and call the resulting model the extended Heston model[4].

The Heston model captures the equity volatility smile well, which implies that there is a good fit with market data for option prices. For interest rate options we use the two-factor HW model. This model is not able to capture the volatility smile sufficiently well since the variable is normally distributed. The reason why we still use this particular model is that many analytical results can be derived which is beneficial from a computational point of view.

Consequently, the model of the asset price and inflation dynamics consists of a four-dimensional SDE. The dynamics for the inflation and asset price are slightly different but can be handled by the same approach. Therefore, we only discuss the dynamics of the asset price. The dynamics of the whole model are given in Appendix A.

Assuming the probability space is given by $(\Omega, \mathcal{F}, \mathbb{P})$, the extended Heston model is defined by the coupled four-dimensional SDE:

$$\begin{array}{lcll} dS(t) & = & (r(t)-\delta)\,S(t)dt + \sqrt{\nu_S(t)}S(t)dW^S(t) & S(0) \geq 0 \\ d\nu(t) & = & \kappa\left(\xi - \nu(t)\right)dt + \sigma_\nu\sqrt{\nu(t)}dW^\nu(t) & \nu(0) \geq 0 \\ dr(t) & = & \left(\theta(t) + u(t) - ar(t)\right)dt + \sigma_r dW^r(t) & r(0) \geq 0 \\ du(t) & = & -bu(t)dt + \sigma_u dW^u(t) & u(0) = 0, \end{array}$$

where at time t, $S(t)$ denotes the asset price, $r(t)$ the short interest rate, $u(t)$ the second interest rate factor and $\nu(t)$ denotes the variance process of the asset price. Furthermore, a, b, and κ are mean reversion parameters of the two-factor HW model and of the Heston model, respectively and ξ denotes the long term mean of the variance process. The function $\theta(t)$ fits the short interest rate to the initial term structure. Next, σ_ν, σ_r and σ_u are volatility parameters of the variance process and of the two-factor HW model, respectively. We have modeled the continuous flow of dividends by a decrease of the stock price in each time interval dt by an amount $\delta S(t)dt$, where $\delta \geq 0$. Finally, $\left(dW^S(t), dW^\nu(t), dW^r(t), dW^u(t)\right)$ are correlated Wiener processes with correlation $\rho_{i,j}$, where $i, j \in \{S, \nu, r, u\}$ and $i = j$ implies $\rho_{i,j} = 1$. The correlation matrix is a symmetric and positive semi-definite matrix, i.e. for all $\mathbf{x} \in \mathbb{R}^4$ the correlation matrix Σ has to satisfy $\mathbf{x}^{tr}\Sigma\mathbf{x} \geq 0$.

[4]Note that the (well-known) Heston model is based on of a fixed interest rate, see Heston (1993).

An important property of this set of SDE's is that the solution of the single SDE's are Markov process (see Shreve (2006)). In other words, the following equalities hold

$$\begin{aligned}\mathbb{E}\left[f(S(t)) \mid \mathcal{F}_s\right] &= g(S(s)), \qquad \mathbb{E}\left[f(\nu(t)) \mid \mathcal{F}_s\right] = g(\nu(s)),\\ \mathbb{E}\left[f(r(t)) \mid \mathcal{F}_s\right] &= g(r(s)), \qquad \mathbb{E}\left[f(u(t)) \mid \mathcal{F}_s\right] = g(u(s)),\end{aligned}$$

where $0 \leq s < t \leq T$ and f and g are two continuous functions. Thus, the processes of interest only depend on their last generated/observed value instead of the whole history, $\mathcal{F}_s$.

3.2 Benchmark setup

Because the CIR process is of particular interest in this thesis, we perform several numerical experiments. For these experiments we use the model parameter settings in Table 1 for the CIR process, which are also used by Glasserman and Kim (2008):

Table 1: Model parameters for the CIR model

	case I	case II	case III	case IV
κ	0.5	0.3	1	6.2
ξ	0.04	0.04	0.09	0.02
σ_ν	1	0.9	1	0.6
$\rho_{S,\nu}$	-0.9	-0.5	-0.3	-0.7

According to Glasserman and Kim (2008), case I is relevant for long-dated foreign exchange (FX) options, case II for long-dated interest rate options, case III for equity options and case IV for S&P 500 index options. According to Andersen (2007), cases I-III are challenging as well as practically relevant.

3.3 Consistency of the Hull-White interest rate model

The time-dependent parameter $\theta(t)$ fits the two-factor HW model to the initial term structure, which implies that the model is arbitrage free. The function $\theta(t)$ defines the average direction in which the interest rate moves at time t, which is independent of the level of the interest rate and can be calculated from the initial term structure. In general, $\theta(t)$ has the following form

$$\theta(t) = \frac{\partial f(0,t)}{\partial T} + af(0,t) + \frac{\partial \phi(0,t)}{\partial T} + a\phi(0,t), \tag{1}$$

where $f(0,T)$ denotes the market instantaneous forward rate at time 0 with maturity T from the initial nominal (continuously compounded) zero coupon yield curve. The function $\phi(0,T)$ is determined in such a way that the function $\theta(t)$ fits the theoretical bond prices to the yield curve observed in the market.

Usually $\frac{\partial \phi(0,t)}{\partial t} + a\phi(0,t)$ is relatively small. If we ignore these terms, equation (1) implies that the drift of the process for r at time t is $\frac{\partial f(0,t)}{\partial t} + a(f(0,t) - r) + u(t)$. This shows that, on average, r follows the slope of the initial instantaneous forward rate curve with disturbance $u(t)$. When the interest rate deviates from that curve, it reverts back to it with rate a.

Result 3.1. *In the case of a one- or two-factor HW model, $\phi(t,T)$ has an analytic expression*[5]*. For the one-factor HW model we have*

$$\phi(t,T) = \frac{\sigma_r^2}{2a^2}\left(1-e^{-a(T-t)}\right)^2, \qquad \frac{\partial\phi(t,T)}{\partial T} + a\phi(t,T) = \frac{\sigma_r^2}{2a}\left(1-e^{-2a(T-t)}\right) \tag{2}$$

and for the two-factor HW model we have

$$\phi(t,T) = \frac{1}{2}\sigma_r^2 B(t,T)^2 + \frac{1}{2}\sigma_u^2 C(t,T)^2 + \rho_{r,u}\sigma_r\sigma_u B(t,T)C(t,T), \tag{3}$$

where $B(t,T) = \frac{1}{a}\left(1-e^{-a(T-t)}\right)$, $C(t,T) = \frac{1}{b}\left(1-e^{-b(T-t)}\right)$. This results in

$$\begin{aligned}\frac{\partial\phi(0,t)}{\partial t} + a\phi(0,t) &= \gamma_1\left(\gamma_2\left(3-4\,e^{a\,t}+e^{2\,a\,t}\right)\sigma_r{}^2 + e^{2\,a\,t}\left(\gamma_3 + b\,e^{a\,t} + \gamma_4\,e^{b\,t}\right)\right.\\ &\quad \left(b - b\,e^{a\,t} + a\left(-1+e^{b\,t}\right)\right)\sigma_u{}^2 + \gamma_5 e^{a\,t}\left(\left(b\left(-1+e^{a\,t}\right)\left(-2+e^{b\,t}\right)\right)\right.\\ &\quad \left.\left.+a\left(-2+e^{a\,t}\right)\left(-1+e^{b\,t}\right)\right)\gamma_6\right),\end{aligned}$$

where

$$\gamma_1 = \frac{1}{2\,a\,(a-b)^2\,b^2\,e^{2\,a\,t}}, \quad \gamma_2 = (a-b)^2\,b^2, \quad \gamma_3 = b-a, \quad \gamma_4 = (a-2b),$$
$$\gamma_5 = 2\,(a-b)\,b, \quad \text{and} \quad \gamma_6 = \sigma_r\sigma_u\rho_{r,u}.$$

The annually compounded zero coupon curve, $R_A(t,T)$, can be fitted exactly by a spline function, but one can also use, for example, a Nelson and Siegel function (see Nelson and Siegel (1987)) to model $R_A(t,T)$.

Since we deal with a continuous interest rate model, we need, however, the continuously compounded zero coupon curve at time t, which is easily derived by

$$R_C(t,T) = \log\left(1+R_A(t,T)\right),$$

with maturity time T. Moreover, when the forward maturity in the continuously compounded forward rate tends to zero, we have $f(0,t) = R_C(0,t) + t\frac{\partial R_C(0,t)}{\partial t} = \log\left(1+R_A(0,t)\right) + t\frac{R_A'(0,t)}{1+R_A(0,t)}$. More information can be found in Bakker (2006).

3.4 Distributional properties of the two-factor HW model

Although this model is discussed in Brigo and Mercurio (2006), the distributional properties are not derived directly for the two-factor HW model, but in terms of the two-factor Gaussian model (see Brigo and Mercurio (2006, Chapter 4)). In order to be complete, we derived the distributional properties of the two-factor HW model, which are given below.

Result 3.2. *Conditional on $\mathcal{F}_s$, $r(t)$ is normally distributed, with expectation and variance given by,*

$$\begin{aligned}\mathbb{E}\left[r(t)\,|\,\mathcal{F}_s\right] &= r(s)e^{-a(t-s)} + \int_s^t \theta(u)e^{-a(t-u)}du,\\ Var\left(r(t)\,|\,\mathcal{F}_s\right) &= \frac{\sigma_r^2}{2a}\left(1-e^{-2a(t-s)}\right) + \frac{\sigma_u^2}{(a-b)^2}\left(\frac{1-e^{-2b(t-s)}}{2b} - 2\frac{1-e^{-(a+b)(t-s)}}{a+b}\right.\\ &\quad \left.+\frac{1-e^{-2a(t-s)}}{2a}\right) + \frac{2\rho_{r,u}\sigma_r\sigma_u}{a-b}\left(\frac{1-e^{-(a+b)(t-s)}}{a+b} - \frac{1-e^{-2a(t-s)}}{2a}\right),\end{aligned}$$

[5]These expressions can also be found in Brigo and Mercurio (2006).

where $0 \leq s < t \leq T$.

This result is derived by extending some results of Brigo and Mercurio (2006, Chapter 4).

Remark. *Using Result 3.2 we note that when we assume* $s = 0$ *and* t *tending to infinity, the variance of the second interest rate factor and of the interest rate tend to* $\frac{\sigma_u^2}{2b}$ *and* $\frac{\sigma_r^2}{2a} - \frac{\sigma_u^2}{2ab(a+b)} + \frac{2\rho_{r,u}\sigma_r\sigma_u}{2a(a+b)}$, *respectively.*

In particular, we have the following new result:

Result 3.3. *It turns out that conditional on* $\mathcal{F}_s$, $u(t)$ *and* $(r(t) \mid u(s) = u_s)$ *are normally distributed, where* $0 \leq s < t \leq T$. *More specifically,*

$$(u(t) \mid \mathcal{F}_s) \sim N\left(u(s)e^{-b(t-s)}, \frac{\sigma_u^2}{2b}\left(1 - e^{-2b(t-s)}\right)\right), \tag{4}$$

$$(r(t) \mid \mathcal{F}_s, u(s) = u_s) \sim N\left(r(s)e^{-a(t-s)} + \int_s^t \widetilde{\theta}(u)e^{-a(t-u)}du, \frac{\sigma_r^2}{2a}\left(1 - e^{-2a(t-s)}\right)\right), \tag{5}$$

where $\widetilde{\theta}(t) = \theta(t) + u_s$.

Remark. *Note that since the one and two-factor HW model are Markov processes, the processes (only) depend on* $u(s)$ *and* $r(s)$ *in case of conditioning on* $\mathcal{F}_s$.

The integral appearing in the expectation of the interest rate can be computed analytically, which we prove in the following new Lemma.

Lemma 3.1. *Conditional on* $\mathcal{F}_s$, *we have the following result:*

$$\begin{aligned}\widetilde{\varphi}(t) &= \varphi(t) + \int_s^t u_s e^{-a(t-u)}du = r(s)e^{-a(t-s)} + \int_s^t \theta(u)e^{-a(t-u)}du + \int_s^t u_s e^{-a(t-u)}du, \\ &= r(s)e^{-a(t-s)} + f(0,t) + \phi(0,t) - \left(f(0,s) + \phi(0,s)\right)e^{-a(t-s)} + \frac{u_s}{a}\left(1 - e^{-a(t-s)}\right),\end{aligned}$$

where the function $\phi(0,t)$ *is given by (2), when a one-factor HW model is used and by (3), when a two-factor HW model is used. In particular, when we condition on initial time, then*

$$\varphi(t) = f(0,t) + \phi(0,t).$$

Proof. First we compute

$$\int_s^t u_s e^{-a(t-u)}du = \left[\frac{u_s}{a}e^{-a(t-u)}\right]_{u=s}^t = \frac{u_s}{a}\left(1 - e^{-a(t-s)}\right).$$

Next, by a basic application of integration by parts we obtain an expression for $\varphi(t)$:

$$\begin{aligned}\varphi(t) &= r(s)e^{-a(t-s)} + \int_s^t \theta(v)e^{-a(t-v)}dv \\ &= r(s)e^{-a(t-s)} + \int_s^t \left(\frac{\partial f(0,v)}{\partial v} + af(0,v)\right)e^{-a(t-v)}dv \\ &+ \int_s^t \left(\frac{\partial \phi(0,v)}{\partial v} + a\phi(0,v)\right)e^{-a(t-v)}dv \\ &= r(s)e^{-a(t-s)} + \left[f(0,v)e^{-a(t-v)}\right]_{v=s}^t + \left[\phi(0,v)e^{-a(t-v)}\right]_{v=s}^t \\ &= r(s)e^{-a(t-s)} + f(0,t) + \phi(0,t) - f(0,s)e^{-a(t-s)} - \phi(0,s)e^{-a(t-s)}. \end{aligned} \tag{6}$$

Combining these two results gives the desired result. Depending on which factor model is used, either (2) or (3) is used to compute $\phi(0,t)$. The special case is obtained by setting $s = 0$, which implies $\phi(0,0) = 0$, $\lim_{\beta \to 0} f(0,\beta) = r(0)$ and $u(0) = 0$ and the desired result follows directly. □

3.5 Transformation to the Gaussian interest rate model

In general, it is convenient to transform the two-factor HW model to the two-factor Gaussian interest rate model, because this model has a simpler form. Note that when one only considers the one-factor HW model this transformation is not needed since the distributional properties are easily derivable. In the case of the one-factor model, it turns out that $r(t)$, conditional on $\mathcal{F}_s$, with $0 \leq s < t \leq T$, is normally distributed with mean and variance given by

$$\mathbb{E}\left[r(t) \mid \mathcal{F}_s\right] = r(s)e^{-a(t-s)} + \alpha(t) - \alpha(s)e^{-a(t-s)}, \qquad Var\left(r(t) \mid \mathcal{F}_s\right) = \frac{\sigma_r^2}{2a}\left(1 - e^{-2a(t-s)}\right),$$

where $\alpha(t) = f(0,t) + \frac{\sigma_r^2}{2a^2}\left(1 - e^{-at}\right)^2$ and $0 \leq s < t \leq T$.

The two-factor Gaussian model is given by

$$\begin{aligned} r(t) &= x(t) + y(y) + \varphi(t) && (r(0) = \varphi(0)) \\ dx(t) &= -b_x x(t)dt + \sigma_x dW^x(t) && (x(0) = 0) \\ dy(t) &= -b_y y(t)dt + \sigma_y dW^y(t) && (y(0) = 0), \end{aligned}$$

where conditional on $\mathcal{F}_s$, $\varphi(t) = r(s)e^{-a(t-s)} + \int_s^t \theta(u)e^{-a(t-u)}du$, with $\varphi(0) = \lim_{\beta\to 0} f(0,\beta) = r(0)$, b_x and b_y are mean reversion parameters, σ_x and σ_y are volatility parameters and $(dW^x(t), dW^y(t))$ are Brownian motion processes with instantaneous correlation $\rho_{x,y}$. The parameters of the two-factor Gaussian interest rate model are obtained from the two-factor HW model by (see Brigo and Mercurio (2006, Chapter 4))

$$\begin{aligned} b_x &= a, \quad b_y = b, \quad \sigma_x = \sigma_1, \quad \sigma_y = \sigma_2, \\ \rho_{x,y} &= \frac{\sigma_r\rho_{r,u} - \sigma_1}{\sigma_2}, \qquad \varphi(t) = r(0)e^{-at} + \int_0^t \theta(u)e^{-a(t-u)}du, \end{aligned}$$

where $\sigma_1 = \sqrt{\sigma_r^2 + \frac{\sigma_u^2}{(a-b)^2} + 2\rho_{r,u}\frac{\sigma_r\sigma_u}{b-a}}$, $\sigma_2 = \frac{\sigma_u}{a-b}$.

Note that the SDE's for $x(t)$ and $y(t)$ are of the same simple form with solutions,

$$x(t) = x(s)e^{-b_x(t-s)} + \sigma_x\int_s^t e^{-b_x(t-u)}dW^x(u),$$
$$y(t) = y(s)e^{-b_y(t-s)} + \sigma_y\int_s^t e^{-b_y(t-u)}dW^y(u),$$

where $0 \leq s < t \leq T$. Thus, conditional on $\mathcal{F}_s$, $x(t)$ and $y(t)$ are normally distributed with mean and variance, respectively, equal to

$$\mathbb{E}\left[x(t) \mid \mathcal{F}_s\right] = x(s)e^{-b_x(t-s)}, \qquad \mathbb{E}\left[y(t) \mid \mathcal{F}_s\right] = y(s)e^{-b_y(t-s)}, \tag{7}$$

$$Var\left(x(t) \mid \mathcal{F}_s\right) = \frac{\sigma_x^2}{2b_x}\left(1 - e^{-2b_x(t-s)}\right), \qquad Var\left[y(t) \mid \mathcal{F}_s\right] = \frac{\sigma_y^2}{2b_y}\left(1 - e^{-2b_y(t-s)}\right), \tag{8}$$

with $0 \leq s < t \leq T$. Working out the distributional properties of the interest rate, we obtain

$$\mathbb{E}\left[r(t) \mid \mathcal{F}_s\right] = x(s)e^{-b_x(t-s)} + y(s)e^{-b_y(t-s)} + \varphi(t), \tag{9}$$

$$\begin{aligned} Var\left(r(t) \mid \mathcal{F}_s\right) &= \frac{\sigma_x^2}{2b_x}\left(1 - e^{-2b_x(t-s)}\right) + \frac{\sigma_y^2}{2b_y}\left(1 - e^{-2b_y(t-s)}\right) \\ &\quad + 2\rho_{x,y}\frac{\sigma_x\sigma_y}{b_x + b_y}\left(1 - e^{-(b_x+b_y)(t-s)}\right). \end{aligned} \tag{10}$$

3.6 Derived analytic correlation of Gaussian interest rate model with asset price and the variance process

When we transform the HW model to the Gaussian model, the correlation parameters of the interest rate model with the asset price and variance process also change. I.e. we wish to use the information of

$$d\begin{pmatrix} dW^S \\ dW^\nu \\ dW^r \\ dW^u \end{pmatrix} d\left(dW^S, dW^\nu, dW^r, dW^u\right) = \begin{pmatrix} 1 & . & . & . \\ \rho_{S,\nu} & 1 & . & . \\ \rho_{S,r} & \rho_{\nu_S,r} & 1 & . \\ \rho_{S,u} & \rho_{\nu_S,u} & \rho_{r,u} & 1 \end{pmatrix} dt$$

to obtain

$$d\begin{pmatrix} dW^S \\ dW^\nu \\ dW^x \\ dW^y \end{pmatrix} d\left(dW^S, dW^\nu, dW^x, dW^y\right) = \begin{pmatrix} 1 & . & . & . \\ \rho_{S,\nu} & 1 & . & . \\ \gamma_{S,x} & \gamma_{\nu,x} & 1 & . \\ \gamma_{S,y} & \gamma_{\nu,y} & \rho_{x,y} & 1 \end{pmatrix} dt,$$

where $\gamma_{S,x}$, $\gamma_{S,y}$, $\gamma_{\nu,x}$ and $\gamma_{\nu,y}$ are unknown. In the following new Lemma we derive the analytical correlations between the two-factor Gaussian interest rate model with the asset price and the variance process.

Lemma 3.2. *When we transform the two-factor HW model to the two-factor Gaussian model then the correlations, $\rho_{S,x}$, $\rho_{S,y}$, $\rho_{\nu,y}$, $\rho_{\nu,y}$ are given by*

$$\rho_{S,x} = \frac{\sigma_r\rho_{S,r} - \sigma_2\rho_{S,u}}{\sigma_1}, \qquad \rho_{S,y} = \rho_{S,u}, \qquad \rho_{\nu,x} = \frac{\sigma_r\rho_{\nu,r} - \sigma_2\rho_{\nu,u}}{\sigma_1}, \qquad \rho_{\nu,y} = \rho_{\nu,u}.$$

Proof. From Brigo and Mercurio (2006, Chapter 4) we have that $dW^x(t) = \frac{\sigma_r}{\sigma_1}dW^r(t) + \frac{\sigma_2}{\sigma_1}dW^u(t)$ and $dW^y(t) = \sigma_2 dW^u(t)$. Using these expressions, the correlations can be easily worked out. To be complete, the derivations are given by

$$\begin{aligned} \rho_{S,x} &= \frac{\frac{\sigma_r}{\sigma 1}\rho_{S,r} - \frac{\sigma_2}{\sigma 1}\rho_{S,u}}{\sqrt{\left(\frac{\sigma_r}{\sigma 1}\right)^2 + \left(\frac{\sigma_2}{\sigma 1}\right)^2 + 2\rho_{r,u}\frac{\sigma_r\sigma_2}{\sigma_1^2}}} = \frac{\sigma_r\rho_{S,r} - \sigma_2\rho_{S,u}}{\sigma_1}, \\ \rho_{\nu,x} &= \frac{\frac{\sigma_r}{\sigma 1}\rho_{\nu,r} - \frac{\sigma_2}{\sigma 1}\rho_{\nu,u}}{\sqrt{\left(\frac{\sigma_r}{\sigma 1}\right)^2 + \left(\frac{\sigma_2}{\sigma 1}\right)^2 + 2\rho_{r,u}\frac{\sigma_r\sigma_2}{\sigma_1^2}}} = \frac{\sigma_r\rho_{\nu,r} - \sigma_2\rho_{\nu,u}}{\sigma_1}, \end{aligned}$$

where we have used the definition of correlation, $Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_x\sigma_y}$ for two random variables X and Y. The correlations of the asset price and the variance level with the second interest rate factor are easily derived since a constant times a random variable does not affect the correlation structure, i.e. $Corr(X,aY) = \frac{Cov(X,Y)}{\sigma_x\sigma_y}$ for $a \in \mathbb{R}$. □

3.7 Analysis and distributional properties of the variance process

Next, we consider the variance process, which is modeled by a CIR process. The dynamics of the CIR process are significantly more difficult than the dynamics of the interest rate. The density function of the variance process is just as the interest rate process known in advance. Using the Fokker-Planck equation[6] for the CIR model, which is given by the following partial differential equation

$$\frac{\partial}{\partial t}f(t,x) = -\frac{\partial}{\partial x}\left(\kappa(\xi - x)f(t,x)\right) + \frac{\sigma_\nu^2}{2}\frac{\partial^2}{\partial^2 x}\left(xf(t,x)\right) \qquad \left(f(0,x) = \delta(x)\right), \tag{11}$$

[6]The Fokker-Planck equation describes the time evolution of the probability density function.

the probability density function, $f(t,x)$, of the variance process, where $\delta(\cdot)$ is the delta function, can be derived.

It turns out that the non-stationary solution, conditional on $\mathcal{F}_s$ $(0 \leq s < t \leq T)$, of the above partial differential equation is a constant times a non-central chi-squared distribution. More specifically, the variance is distributed as a constant, C, times a non-central chi-squared distribution with d degrees of freedom and non-centrality parameter λ, i.e.

$$\mathbb{P}\left(\nu(t) \leq x \mid \nu(s)\right) = F_{nccs}\left(\frac{x}{C}; d, \lambda\right),$$

where $F_{nccs}(x)$ represents the cumulative distribution of the non-central chi-squared distribution, which is given below,

$$F_{nccs}(x; d, \lambda) = \sum_{j=1}^{\infty} e^{-\frac{\lambda}{2}} \frac{(\frac{\lambda}{2})^j}{j!} \frac{\Gamma(j + \frac{d}{2}, \frac{x}{2})}{\Gamma(j + \frac{d}{2})} \qquad \text{and} \tag{12}$$

$$f_{nccs}(x; d, \lambda) = \frac{1}{2} e^{-\frac{1}{2}(x+\lambda)} \left(\frac{x}{\lambda}\right)^{\frac{1}{2}(\frac{d}{2}-1)} B_{\frac{d}{2}-1}\left(\sqrt{\lambda x}\right), \tag{13}$$

where $f_{nccs}(x; d, \lambda)$ denotes the corresponding probability density function. The parameters C, d, λ are given by:

$$C = \frac{\sigma_\nu^2\left(1 - e^{-\kappa(t-s)}\right)}{4\kappa}, \quad d = \frac{4\kappa\xi}{\sigma_\nu^2}, \quad \lambda = \frac{4\kappa e^{-\kappa(t-s)}\nu(s)}{\sigma_\nu^2\left(1 - e^{-\kappa(t-s)}\right)}.$$

Here $\Gamma(y,x) = \int_0^x t^{y-1}e^{-t}dt$ denotes the incomplete gamma function, $\Gamma(x)$ the gamma function and $B_x(y)$ denotes the modified Bessel function of the first kind (see Abramowitz and Stegun (1972)).

Result 3.4. *The distribution of the variance process is given by*

$$F_{var}(x; d, \lambda) = F_{nccs}\left(\frac{x}{C}; d, \lambda\right) = \sum_{j=1}^{\infty} e^{-\frac{\lambda}{2}} \frac{(\frac{\lambda}{2})^j}{j!} \frac{\Gamma(j + \frac{d}{2}, \frac{x}{2C})}{\Gamma(j + \frac{d}{2})},$$

where we have applied the Change of Variables Theorem. The corresponding density function, $f_{var}(x; d, \lambda)$, is given by

$$f_{var}(x) = \frac{dF_{var}(x; d, \lambda)}{dx} = \frac{1}{C} f_{nccs}\left(\frac{x}{C}; d, \lambda\right).$$

This density function, f_{var}, can be used to compute the moments of both the variance and the volatility model. The mean and variance of $\nu(t)$, conditional on $\mathcal{F}_s$, are given by

$$\mathbb{E}\left[\nu(t) \mid \mathcal{F}_s\right] = \int_0^\infty x f_{var}(x; d, \lambda)dx = C(d + \lambda) = \xi + (\nu(s) - \xi)e^{-\kappa(t-s)}, \tag{14}$$

$$\begin{aligned} Var\left(\nu(t) \mid \mathcal{F}_s\right) &= \int_0^\infty x^2 f_{var}(x; d, \lambda)dx - \left(\int_0^\infty x f_{var}(x; d, \lambda)dx\right)^2 = C^2(2d + 4\lambda) \\ &= \frac{\nu(s)\sigma_\nu^2 e^{-\kappa(t-s)}}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) + \frac{\xi\sigma_\nu^2}{2\kappa}\left(1 - e^{-\kappa(t-s)}\right)^2. \end{aligned} \tag{15}$$

The constant C is positive and a non-central chi-squared random variable has positive support for strictly positive parameters d and λ, which implies that the variance process is always well

defined because in realistic applications the parameters of the CIR model are strictly positive resulting in strictly positive values of d and λ. To ensure that the variance process remains strictly positive at time t with $0 < t \leq T$ and conditional on $0 < \nu(0)$, we need

$$1 \leq \frac{2\kappa\xi}{\sigma_\nu^2}, \tag{16}$$

(see Cairns (2004)), which is called the Feller condition. When (16) is not fulfilled, then the origin is accessible and strongly reflecting, i.e. most of the probability mass is around zero. In practical applications (16) is often not fulfilled which implies that a simulation scheme, such as the Euler scheme, can break down due to the square root process.

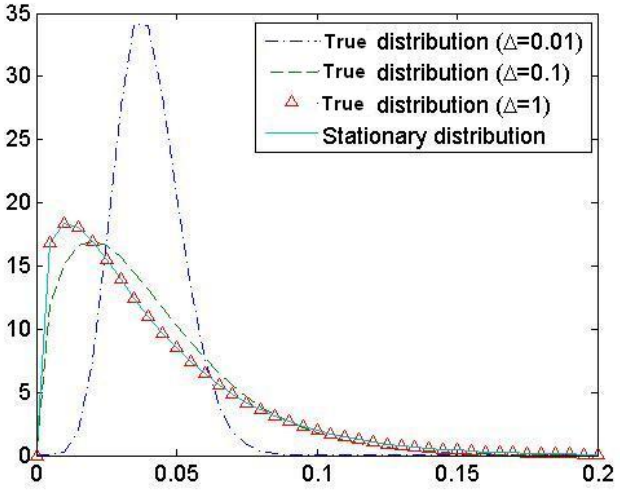
Broady and Kaya (2006) gave two representations of a non-central chi-squared random variable, $\chi_d^2(\lambda)$, with d degrees of freedom and non-centrality parameter λ:

$$\chi_d^2(\lambda) \sim \begin{cases} \left(Z + \sqrt{\lambda}\right)^2 + \chi_{d-1}^2 & \text{for } d > 1, \\ \chi_{d+2N}^2 & \text{for } d > 0, \end{cases} \tag{17}$$

with $Z \sim N(0,1)$, χ_v^2 an ordinary chi-squared distribution with v degrees of freedom and N is Poisson distributed with mean $\frac{1}{2}\lambda$. Because in most applications $d < 1$, the second representation is often used in practice.

The stationary distribution of (11) is according to Drăgulescu and Yakovenko (2002) equal to a gamma distribution (provided that $\xi > 0$), with scale and shape parameters equal to $\frac{2\xi}{d}$ and $\frac{d}{2}$, respectively. Drăgulescu and Yakovenko (2002) obtained this result by working out the Fokker-Planck equation and setting $\frac{\partial}{\partial t} f(t,x)$ equal to zero. For illustration we show Figure 2, where we have plotted the stationary distribution versus the true distribution with different values of the difference $t - s$ with $0 \leq s < t \leq T$ and where we used $\kappa = 6.2$, $\xi = 0.04$, $\sigma_\nu = 0.6$, $\nu(0) = \xi$:

Figure 2: True distribution versus the stationary distribution of the variance process



Analyzing Figure 2, we see that when $t - s$ tends to infinity, the distribution of the variance converges to the stationary distribution.

3.7.1 Analysis of the volatility model

For reasons of convergence, it is often preferable to transform a SDE so that it contains a constant volatility term. So, instead of investigating the variance model, one could investigate the volatility model[7]. To derive the SDE for the volatility we apply Itô's lemma to $\eta(t) = \sqrt{\nu(t)}$. The resulting SDE is given by

$$d\eta(t) = d\sqrt{\nu(t)} = \frac{1}{2}\kappa\left(\left(\xi - \frac{\sigma_\nu}{2\kappa}\right)\frac{1}{\eta(t)} - \eta(t)\right)dt + \frac{1}{2}\sigma_\nu dW^\eta(t).$$

However, a requirement for applying Itô's lemma is that the function $f(x) = \sqrt{x}$ is twice differentiable on the interval $[0, \infty)$, which is obviously not the case for $x = 0$. When (16) is not fulfilled, i.e. when the origin is attainable, and since the square root function is not differentiable at zero, the process obtained by incorrectly using Itô's lemma is structurally different.

Instead of transforming the SDE using Itô's lemma, we analyze the distribution function of the volatility model. This could be convenient for constructing an efficient simulation scheme. We have already remarked that the variance process is a non-central chi-squared distributed process. Next, we derive in Lemma 3.3 the distribution of the volatility process, which has not been found in the existing literature.

Lemma 3.3. *The distribution and density function of the volatility process, $\eta(t)$, are given by*

$$\begin{aligned} F_{vol}(x) &= F_{nccs}\left(\frac{x^2}{C}\right) = \sum_{j=1}^{\infty} e^{-\frac{\lambda}{2}}\frac{(\frac{\lambda}{2})^j}{j!}\frac{\Gamma\left(j+\frac{d}{2}, \frac{x^2}{2C}\right)}{\Gamma(j+\frac{d}{2})} \qquad \text{and} \\ f_{vol}(x) &= \frac{2x}{C} f_{nccs}\left(\frac{x^2}{C}\right). \end{aligned}$$

The mean and variance are given by the following series representations,

$$\mathbb{E}\left[\eta(t) \mid \eta(s)\right] = \sqrt{2C}e^{-\frac{\lambda}{2}}\sum_{j=0}^{\infty}\frac{\Gamma\left(\frac{1+d}{2}+j\right)}{\Gamma\left(\frac{d}{2}+j\right)}\frac{\lambda^j}{j!2^j}, \tag{18}$$

$$Var\left(\eta(t) \mid \eta(s)\right) = C(d+\lambda) - C2e^{-\lambda}\sum_{j=0}^{\infty}\frac{\Gamma\left(\frac{1+d}{2}+j\right)}{\Gamma\left(\frac{d}{2}+j\right)}\frac{\lambda^j}{j!2^j}. \tag{19}$$

Proof. We begin by working out the distribution of the volatility process,

$$\begin{aligned} F_{vol}\left(x\right) &= \mathbb{P}\left(0 \le \sqrt{\nu(t)} \le x \mid \nu(s)\right), \\ &= \mathbb{P}\left(0 \le \nu(t) \le x^2 \mid \nu(s)\right), \\ &= F_{nccs}\left(\frac{x^2}{C}\right). \end{aligned}$$

The density function is computed by

$$f_{vol}(x) = \frac{dF_{vol}\left(x\right)}{dx} = \frac{2x}{C} f_{nccs}\left(\frac{x^2}{C}\right).$$

[7]For constructing an efficient simulation method this could be convenient (see Zhu (2008)).

The expectation and variance of the distribution of the volatility are computed by[8]

$$\begin{aligned}
\mathbb{E}\left[\eta(t) \mid \eta(s)\right] &= \int_0^\infty \sqrt{x} f_{var}(x;d,\lambda)dx = \sqrt{C}\sqrt{2}e^{-\frac{\lambda}{2}}\Gamma\left(\frac{1+d}{2}\right) H\left(\frac{1+d}{2}, \frac{d}{2}, \frac{\lambda}{2}\right), \\
Var\left(\eta(t) \mid \eta(s)\right) &= \int_0^\infty x f_{var}(x;d,\lambda)dx - \left(\int_0^\infty \sqrt{x} f_{var}(x;d,\lambda)dx\right)^2 \\
&= C(d+\lambda) - C2e^{-\lambda}\Gamma\left(\frac{1+d}{2}\right)^2 H\left(\frac{1+d}{2}, \frac{d}{2}, \frac{\lambda}{2}\right)^2,
\end{aligned}$$

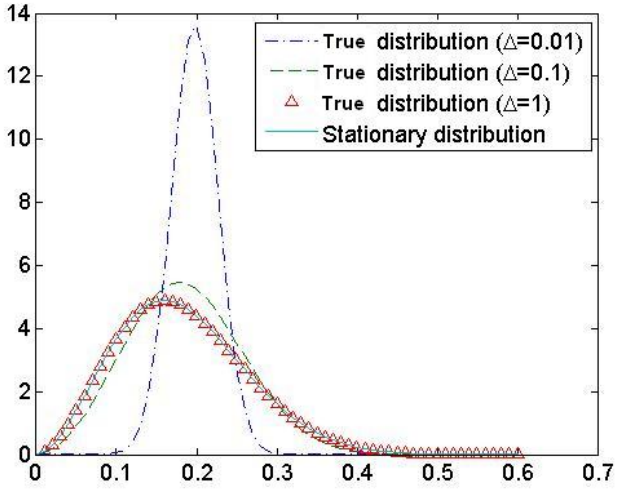
where $H(x,y,z) = \frac{1}{\Gamma(y)}\sum_{j=0}^{\infty} \frac{(x)_j}{(y)_j}\frac{z^j}{j!}$ denotes the regularized confluent hypergeometric function of the first kind, with $(x)_j = \frac{\Gamma(x+j)}{\Gamma(x)}$ and $(y)_j = \frac{\Gamma(y+j)}{\Gamma(y)}$ for $j \in \mathbb{N}$ (see Abramowitz and Stegun (1972)). We observe that the $\Gamma(x)$ and $\Gamma(y)$ terms cancel, which is beneficial for computational reasons. Simplifying the results, we obtain (18) and (19). □

The stationary distribution of the volatility model, $f_{stat}(x)$, can also be obtained using the Fokker-Planck equations. It turns out, according to Drăgulescu and Yakovenko (2002), that the stationary distribution of the volatility model has the following probability density function

$$f_{stat}(x) = \frac{2\frac{d}{2}^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}\frac{x^{d-1}}{\xi^{\frac{d}{2}}}e^{-\frac{d}{2\xi}x^2}.$$

In Figure 3 we have plotted the stationary distribution versus the true distribution with different values of the difference $t-s$, with $\kappa = 6.2$, $\xi = 0.04$, $\sigma_\nu = 0.6$, $\nu(0) = \xi$:

Figure 3: True distribution versus the stationary distribution of the volatility process



We again see that when $t-s$ tends to get larger, the distribution of the volatility converges to the stationary distribution.

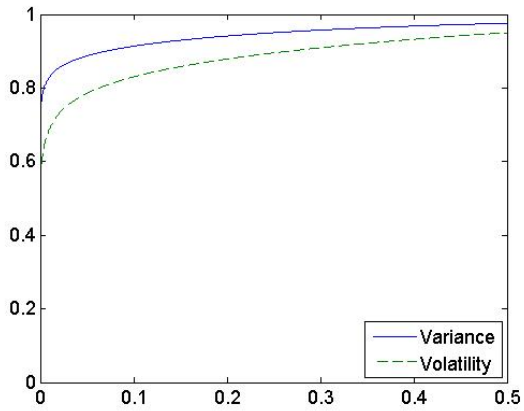
3.7.2 Comparison and convergence

Using cases I-IV described in Table 1, we compare the distribution function of the volatility and the variance process. We obtain the results in Figure 4:

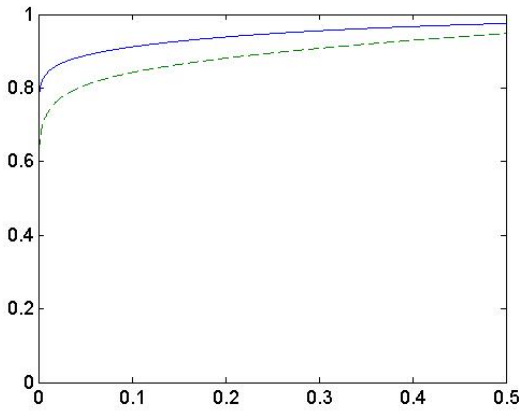
[8]These expressions are obtained using Mathematica.

Figure 4: Distribution of the variance and volatility process

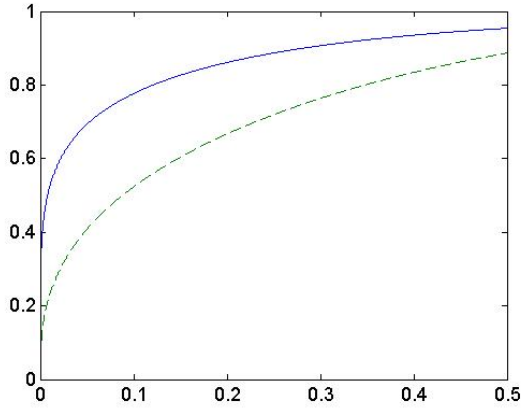
(a) Case I: $\nu(0) = \xi$, $T = 1$



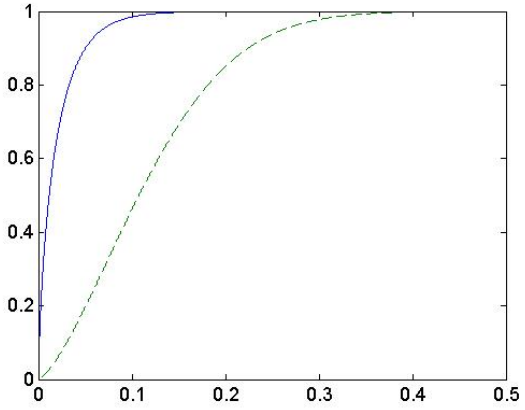
(b) Case II: $\nu(0) = \xi$, $T = 1$



(c) Case III: $\nu(0) = \xi$, $T = 1$



(d) Case IV: $\nu(0) = \xi$, $T = 1$

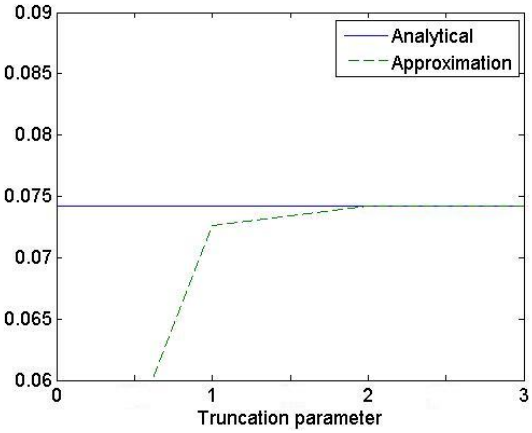


Analyzing these figures, we see that the distribution of the volatility has less probability mass around zero than the distribution of the variance process, which can be preferable for constructing an appropriate simulation scheme.

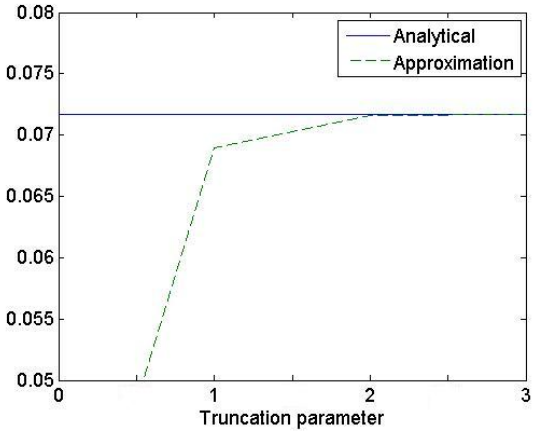
In order to compute the expectation and the variance of the volatility model we have to introduce a truncation parameter of the series in (18) and (19), to obtain values of the summation. In Figure 5, we have listed the convergence results:

Figure 5: Expectation of the volatility process using different truncation parameters

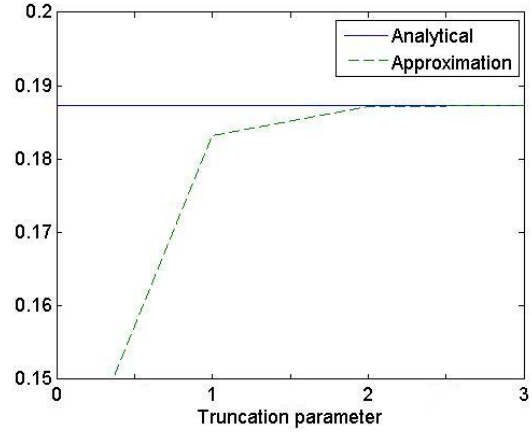
(a) Case I: $\nu(0) = \xi$, $T = 1$



(b) Case II: $\nu(0) = \xi$, $T = 1$



(c) Case III: $\nu(0) = \xi$, $T = 1$



(d) Case IV: $\nu(0) = \xi$, $T = 1$

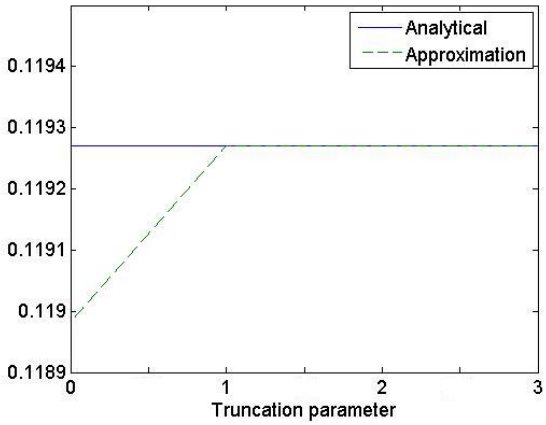


Figure 5 shows that a relatively low truncation parameter can be chosen to obtain sufficiently accurate results. Similar results are obtained for the variance of the volatility model.

3.8 Distributional properties of the asset price

By applying Itô's formula to $\log(S(t))$ and by writing the resulting SDE in integral form, it can be shown that the exact solution of the asset price dynamics is given by:

$$S(t) = S(s)e^{\int_s^t \left(r(u) - \frac{1}{2}\nu(u) - \delta\right)du + \int_s^t \sqrt{\nu(u)}\, dW^S(u)}, \tag{20}$$

where $0 \leq s < t \leq T$.

Brigo and Mercurio (2006, Appendix B) worked out the distributional properties of $\log\left(\frac{S(t)}{S(s)}\right)$, conditional on $\mathcal{F}_s$, when the asset price is modeled by a one-factor HWBS model. These results are easily extended to the case when the asset price is modeled by a two-factor HWBS model, by using the results in Appendix C. However, when the variance process is modeled by a CIR process, the derivations of the expectation and the variance are significantly more involved, due to the correlated variables. For that reason we do not proceed in this way.

4 Efficient path simulation of the extended Heston model

For the valuation of embedded options and for simulation purposes using the extended Heston model (as described in Chapter 3) one has to make use of simulation tools. The underlying model is given by a set of SDE's which cannot be used directly for simulation since it is a continuous model. Thus, numerical discretization is needed to approximate the underlying model.

In this chapter we discuss the simulation of the extended Heston model. We start with the discussion of existing simulation methods of the Heston model, which forms the basis of our new simulation method of the extended Heston model. In particular, we derive simulation methods for the interest rate process, variance process and the asset price process[9]. These simulation methods are based on the underlying distribution of the stochastic variables. So, the theory presented in Chapter 3 is used to set up an efficient simulation scheme. Furthermore, we discuss a martingale correction, an empirical martingale simulation and some convergence properties.

We suppose that we are given an arbitrary set of discrete times $\mathcal{I} = \{t_i\}_{i=1}^{M}$ and we consider the problem of generating random paths of the pairs $(S(t), \nu(t), r(t), u(t))$ for all $t \in \mathcal{I}$. This setup will be required, for instance, in the pricing of path-dependent options, where the pay-off function is dependent on different dates of, for example, the asset price. We are now interested in generating random variables $(S(t+\Delta t), \nu(t+\Delta t), r(t+\Delta t), u(t+\Delta t))$, conditional on $(S(t), \nu(t), r(t), u(t))$, for an arbitrary increment Δt. Repeated application of this one-period scheme will produce a full path $(S(t), \nu(t), r(t), u(t))_{t\in\mathcal{I}}$. Throughout the thesis, by $\left(\widetilde{S}, \widetilde{\nu}, \widetilde{r}, \widetilde{u}\right)$ we will denote a discrete-time approximation to (S, ν, r, u).

4.1 Existing simulation methods for the Heston model

In recent years several papers have appeared about the efficient (exact) simulation of the well-known Heston model. Although the application purpose in these papers is the same, namely the pricing of call options as efficient as possible in a plain MC framework, they all use different approaches for simulating the variance process, the integrated variance process and, thus, for the asset price process.

In Chapter 3 we have discussed the distributional properties of the variance process, which is represented by a non-central chi-squared distribution times a constant C. The generation of non-central chi-squared distributed random numbers is time-consuming because an acceptance-rejection method or a direct inversion has to be performed. Therefore, various alternatives are proposed so that the computation time is improved. The difficulty of constructing an appropriate scheme is when the Feller condition, (16), is not fulfilled, because then the origin is accessible and strongly reflecting so that the probability is high that the simulation scheme breaks down when negative variances are not accounted for. Below we briefly describe existing discretization methods of simulating the asset price when it is modeled by the Heston model.

4.1.1 Exact simulation

Broady and Kaya (2006) wrote a popular paper about the exact simulation of the Heston model. To simulate the variance process, $\nu(t+\Delta t)$ given $\nu(t)$, they use the representations given by (17). This involves generating from an ordinary chi-squared random variable, for which an

[9] In Appendix E and F we discuss alternative simulation method of the variance and the integrated variance, which, however, do not meet the performance of existing simulation schemes.

acceptance-rejection algorithm has to be used. Since ν is drawn from its exact probability distribution, the resulting sampling scheme for the ν process is completely free of bias.

To obtain a bias-free scheme for sampling the asset price process, Broady and Kaya (2006) use $\nu(t+\Delta t) = \nu(t) + \kappa\xi\Delta t - \kappa\int_t^{t+\Delta t}\nu(u)du + \sigma_\nu\int_t^{t+\Delta t}\sqrt{\nu(u)}dW^\nu(u)$ to obtain

$$\int_t^{t+\Delta t}\sqrt{\nu(u)}dW^\nu(u) = \left(\frac{1}{\sigma_\nu}\right)\left(\nu(t+\Delta t) - \nu(t) - \kappa\xi\Delta t + \kappa\int_t^{t+\Delta t}\nu(u)du\right). \tag{21}$$

The observation that the process for $\nu(t+\Delta t)$ is independent of the Brownian motion $W^S(t+\Delta t)$ gives the insight that the distribution of $\int_t^{t+\Delta t}\sqrt{\nu(u)}dW^S(u)$ given the path generated by $\nu(t+\Delta t)$ is normal with zero mean and variance equal to $\int_t^{t+\Delta t}\nu(u)du$.

A Cholesky decomposition shows that

$$\begin{aligned}\log\left(S(t+\Delta t)\right) &= \log\left(S(t) + r\Delta t - \frac{1}{2}\int_t^{t+\Delta t}\nu(u)du\Delta t\right.\\ &\quad \left.+\rho_{S,\nu}\int_t^{t+\Delta t}\sqrt{\nu(u)}dW^\nu(u) + \sqrt{1-\rho_{S,\nu}^2}\int_t^{t+\Delta t}\sqrt{\nu(u)}dW(u),\right.\end{aligned}$$

where W is a Brownian motion independent of W^ν. When we write this in integral form, we get

$$\begin{aligned}\log\left(S(t+\Delta t)\right) &= \log\left(S(t) + r\Delta t + \frac{\rho_{S,\nu}}{\sigma_\nu}\left(\nu(t+\Delta t) - \nu(t) - \kappa\xi\Delta t\right)\right.\\ &\quad \left.+\left(\frac{\kappa\rho_{S,\nu}}{\sigma_\nu} - \frac{1}{2}\right)\int_t^{t+\Delta t}\nu(u)du + \sqrt{1-\rho_{S,\nu}^2}\int_t^{t+\Delta t}\sqrt{\nu(u)}dW(u).\right.\end{aligned} \tag{22}$$

This shows the presence of an integrated variance process. The main result in their paper is the formulation of the characteristic function of the integrated variance process, conditional on its begin and endpoint, which we denote by $Iv(s,t)$ and is given by

$$\begin{aligned}\Psi_{Iv}(c,\nu_s,\nu_t) &\equiv \mathbb{E}\left[e^{\int_s^t\nu(u)du} \mid \nu(s)=\nu_s, \nu(t)=\nu_t\right] \\ &= \frac{\gamma_1(c)e^{\frac{1}{2}(\gamma_1(c)-\kappa)(t-s)}\left(1-e^{-\kappa(t-s)}\right)}{\kappa\left(1-e^{-\gamma_1(c)(t-s)}\right)}\\ &\quad e^{\gamma_2(\gamma_3-\gamma_4)}\cdot\frac{B_{\frac{1}{2}d-1}\left(\sqrt{\nu_s\nu_t}4\gamma_1(c)e^{-\frac{\gamma_1(c)}{2}(t-s)/\sigma_\nu\left(1-e^{\gamma(c)(t-s)}\right)}\right)}{B_{\frac{1}{2}d-1}\left(\sqrt{\nu_s\nu_t}4\kappa e^{-\frac{\gamma_1(c)}{2}(t-s)/\sigma_\nu\left(1-e^{\kappa(t-s)}\right)}\right)},\end{aligned} \tag{23}$$

where

$$\begin{aligned}\gamma_1(c) = \sqrt{\kappa^2 - 2\sigma_\nu^2 ic}, \quad d &= \frac{4\kappa\xi}{\sigma_\nu^2}, \quad \gamma_2 = \frac{\nu_s+\nu_t}{\sigma_\nu^2}, \quad \gamma_3 = \frac{\kappa\left(1+e^{-\kappa(t-s)}\right)}{1-e^{-\kappa(t-s)}},\\ \gamma_4 &= \frac{\gamma(c)\left(1+e^{-\gamma(c)(t-s)}\right)}{1-e^{-\gamma(c)(t-s)}}\end{aligned}$$

and $B_{(.)}(x)$ denotes the modified Bessel function of the first kind. The characteristic function can be used to obtain a value of the distribution function by evaluation the following integral representation

$$F_{Iv}(x) = \mathbb{P}\left(Iv(s,t)\le x\right) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin(ux)}{u}\Psi(u)du = \frac{2}{\pi}\int_0^{\infty}\frac{\sin(ux)}{u}\Re(\Psi(u))du, \tag{24}$$

where $\Re$ denotes the real part. Because this integral cannot be computed analytically, a value of the distribution is obtained numerically by a trapezoidal integration method, i.e.

$$F_{Iv}(x) \approx \frac{hx}{\pi} + \frac{2}{\pi} \sum_{j=1}^{N_{trap}} \frac{\sin(hjx)}{j} \Re(\Psi(hj)), \tag{25}$$

where $h = \frac{2\pi}{x+\mu_\epsilon}$, μ_ϵ denotes the expectation plus five or more standard deviations and N_{trap} should be chosen such that $\frac{|\Psi(hN_{trap})|}{N_{trap}} < \frac{\pi\epsilon}{2}$ is sharp (ϵ is the required accuracy). One could also specify h and N_{trap} by trial and error.

Obtaining a value of the inverse distribution function a time consuming iterative inversion method (a second-order Newton method) is used, to approximate x from $F_{Iv}(x) = U$, where U is a uniform random variable, with prespecified error bound. Although this is an exact simulation of a value of the inverse distribution (with errors controlled by the step size and the tolerance of the inversion method) the method is very time consuming due to the many modified Bessel function evaluations[10].

Smith (2008) proposes an adjustment to the famous Broady and Kaya scheme. Although Smith (2008) constructs an approximation of the characteristic function such that the characteristic function can be cached in two-dimensions, a (time-consuming) Fourier inversion still has to be performed. More specifically, $\sqrt{\nu(t)\nu(t+\Delta t)}$ as well as $\frac{1}{2}(\nu(t) + \nu(t+\Delta t))$ in (23) are replaced by $\omega\frac{1}{2}(\nu(t) + \nu(t+\Delta t)) + (1-\omega)\sqrt{\nu(t)\nu(t+\Delta t)}$ where $\omega \in [0,1]$. Smith (2008) concludes that the method works well up to a maturity of 5 years.

In order to simulate $\log(S(t+\Delta t)$ using the Broady and Kaya scheme, one has to perform the following steps:

1. Generate $\nu(t+\Delta t)$, conditional on $\nu(t)$, using the representations of a non-central chi-squared random variable given in (17), times the constant C;
2. Generate $Iv(t, t+\Delta t)$, conditional on $(\nu(t), \nu(t+\Delta t))$, by numerically solving x from $F_{Iv}(x) = U$, where U is uniformly distributed and $F_{Iv}(x)$ is approximated by (25);
3. Conditional on the pair $\left(\nu(t+\Delta t, \int_t^{t+\Delta t} \nu(u)du\right)$, use (22) to generate a sample of the asset price, $S(t+\Delta t)$.

A similar approach is the approach by Glasserman and Kim (2008). They propose the gamma expansion method for simulating the integrated variance process. Using the gamma expansion method the simulation of the integrated variance process can be done more efficient compared to the exact simulation scheme of Broady and Kaya (2006). Their main result is an explicit representation of the integrated variance process in terms of infinite sums and mixtures of gamma random variables. Glasserman and Kim show that their simulation scheme outperforms the Broady and Kaya scheme and competes with the QE scheme of Andersen (2008), which is discussed below. Although the implementation of the simulation scheme is not straightforward, this scheme might be preferred when dealing with long (fixed) maturity options.

[10] In the evaluation of a modified Bessel function of the first kind, special care has to go to the argument of z in $B(v,z)$, because otherwise $B(v,z)$ is discontinuous at each point along the negative x-axis.

4.1.2 Euler discretization

A general approach to simulate the Heston model is using the Euler discretization scheme. Applying the Euler scheme to the variance process leads to a positive probability of the variance being negative, i.e.

$$\mathbb{P}\left(\widetilde{\nu}(t+\Delta t)<0 \mid \widetilde{\nu}(t)>0\right)=1-\Phi\left(\frac{(1-\kappa\Delta t)\widetilde{\nu}(t)+\kappa\xi\Delta t}{\sigma_\nu\sqrt{\widetilde{\nu}(t)\Delta t}}\right),$$

and thus the scheme can break down since the square root is not defined for negative values. Especially when (16) is not fulfilled most probability mass of the variance is concentrated around zero, which results in a high probability of the variance being negative using the Euler scheme. Lord and Koekkoek *et al.* (2008) discuss several adjustments of the Euler discretization schemes. They unify all Euler fixes into a single general framework and recommend the so-called full truncation (FT) scheme, tailored to minimize the positive bias found when pricing European options.

The FT scheme for the square-root process reads,

$$\widetilde{\nu}(t+\Delta t)=\max\left(\widetilde{\nu}(t)+\kappa\Delta t(\xi-\max\left(\widetilde{\nu}(t),0\right))+\sigma_\nu\max\left(\widetilde{\nu}(t),0\right)\Delta W_\nu(t),0\right).$$

Although the scheme cannot break down, the scheme is parameter dependent, i.e. when (16) is not fulfilled then the FT scheme produces unsatisfactory results due to the fact that the scheme is not able to produce the true dynamics.

The asset price is simulated by using the fact that the log-asset price is log-normally distributed, i.e.

$$\widetilde{S}(t+\Delta t)=\widetilde{S}(t)e^{(r-\frac{1}{2}\nu(t))\Delta t+\sqrt{\nu(t)}\Delta W^S(t)}.$$

Implementing this scheme can be done using a Cholesky decomposition, i.e. $\Delta W^S(t)=\rho_{S,\nu}\Delta W^\nu(t)+\sqrt{1-\rho_{S,\nu}^2}\Delta W(t)$, with $W(t)$ independent of $W^\nu(t)$.

Lord and Koekkoek *et al.* (2008) show by means of numerical experiments that their FT scheme outperforms existing Taylor discretization schemes such as the quasi second-order schemes of Kahl and Jäckel (2006) and Ninomiya and Victoir (2004). The method is also more efficient than the exact scheme of Broadie and Kaya explained in the previous chapter.

4.1.3 QE and TG scheme

Andersen (2007) propose the Truncated Gaussian (TR) and the Quadratic Exponential (QE) scheme for simulating the variance process. Both are discretization schemes and constructed using moment matching techniques. Much attention is paid to the generation of the variance process because, as already remarked, this is the most difficult step in the simulation procedure together with the simulation of the integrated variance process. The TG method has the (simple) form $\widetilde{\nu}(t+\Delta t)=(\mu_{TG}+\sigma_{TG}Z_\nu)^+$, where $(\cdot)^+=\max(\cdot,0)$, Z_ν is normally distributed and μ_{TG} and σ_{TG} are solved using a moment matching technique. More specifically, they are given by

$$\mu_{TG} = f_{\mu_{TG}}(\psi)\cdot m, \qquad f_{\mu_{TG}}(\psi)=\frac{r(\psi)}{f_{norm}(r(\psi))+r(\psi)F_{norm}(r(\psi))}, \tag{26}$$

$$\sigma_{TG} = f_{\sigma_{TG}}(\psi)\cdot s, \qquad f_{\sigma_{TG}}(\psi)=\frac{1}{\sqrt{\psi}\left(f_{norm}(r(\psi))+r(\psi)F_{norm}(r(\psi))\right)}, \tag{27}$$

where $\psi = \frac{s^2}{m^2} = \frac{Var(\nu(t+\Delta t)|\,\nu(t))}{\mathbb{E}[\nu(t+\Delta t)|\,\nu(t)]^2}$, where $Var\left(\nu(t+\Delta t)\,|\,\nu(t)\right)$ and $\mathbb{E}\left[\nu(t+\Delta t)\,|\,\nu(t)\right]$ can be found in Chapter 3.7. The function $r(\psi)$ has to be recovered from

$$rf_{norm}(r) + F_{norm}(r)(1+r^2) = (1+\psi)(f_{norm}(r) + rF_{norm}(r))^2,$$

using numerical tools. The functions $f_{\mu_{TG}}(\psi)$ and $f_{\sigma_{TG}}(\psi)$ can be prestored on a computer once, for different values of ψ.

Generating $\widetilde{\nu}(t+\Delta t)$ given $\widetilde{\nu}(t)$ using the TG algorithm consists of the following steps:

1. Given $\widetilde{\nu}(t)$, compute m and s^2 using (14) and (15);
2. Compute $\psi = \frac{s^2}{m^2}$ and look up $f_{\mu_{TG}}(\psi)$ and $f_{\sigma_{TG}}(\psi)$ from cache;
3. Compute μ_{TG} and σ_{TG} according to (26) and (27);
4. Generate a random sample, Z_ν, of the standard normal distribution;
5. Compute $\widetilde{\nu}(t+\Delta t) = (\mu_{TG} + \sigma_{TG} Z_\nu)^+$.

The QE scheme distinguishes between low and high variance values using a switching rule. This is mainly due to the dramatic behavior of the distribution function of the variance process. For sufficiently large values of the variance they propose

$$\widetilde{\nu}(t+\Delta t) = a(b+Z_\nu)^2, \tag{28}$$

to simulate the variance process and for sufficiently low values of the variance they propose to simulate from the following distribution function

$$P\left(\widetilde{\nu}(t+\Delta t) \leq x\right) = p + (1-p)(1-e^{-\beta x}), \quad x \geq 0 \tag{29}$$

where a, b, p and β are computed using moment matching techniques. Provided that $\psi = \frac{s^2}{m^2} \leq 2$, we have

$$b^2 = \frac{2}{\psi} - 1 + \sqrt{\frac{2}{\psi}}\sqrt{\frac{2}{\psi} - 1} \geq 0, \qquad a = \frac{m}{1+b^2}. \tag{30}$$

Provided that $\psi \geq 1$, we have

$$p = \frac{\psi - 1}{\psi + 1} \in [0,1), \qquad \beta = \frac{2}{m(\psi+1)} \geq 0. \tag{31}$$

Introducing a switching rule through a critical level $\psi_c \in [1,2]$ and use (28) when $\psi \leq \psi_c$ and (29) otherwise. As a critical level, Andersen (2007) recommends using $\psi_c = 1.5$.

The algorithm for the QE simulation step from $\widetilde{\nu}(t)$ to $\widetilde{\nu}(t+\Delta t)$ is then given as follows,

1. Given $\widetilde{\nu}(t)$, compute m and s^2 from (14) and (15);
2. Compute $\psi = \frac{s^2}{m^2}$;
3. Draw a uniform random number[11] U_ν;
4. If $\psi \leq \psi_c$:

[11] Note that both pseudo-random numbers as well as quasi-random numbers can be used.

- Compute a and b from equations in (30);
- Compute $Z_\nu = \Phi^{-1}(U_\nu)$ using an appropriate transformation method;
- Compute $\widetilde{\nu}(t+\Delta t) = a(b+Z_\nu)^2$.

5. Otherwise, if $\psi > \psi_c$:

   - Compute β and p from equations in (31);
   - Compute $\widetilde{\nu}(t+\Delta t) = \Psi^{-1}(U_\nu; p, \beta)$, where $\Psi(x : p, \beta)$ denotes the inverse of (29).

Next, he simulates the asset price using the log-asset price, with a drift interpolation method to discretize the integrated variance process, conditional on its begin and endpoints, i.e.

$$\int_t^{t+\Delta t} \nu(u)du = \gamma_1 \Delta t \nu(t) + \gamma_2 \Delta t \nu(t+\Delta t),$$

where γ_1 and γ_2 are to be chosen, with the requirement that their sum is equal to one. They remark that for relative small time steps the variability of the integrated variance process is relatively low so that a drift interpolation is sufficiently accurate. The sampling scheme of the asset price is then given by

$$\widetilde{S}(t+\Delta t) = \widetilde{S}(t) e^{K_0 + K_1 \nu(t) + K_2 \nu(t+\Delta t) + \sqrt{K_3 \nu(t) + K_4 \nu(t+\Delta t)} Z_S},$$

where Z_S is standard normally distributed and

$$\begin{aligned} K_0 &= \left(r - \frac{\rho_{S,\nu}\kappa\xi}{\sigma_\nu}\right)\Delta t, \qquad K_1 = \gamma_1 \Delta t \left(\frac{\kappa \rho_{S,\nu}}{\sigma_\nu}\right) - \frac{\rho_{S,\nu}}{\sigma_\nu}, \\ K_2 &= \gamma_2 \Delta t \left(\frac{\kappa \rho_{S,\nu}}{\sigma_\nu}\right) - \frac{\rho_{S,\nu}}{\sigma_\nu}, \qquad K_3 = \gamma_1 \Delta t (1-\rho_{S,\nu}^2), \qquad K_4 = \gamma_2 \Delta t (1-\rho_{S,\nu}^2). \end{aligned}$$

A martingale correction is constructed and can be used to adjust the asset price so that the discretized asset price satisfies the martingale condition. This is done by replacing K_0 by K_0^*, which can be found in Andersen (2007). Using numerical experiments they show that the TG and QE schemes are much more efficient than the FT scheme and other Taylor discretization schemes. Their experiments also indicate that the QE scheme is to be preferred as a simulation scheme and should be the default choice.

4.1.4 NCI scheme

Haastrecht and Pelsser (2008) propose an approximation of the inverse non-central chi-squared distribution to simulate the variance process. They use the second representation of (17) for generating the variance process. Thus, this involves generating from a Poisson distribution as well as from a chi-squared distribution. To construct an efficient simulation algorithm they make use of interpolation methods to cache the chi-squared distributions. They introduce two grids,

$$\mathcal{N} = \{0, \ldots, N_j, \ldots, N_{max}\} \quad \text{and} \quad \mathcal{U} = \{0, \ldots, 1-\delta\},$$

where a parameter δ is introduced to avoid numerical difficulties[12]. The grid $\mathcal{N}$ represents the set of Poisson-values, where N_{max} is chosen such that $\mathbb{P}(N > N_{max})$ is sufficiently small, of which the inverse of the corresponding conditional chi-squared distributions are cached. The

[12]This is due to the fact that the quantile function of a chi-squared random variable evaluated in 1 is equal to infinity.

inverse chi-squared distributions are precomputed using the grid $\mathcal{U}$ so that subsequently an interpolation method can be performed to generate a sample of the chi-squared distribution. Haastecht and Pelsser (2008) use a monotone cubic Hermite spline interpolation in combination with an equidistantly chosen grid, $\mathcal{U}$. The NCI algorithm is as follows:

1. Conditional on the values of $\mathcal{N}$, precompute the inverse chi-squared distributions using an equidistant $\mathcal{U}$;
2. Conditional on $\widetilde{\nu}(t)$, compute the parameter λ, which can be found in Chapter 3.7;
3. Generate a Poisson distributed random number, N^P, of a Poisson random variable with mean $\frac{1}{2}\lambda$[13];
4. If $N^P \leq N_{max}$, then:
   - Conditional on the value N^P, use grid $\mathcal{N}$ to look up the corresponding inverse chi-squared distribution values, $F_{chi}^{-1}(U; d+2N^P)$, where $U \in \mathcal{U}$ and with $d+2N^P$ degrees of freedom;
   - Generate a uniformly distributed sample, U_ν, and compute $\widetilde{F}_{chi}^{-1}(U_\nu; d+2N^P)$ by using a monotone cubic Hermite spline interpolation in combination with an equidistantly chosen grid, $\mathcal{U}$, and the corresponding precomputed inverse chi-squared distribution function values $F_{chi}^{-1}(U; d+2N^P)$, with $U \in \mathcal{U}$, where $\widetilde{F}^{-1}$ denotes the approximated inverse chi-squared distribution function;
   - Then, compute $\nu(t+\Delta t) = C\widetilde{F}_{chi}^{-1}(U_\nu; d+2N^P)$, where the constant C can be found in Chapter 3.7.
5. Otherwise:
   - Generate a sample of $\nu(t+\Delta t)$, by generating a sample of a chi-squared random variable with $d+2N^P$ degrees of freedom, conditional on N^P, using known generation schemes (see Glasserman (2003)).

The NCI scheme can also be used in combination with the QE scheme, because the NCI is especially accurate for low values of the non-centrality parameter and the QE scheme especially for high values of the non-centrality parameter. However, numerical experiments show that the QE scheme with a martingale correction still outperforms the NCI scheme.

For the simulation of the asset price, the log-asset price in combination with a drift interpolation method for the integrated variance process is used, which is explained in the previous chapter. Since the variability of the integrated variance process is relatively small, the drift interpolation method is sufficiently accurate for relatively small time steps.

4.1.5 Transformation to volatility

Zhu (2008) transforms the variance model to the volatility model using Itô's lemma and simulates thereof. Using Itô's lemma, the SDE for the volatility is given by

$$d\sqrt{\nu(t)} = \frac{1}{2}\kappa\left(\left(\xi - \frac{\sigma_\nu^2}{4\kappa}\right)\frac{1}{\sqrt{\nu(t)}} - \sqrt{\nu(t)}\right)\Delta t + \frac{1}{2}\sigma_\nu dW^\nu(t).$$

[13] See for more information Knuth (1981) and Ahrens and Dieter (1982) for generating Poisson random variables.

Zhu (2008) argues that this scheme cannot break down since there is no square root process involved. For simulation, two methods are suggested. On the one hand they propose a central discretization to simulate the (stochastic) long mean level properly. They recommend the central discretization scheme only in case the Feller condition, (16), is fulfilled. On the other hand they use a moment matching technique to simulate the (stochastic) long mean level properly. Even when the Feller condition is not fulfilled, this scheme produces good simulation results and is preferred to the central discretization scheme.

Their numerical results show that the results of their proposed scheme are similar to those of the QE scheme of Andersen (2007) and thus provides a good alternative to existing simulation schemes for the Heston model.

4.2 Contribution to the simulation of the interest rate

In this chapter we discuss the proposed simulation scheme of the interest rate. This scheme is based on the distributional properties of the interest rate, which are discussed in Chapter 3.

4.2.1 Simulation of the two-factor Hull-White model

We now use the distributional properties of the interest rate to simulate the interest rate process in an efficient way. We gave two representations of the interest rate process given in Result 3.2 and Result 3.3, which can be used for simulation. Using representation (3.2), we get the following algorithm for simulation:

- Generate a sample, Z, from the standard normal distribution;
- Compute the mean, μ_r, and variance, σ_r^2, conditional on $\mathcal{F}_t$, of the interest rate given by Result 3.2;
- Compute $\widetilde{r}(t+\Delta t) = \mu_r + \sigma_r Z$.

On the other hand, in the two-factor HW model we see that the first factor, r, is dependent on the second factor, u. According to Chapter 3, the second factor is normally distributed and, conditional on the pair $(\widetilde{u}(t), \widetilde{r}(t))$, the first factor is also normally distributed. This having remarked, we use (4) and (5) to simulate the interest rate, which results in the following algorithm:

- Generate $\widetilde{u}(t+\Delta t)$, conditional on $\widetilde{u}(t)$, using

$$\widetilde{u}(t+\Delta t) = \widetilde{u}(t)e^{-b\Delta t} + \sqrt{\frac{\sigma_u^2}{2b}\left(1-e^{-2b\Delta t}\right)}\, Z_u;$$

- Set $\widetilde{\theta}(t+\Delta t) = \theta(t+\Delta t) + \widetilde{u}(t)$;
- Generate $\widetilde{r}(t+\Delta t)$, conditional on $(\widetilde{u}(t), \widetilde{r}(t))$, by using

$$\widetilde{r}(t+\Delta t) = \widetilde{r}(t)e^{-a\Delta t} + \int_t^{t+\Delta t} \widetilde{\theta}(u)e^{-a(t-u)}du + \sqrt{\frac{\sigma_r^2}{2a}\left(1-e^{-2a\Delta t}\right)}\, Z_r,$$

  where $\int_t^{t+\Delta t} \widetilde{\theta}(u)e^{-a(t-u)}du$ can be worked out using Lemma 3.1 and Z_u and Z_r are correlated normal random variables.

The first simulation algorithm is to be preferred because it it is a direct generation of the true dynamics of the interest rate. However, the second algorithm can be combined with the simulation of the inflation dynamics in the full risk-neutral model, which is given in Appendix A. This can be done by performing some simple computations, which result in a reduction of computation time. Therefore, we prefer the second generation method to simulate the interest rate dynamics.

4.2.2 Simulation of the two-factor Gaussian model

In practice it is often more convenient to use the Gaussian interest rate model, because of its simple form. Because there is a one-to-one relation between the HW and the Gaussian interest rate model (see Brigo and Mercurio (2006, Chapter 4)), we can transform the two-factor HW model to the two-factor Gaussian model and simulate thereof[14]. We have already mentioned in Chapter 3.5 that the factors x and y of the two-factor Gaussian interest rate model are both normally distributed. This makes the simulation of the interest rate rather easy, because we only have to generate a sample of normally distributed random numbers with corresponding mean and variance. To simulate the interest rate using the two-factor Gaussian model, one has to perform the following steps:

- Transform the two-factor HW model to the two-factor Gaussian model;
- Conditional on $\widetilde{x}(t)$ and $\widetilde{y}(t)$, generate $\widetilde{x}(t+\Delta t)$ and $\widetilde{y}(t+\Delta t)$ by using

$$\begin{aligned}\widetilde{x}(t+\Delta t) &= \widetilde{x}(t)e^{-b_x\Delta t} + \sqrt{\frac{\sigma_x^2}{2b_x}\left(1-e^{-2b_x\Delta t}\right)}\,Z_x,\\ \widetilde{y}(t+\Delta t) &= \widetilde{y}(t)e^{-b_y\Delta t} + \sqrt{\frac{\sigma_y^2}{2b_y}\left(1-e^{-2b_y\Delta t}\right)}\,Z_y,\end{aligned}$$

  where Z_x and Z_y are correlated normal random variables;
- Conditional on $\mathcal{F}_t$, compute $\varphi(t+\Delta t)$ by using Lemma 3.1;
- Finally, set $\widetilde{r}(t+\Delta t) = \widetilde{x}(t+\Delta t) + \widetilde{y}(t+\Delta t) + \varphi(t+\Delta t)$.

Because the two-factor Gaussian model has a more simple form than the two-factor HW model, we prefer using the two-factor Gaussian model for simulation. Although the generation of the inflation dynamics is not as simple as by using the two-factor HW model, we proceed using the two-factor Gaussian model, because of its simple form, for deriving a discretization method for the asset price, which is discussed in Chapter 4.4.

4.2.3 Numerical experiment

Using the algorithms discussed above for generating the interest rate process, the interest rate is sampled directly from the known distribution derived in Chapter 3. As the generation of standard normally distributed random numbers can be done very fast, the generation of the interest rate can be done in a highly efficient way.

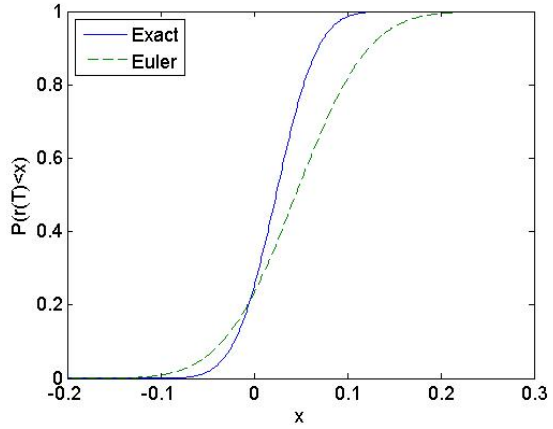
To show the difference of the exact scheme with the Euler scheme we compute the cumulative distribution function of the interest rate at a time T conditional on time $t = 0$, where we use for simplicity the one-factor HW model. Figure 6 shows the cumulative distribution function for

[14] Note that our starting point is the two-factor HW interest rate model.

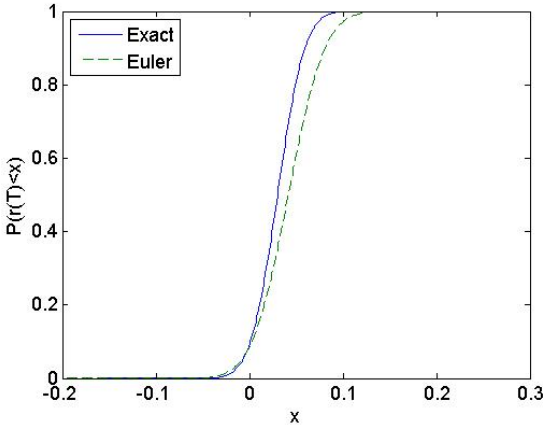
$r(T)$ given $r(0)$, $a = 5$, $r(0) = 0.04$, $\sigma_r = 0.7$ and we assume that the market's instantaneous forward rate at time 0 with maturity t, $f(0,t)$, is equal to 0.04. We vary time T to compare the accuracy of the Euler scheme. We use the Euler discretization scheme with one step so that at time T the interest rate is normally distributed with mean $r(0) + (\theta(0) - ar(0))T$ and variance $\sigma_r^2 T$.

Figure 6: Cumulative distribution of r for different values of T

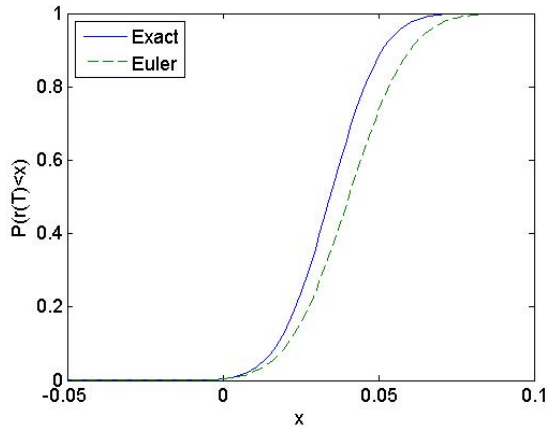
(a) $T = \frac{1}{8}$



(b) $T = \frac{1}{16}$



(c) $T = \frac{1}{32}$



(d) $T = \frac{1}{1024}$

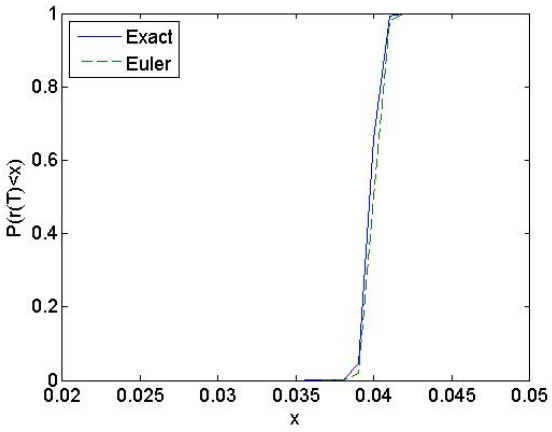


Figure 6 shows that the Euler scheme is only accurate for a relative small time step. We remark that the results in Figure 6 are sensitive to the parameter settings. For example, using a lower value of a the Euler scheme can be as accurate as the example discussed using a larger time step.

Next, we show the computation time to generate scenarios with the same parameters as discussed above but instead of using a small value of T, we use $T = 5$. We use the Euler scheme with 100 steps per year and the exact scheme with 10 steps per year. We get the results in Table 2:

Table 2: Computation times

N	Exact	Euler
10^3	0.0026	0.0407
10^4	0.0138	0.1492
10^5	0.2377	1.7460
10^6	2.0526	18.6022

Table 2 shows that the computation time is improved by approximately a factor 10. Similar results are obtained when using the two-factor HW model.

4.3 Contribution to the simulation of the variance process

In Haastrecht and Pelsser (2008) the application of the NCI scheme is explained and tested against other simulation methods. As already explained, they make use of interpolation techniques to set up an efficient simulation scheme based on the second representation of (17). This involves generation of a sample of a Poisson random variable and subsequently of a chi-squared random variable using an monotone cubic Hermite spline interpolation techniques in combination with an equidistantly chosen grid $\mathcal{U}$, which is a discretization of the uniform interval $[0, 1)$.

A way of improving the NCI scheme is by using the first representation given in (17) for simulating the variance process which holds for $d > 1$, where $d = \frac{4\kappa\xi}{\sigma_\nu^2}$. When $d > 1$, the sampling of the variance can be done significantly more efficient since the standard normal distribution and the chi-squared random variable with $d - 1$ degrees of freedom do not depend on $\widetilde{\nu}(t)$. Consequently, we can prestore the inverse normal distribution and the inverse chi-squared distribution by using the grid $\mathcal{U}$.

Whereas the generation of a standard normal random variable can be done very fast by using, for example, Moro's inversion or the Marsaglia polar method, the generation of a chi-squared random variable has to be done using an acceptance-rejection algorithm (see Glasserman (2003)). Therefore, we generate a sample of a chi-squared random variable by means of interpolation techniques as explained by Haastrecht and Pelsser (2008). I.e. we precompute the inverse chi-squared distribution based on a chosen grid $\mathcal{U}$ and interpolate between these points by generating (only) an uniform sample[15].

Since Haastrecht and Pelsser (2008) use an advanced interpolation technique based on an equidistant grid, a relatively fine grid has to be used in order to obtain accurate results. This is due to the often (typical) dramatic behavior of the quantile function near the tails of the distribution function. We therefore propose a non-equidistantly chosen grid $\mathcal{U}$ to discretize the uniform interval $[0, 1)$ in combination with a (simple) linear interpolation method, which we call an adapted linear interpolation method. Using this approach interpolation can be done more efficiently. As an illustration we perform a linear interpolation technique, a cubic interpolation method[16] and an adaptive linear interpolation method, where the non-equidistantly chosen grid is specified by trial and error, by precomputing 10 inverse values[17] of the inverse chi-squared distribution with parameter 0.1. The results are shown in Figure 7:

[15]Note that this approach can also be used to generate the standard random variable.

[16]In Matlab the cubic interpolation method produces the same results as the cubic Hermite interpolation method.

[17]In case of linear and cubic interpolation we use equally spaced values.

Figure 7: Approximation of the inverse chi-squared distribution using interpolation methods

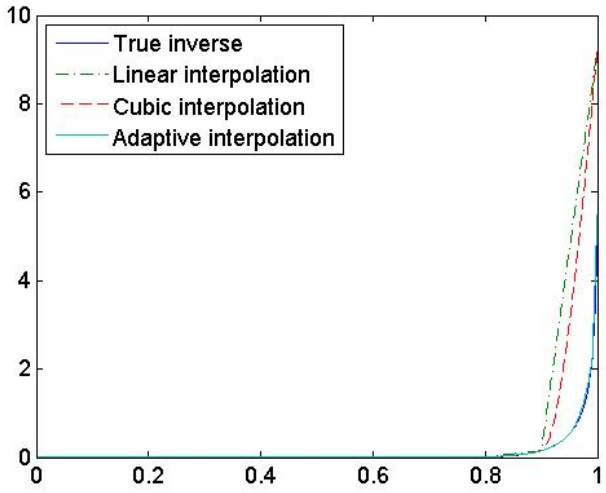


Figure 7 shows that an adaptive interpolation method is to be preferred, because an almost exact recovery of the true distribution is obtained. To achieve accurate results using an equidistant grid in combination with the linear interpolation or a cubic interpolation method, (significantly) more points have to be precomputed due to the drastically changing behavior of the quantile function near the tails.

Next, interpolating the quantile function of the standard normal distribution can be done easily since it is an odd function around $\frac{1}{2}$. Therefore, one precomputes inverse values only in the interval $[0, \frac{1}{2}]$ or $[\frac{1}{2}, 1]$. In the case of using the interval $[\frac{1}{2}, 1]$ for precomputing the inverse values, the value of $F^{-1}_{norm}(\hat{x})$, for example, is obtained from $-F^{-1}_{norm}(1-\hat{x})$, where $\hat{x} \in [0, \frac{1}{2}]$. As an illustration, we perform a linear interpolation technique, a cubic interpolation method and an adaptive linear interpolation method by precomputing 5 inverse values[18] of the inverse standard normal distribution with parameter, we get the results in Figure 8:

[18]In case of linear and cubic interpolation we use equally spaced values.

Figure 8: Approximation of the inverse standard normal distribution using interpolation methods

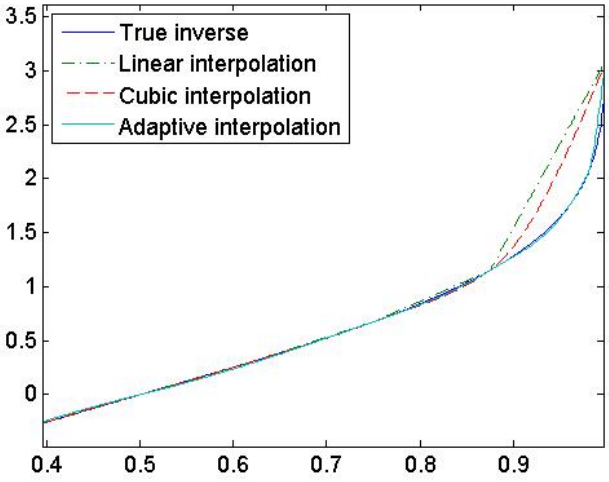


Figure 8 shows that the inverse standard normal distribution can be approximated rather accurately using only 5 precomputed inverse values using a non-equidistantly chosen grid in combination with a linear interpolation method. Since the generation of normally distributed random numbers can already be done rather fast and accurately, we simply use a direct generation, by using, for example, Moro's inversion, instead of using interpolation techniques.

The resulting algorithm for generating the variance process, which we call the adapted NCI algorithm, is as follows:

1. If $d > 1$:
   - Specify a non-equidistant grid $\mathcal{U} = \{0, \ldots, 1 - \delta\}$, where the parameter δ is used to avoid numerical difficulties;
   - Precompute the inverse chi-squared random distribution values, $F_{chi}^{-1}(U; d-1)$, with $U \in \mathcal{U}$;
   - Generate a uniformly distributed sample, U_ν, and use a linear interpolation method in combination with $F_{chi}^{-1}(U; d-1)$, for $U \in \mathcal{U}$, to obtain an approximate sample, $\widetilde{F}_{chi}^{-1}(U_\nu; d-1)$ of the chi-squared random variable with $d-1$ degrees of freedom;
   - Generate a standard normal random variable, Z;
   - Given $\widetilde{\nu}(t)$ compute λ, which can be found in Chapter 3.7;
   - Generate a sample for $\widetilde{\nu}(t + \Delta t)$ by computing
     $$\widetilde{\nu}(t + \Delta t) = C((Z + \sqrt{\lambda})^2 + \widetilde{F}_{chi}^{-1}(U_\nu; d-1)),$$
     where the constant C can be found in Chapter 3.7.
2. Otherwise:
   - Use the NCI algorithm of Haastrecht and Pelsser (2008) by using a non-equidistantly chosen grid, $\mathcal{U}$, in combination with a linear interpolation method.

Obviously, using a linear interpolation method in combination with a non-equidistant grid will result in much better computation times than using an advanced interpolation method in combination with an equidistantly chosen grid.

4.4 Contribution to the simulation of the asset price

To derive a simulation scheme of the asset price we use the Gaussian interest rate model, because of its simple form. Since our starting point is the two-factor HW model, we transform this model to the two-factor Gaussian model using Chapter 3.5. Chapter 3.6 is used to obtain the true correlations between the interest rate factors wand the asset price and the variance process.

We denote the correlation matrix between the stochastic variables with Σ, which has the following form

$$\Sigma = \begin{pmatrix} 1 & . & . & . \\ \rho_{y,x} & 1 & . & . \\ \rho_{y,\nu} & \rho_{x,\nu} & 1 & . \\ \rho_{y,S} & \rho_{x,S} & \rho_{\nu,S} & 1 \end{pmatrix}.$$

To simulate from such a model we compute the Cholesky decomposition of Σ, which is a matrix $\mathbf{Q}$ such that $\mathbf{Q}\mathbf{Q}^{tr} = \Sigma$. Since Σ is symmetric and positive semi-definite the Cholesky decomposition exists but is not unique.

The log-asset price has the following exact representation,

$$\log\left(S(t+\Delta t)\right) = \log\left(S(t)\right) + \int_t^{t+\Delta t} \left(r(u) - \frac{1}{2}\nu(u) - \delta\right) du + \int_t^{t+\Delta t} \sqrt{\nu(u)} dW^S(u).$$

To obtain a bias-free scheme for the asset price we first integrate the SDE of $x(t+\Delta t)$, $y(t+\Delta t)$ and $\nu(t+\Delta t)$, to yield

$$\begin{aligned} y(t+\Delta t) &= y(t) - b_y \int_t^{t+\Delta t} y(u)du + \sigma_y \int_t^{t+\Delta t} dW^y(u), \\ x(t+\Delta t) &= x(t) - b_x \int_t^{t+\Delta t} x(u)du + \sigma_x \int_t^{t+\Delta t} dW^x(u), \\ \nu(t+\Delta t) &= \nu(t) + \int_t^{t+\Delta t} \kappa\left(\xi - \nu(u)\right) du + \sigma_\nu \int_t^{t+\Delta t} \sqrt{\nu(u)} dW^\nu, \end{aligned}$$

where $r(t+\Delta t) = x(t+\Delta t) + y(t+\Delta t) + \varphi(t+\Delta t)$. Rearranging terms, gives

$$\int_t^{t+\Delta t} dW^y(u) = \frac{1}{\sigma_y}\left(y(t+\Delta t) - y(t) + b_y \int_t^{t+\Delta t} y(u)du\right), \tag{32}$$

$$\int_t^{t+\Delta t} dW^x(u) = \frac{1}{\sigma_x}\left(x(t+\Delta t) - x(t) + b_x \int_t^{t+\Delta t} x(u)du\right), \tag{33}$$

$$\int_t^{t+\Delta t} \sqrt{\nu(u)} dW^\nu(u) = \frac{1}{\sigma_\nu}\left(\nu(t+\Delta t) - \nu(t) - \kappa\xi\Delta t + \kappa \int_t^{t+\Delta t} \nu(u)du\right). \tag{34}$$

Result 4.1. *Next, using a Cholesky decomposition and (32), (33) and (34), we obtain the*

following exact representation of the asset price[19]

$$\begin{aligned}
\log\left(S(t+\Delta t)\right) &= \log\left(S(t)\right) + \int_t^{t+\Delta t}\left(r(u) - \frac{1}{2}\nu_S(u) - \delta\right)du \qquad (35)\\
&\quad + Q_{(4,1)}\int_t^{t+\Delta t} dW^y(u) + Q_{(4,2)}\int_t^{t+\Delta t} dW^x(u)\\
&\quad + Q_{(4,3)}\int_t^{t+\Delta t}\sqrt{\nu(u)}dW^\nu(u) + Q_{(4,4)}\int_t^{t+\Delta t}\sqrt{\nu(u)}dW^S(u),\\
&= \log\left(S(t)\right) + \int_t^{t+\Delta t}\varphi(u)du - \delta\Delta t \qquad (36)\\
&\quad + \left(\frac{b_y Q_{(4,1)}}{\sigma_y} + 1\right)\int_t^{t+\Delta t} y(u)du + \frac{Q_{(4,1)}}{\sigma_y}\left(y(t+\Delta t) - y(t)\right)\\
&\quad + \left(\frac{b_x Q_{(4,2)}}{\sigma_x} + 1\right)\int_t^{t+\Delta t} x(u)du + \frac{Q_{(4,2)}}{\sigma_x}\left(x(t+\Delta t) - x(t)\right)\\
&\quad + \left(\frac{\kappa Q_{(4,3)}}{\sigma_\nu} - \frac{1}{2}\right)\int_t^{t+\Delta t}\nu(u)du\\
&\quad + \frac{Q_{(4,3)}}{\sigma_\nu}\left(\nu(t+\Delta t) - \nu(t) - \kappa\xi\Delta t\right) + Q_{(4,4)}\int_s^t\sqrt{\nu}dW^S(u),
\end{aligned}$$

where W^y, W^x, W^ν and W^S are independent Brownian motions and $\int_t^{t+\Delta t}\varphi(u)du$ can be worked out using the results in Appendix C.

Remark. *The correlations between $S(t+\Delta t)$ and $x(t+\Delta t)$, $y(t+\Delta t)$ and $\nu(t+\Delta t)$ are mainly driven by $\frac{Q_{(4,1)}}{\sigma_y}y(t+\Delta t)$, $\frac{Q_{(4,2)}}{\sigma_x}x(t+\Delta t)$ and $\frac{Q_{(4,3)}}{\sigma_\nu}\nu(t+\Delta t)$, respectively. According to Andersen (2007), avoiding these terms will result in a poor correlation structure during the simulation process of the asset price. Thus, any discretization scheme should contain these terms. As W^S is independent of ν, conditional on $\nu(t)$ and $\int_t^{t+\Delta t}\nu(u)du$, the Itô integral $\int_t^{t+\Delta t}\sqrt{\nu}dW^S(u)$ is normally distributed with zero mean and variance equal to $\int_t^{t+\Delta t}\nu(u)du$. It should be clear using (36) that conditional on $x(t+\Delta t)$, $y(t+\Delta t)$,$\int_t^{t+\Delta t}x(u)du$ $\int_t^{t+\Delta t}y(u)du$, $\nu(t+\Delta t)$ and $\int_t^{t+\Delta t}\nu(u)du$, the distribution of $\log\left(S(t+\Delta t)\right)$ is normally distributed with easily computable moments.*

Representation (35) shows the presence an integrated interest rate process and an integrated variance process. It turns out that the simulation of the integrated variance process is much more involved than the integrated interest rate process, which is due to the different distributional properties of the variance and interest rate process. The importance of simulating these integrated processes as accurate as possible is of high importance in a exact simulation setting[20]. Using a discretization scheme the importance is less high because the simulation is then dependent on the number of time steps such that an interpolation method is already sufficiently accurate.

In many applications where the simulation of the short interest rate, r is required, values of the discount factor, $e^{-\int_0^t r(u)du}$, are required. Glasserman (2003) observes that the pair

[19]This approach also have to be used to impose the true correlation struction between the variance process and the first and second interest rate factor of the Gaussian interest rate model.

[20]In Appendix F we discuss some alternative simulation methods of the integrated variance process, however, these methods do not meet the performance of the method discussed by Glasserman and Kim (2008).

$\left(r(t), \int_0^t r(u)du\right)$ is jointly Gaussian and it is possible to simulate the paths of the pair without discretization error. However, the jointly simulation of $\left(r(t), \int_0^t r(u)du\right)$, where r is modeled by the two-factor Gaussian model, results in complex expressions due to the correlations and, therefore, we do not proceed in this way.

Furthermore, due to the complexity of the characteristic function of the integrated variance process, it is rather time-consuming to apply an exact generation method. Additionally, since we are interested in a discretization scheme of the asset price, an interpolation method is already sufficiently accurate according to Andersen (2007). For that reason, we approximate the integrated interest rate, $\int_s^t r(u)du$, and the integrated variance process, $\int_s^t \nu(u)du$, conditional on its begin and endpoints, by a drift interpolation. More specifically,

$$\int_s^t r(u)du \approx q_1(t-s)r(s) + q_2(t-s)r(t), \qquad \int_s^t \nu(u)du \approx q_1(t-s)\nu(s) + q_2(t-s)\nu(t),$$

where the Euler approximation is given by $q_1 = 1$ and $q_2 = 0$. The predictor-corrector method (central discretization) is given by $q_1 = q_2 = \frac{1}{2}$ and should be the default choice, because it is more accurate than the Euler discretization. In Appendix F we discuss some alternative simulation methods to generate $Iv(s,t)$, which, however, do not meet the performance of a (simple) drift interpolation in case of using a discretization method for simulating the asset price process.

By using the drift interpolation method to approximate the integrated interest rate and the integrated variance process results in the following discretization scheme for the asset price:

$$\begin{aligned}
\log\left(\tilde{S}(t+\Delta t)\right) &= \log\left(\tilde{S}(t)\right) + \int_t^{t+\Delta t} \varphi(u)du - \delta\Delta t \\
&\quad + \left(\frac{b_y Q_{(4,1)}}{\sigma_y} + 1\right)\left(\gamma_1^y \Delta t\, \widetilde{y}(t) + \gamma_2^y \Delta t\, \widetilde{y}(t+\Delta t)\right) + \frac{Q_{(4,1)}}{\sigma_y}\left(\tilde{y}(t+\Delta t) - \tilde{y}(t)\right) \\
&\quad + \left(\frac{b_x Q_{(4,2)}}{\sigma_x} + 1\right)\left(\gamma_1^x \Delta t\, \widetilde{x}(t) + \gamma_2^x \Delta t\, \widetilde{x}(t+\Delta t)\right) + \frac{Q_{(4,2)}}{\sigma_x}\left(\tilde{x}(t+\Delta t) - \tilde{x}(t)\right) \\
&\quad + \left(\frac{\kappa Q_{(4,3)}}{\sigma_\nu} - \frac{1}{2}\right)\left(\gamma_1^\nu \Delta t\, \widetilde{\nu}(t) + \gamma_2^\nu \Delta t\, \widetilde{\nu}(t+\Delta t)\right) \\
&\quad + \frac{Q_{(4,3)}}{\sigma_\nu}\left(\tilde{\nu}(t+\Delta t) - \tilde{\nu}(t) - \kappa\xi\Delta t\right) \\
&\quad + \sqrt{\Delta t} Q_{(4,4)} \sqrt{\left(\gamma_1^\nu \Delta t\, \widetilde{\nu}(t) + \gamma_2^\nu \Delta t\, \widetilde{\nu}(t+\Delta t)\right)}\, Z^S,
\end{aligned}$$

where Z^S is a standard Gaussian random variable. The deterministic integral $\int_t^{t+\Delta t} \varphi(u)du$ can also be approximated by a drift interpolation, i.e. $\int_t^{t+\Delta t} \varphi(u)du \approx \frac{1}{2}\Delta t\left(\varphi(t) + \varphi(t+\Delta t)\right)$. Although this integral can be computed analytically (see Appendix C), the drift interpolation is more efficient, because we use a discretization scheme for simulating the asset price process.

In more compact notation the scheme reads:

$$\begin{aligned}
\log\left(\tilde{S}(t+\Delta t)\right) &= \log\left(\tilde{S}(t)\right) + K_0 + K_1\widetilde{y}(t) + K_2\widetilde{x}(t) + K_3\widetilde{\nu}(t) + K_4\widetilde{y}(t+\Delta t) \\
&\quad + K_5\widetilde{x}(t+\Delta t) + K_6\widetilde{\nu}(t+\Delta t) + \sqrt{K_7\widetilde{\nu}(t) + K_8\widetilde{\nu}(t+\Delta t)}\, Z^S,
\end{aligned} \tag{37}$$

where

$$
\begin{aligned}
K_0 &= \int_t^{t+\Delta t} \varphi(u)du - \delta\Delta t - \frac{\kappa\xi Q_{(4,3)}\Delta t}{\sigma_\nu}, \qquad K_1 = \gamma_1^y \Delta t \left(\frac{b_y Q_{(4,1)}}{\sigma_y} + 1\right) - \frac{Q_{(4,1)}}{\sigma_y}, \\
K_2 &= \gamma_1^x \Delta t \left(\frac{b_x Q_{(4,2)}}{\sigma_x} + 1\right) - \frac{Q_{(4,2)}}{\sigma_x}, \qquad K_3 = \gamma_1^\nu \Delta t \left(\frac{\kappa Q_{(4,3)}}{\sigma_\nu} - \frac{1}{2}\right) - \frac{Q_{(4,3)}}{\sigma_\nu}, \\
K_4 &= \gamma_2^y \Delta t \left(\frac{b_y Q_{(4,1)}}{\sigma_y} + 1\right) + \frac{Q_{(4,1)}}{\sigma_y}, \qquad K_5 = \gamma_2^x \Delta t \left(\frac{b_x Q_{(4,2)}}{\sigma_x} + 1\right) + \frac{Q_{(4,2)}}{\sigma_x}, \\
K_6 &= \gamma_2^\nu \Delta t \left(\frac{\kappa Q_{(4,3)}}{\sigma_\nu} - \frac{1}{2}\right) + \frac{Q_{(4,3)}}{\sigma_\nu}, \qquad K_7 = \gamma_1^\nu \Delta t\, Q_{(4,3)}^2 \quad \text{and} \quad K_8 = \gamma_2^\nu \Delta t\, Q_{(4,3)}^2.
\end{aligned}
$$

The parameters γ_1^y, γ_2^y, γ_1^x, γ_2^x, γ_1^ν, γ_2^ν are preferable set to $\frac{1}{2}$, because then the predictor-corrector discretization method is used instead of the Euler discretization method.

In order to simulate the asset price, one needs to perform the following steps:

1. Set the parameters of the model;
2. Impose the correct correlation structure for x, y and ν by using the Cholesky decomposition and generate $(\widetilde{y}(t+\Delta t) \mid \widetilde{y}(t))$ and $(\widetilde{x}(t+\Delta t) \mid \widetilde{x}(t))$ as discussed in Chapter 4.2 and generate $(\widetilde{\nu}(t+\Delta t) \mid \widetilde{\nu}(t))$ using a method discussed in Chapter 4.1 and 4.3;
3. Conditional on $\widetilde{S}(t)$, $\widetilde{x}(t)$, $\widetilde{y}(t)$, $\widetilde{\nu}(t)$ and the values for $\widetilde{x}(t+\Delta t)$, $\widetilde{y}(t+\Delta t)$ and $\widetilde{\nu}(t+\Delta t)$ computed in step 3, compute $\widetilde{S}(t+\Delta t)$ from 37.

4.5 Martingale correction

Since we investigate MC methods for risk-neutral valuation, the scenarios we generate have to be arbitrage-free. Thus, the martingale condition under the risk-neutral measure

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^{t+\Delta t} r(u)du} S(t+\Delta t) \mid S(t)\right] = S(t) < \infty,$$

has to be satisfied. When we perform an exact simulation for both the integrated interest rate and variance this condition is fulfilled, because by definition the asset price is a martingale and an unbiased estimator is used. But in the case of using a discretization method the martingale condition is typically not fulfilled, because we approximate a continuous model by a discretization method. For a deterministic interest rate, martingale corrections are constructed for the QE scheme by Andersen (2007) and for the NCI scheme by Haastrecht and Pelsser (2008), so that the discretized asset price fulfills the martingale condition. They show that the accuracy of their schemes is improved significantly.

When we consider the adapted NCI scheme, the martingale correction is the same as the martingale correction constructed by Haastrecht and Pelsser (2008), because a representation of a non-central chi-squared random variable is used. We therefore refer to Haastrecht and Pelsser (2008) for more information about the martingale correction.

When we consider the extended Heston model, the construction of a martingale correction is much more involved due to the correlation between the interest rate and the variance process. Although the approach of Andersen (2007) seems not be applicable to construct a martingale correction, due to the correlated variables, we do not doubt that another approach could result in a (approximate) martingale correction. However, we do not proceed in this way and simply use the martingale corrections derived by Andersen (2007) and Haastrecht and Pelsser (2008) when dealing with the QE or the NCI/adapted NCI scheme, respectively.

Empirical martingale simulation

For the asset price to fulfill the martingale property we can also apply path adjustments to the generated asset price. When we generate M samples of length N_{MC} of the asset price at different times, i.e. $S(t_1), \ldots, S(t_M)$ then the discounted sample mean at different times coincides with the initial price. Duan and Simonato (1995) propose the empirical martingale simulation (EMS) method, which is a path adjusted method which transforms the simulated asset price to another asset price with the additional property that the martingale property is fulfilled.

Result 4.2. *The EMS method transforms the generated asset price $S_i(t_j)$ to an asset price $S_i^*(t_j)$ in the following way*

$$S_i^*(t_j) = S_0 \frac{Z_i(t_j)}{Z_0(t_j)}, \tag{38}$$

where

$$Z_i(t_j) = S_i^*(t_j)\frac{S_i(t_j)}{S_i(t_{j-1})} \qquad Z_0(t_j) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} e^{-rt_j} Z_i(t_j),$$

where r denotes the fixed interest rate. The resulting samples $S^*(t_j)$ fulfill the martingale property.

Duan and Simonato (1995) construct their path adjustment for fixed interest rate. This method can, however, be extended to stochastic interest rates by replacing the fixed discount factor e^{-rt_j} by

$$\begin{cases} e^{-\Delta t \sum_{k=1}^{j-1} r_i(t_k)}, & \text{when a Euler discretization is used;} \\ e^{-\frac{1}{2}\Delta t \sum_{k=1}^{j-1}(r_i(t_k)+r_i(t_{k+1}))}, & \text{when a central discretization is used} \end{cases}$$

where $i = 1, \ldots, N_{MC}$.

The EMS method can also be extended to stochastic variance. Duan and Simonato (1995) apply the EMS method to the BS model extended with stochastic variance driven by a GARCH model. We are, however, interested in the case when the variance is driven by a CIR process. This is more or less a trivial extension because the EMS method only adjusts the asset price, so the EMS method is independent of the generation of the asset price.

In order to perform the EMS method to the extended Heston model, one has to perform the following steps:

- Generate the interest rate and variance process up to a certain time T;
- Generate the asset price using the interest rate and variance scenarios up to time T;
- Compute the adjusted asset price S^* by computing

$$S_i^*(t_j) = S_0 \frac{Z_i(t_j)}{Z_0(t_j)}, \tag{39}$$

  where

$$Z_i(t_j) = S_i^*(t_j)\frac{S_i(t_j)}{S_i(t_{j-1})} \qquad Z_0(t_j) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} e^{-\int_0^{t_j} r(u)du} Z_i(t_j),$$

  where $e^{-\int_0^{t_j} r(u)du}$ is approximated by a drift interpolation method.

Although we have not corrected the interest rate scenarios, so they are not fully arbitrage free, the asset price exactly fulfills the martingale correction.

4.6 Numerical experiment: martingale correction versus EMS

We wish to test the martingale correction, which Andersen (2007) constructed for the QE scheme (QEmc), and the EMS method. We first simulate the asset price over a horizon of 10 years using $10,000$ simulations and then we discount the scenarios with respect to the interest rate. We then should get a flat line equal to the initial price of the asset price. The figure on the left-hand side in Figure 9 is obtained when we assume the interest rate equal to $r(t) = 0.05$ and use the Heston model in combination with Case III, $\Delta t = 1/4$ and $S(0) = 1$ and the figure on the right-hand side in Figure 9 is obtained when using the same parameter settings except modeling the interest rate by a two-factor HW model with parameters $a = 0.028$, $b = 0.1$, $\sigma_r = 0.005$, $\sigma_u = 0.001$, $r(0) = 0.043$ and $\rho_{r,u} = -0.1$, and for simplicity we take $\rho_{S,u} = \rho_{S,r} = \rho_{\nu,u} = \rho_{\nu,r} = 0$.

Figure 9: Effect of the martingale correction and the EMS method

(a) EMS versus QEmc

(b) EMS versus extended QEmc

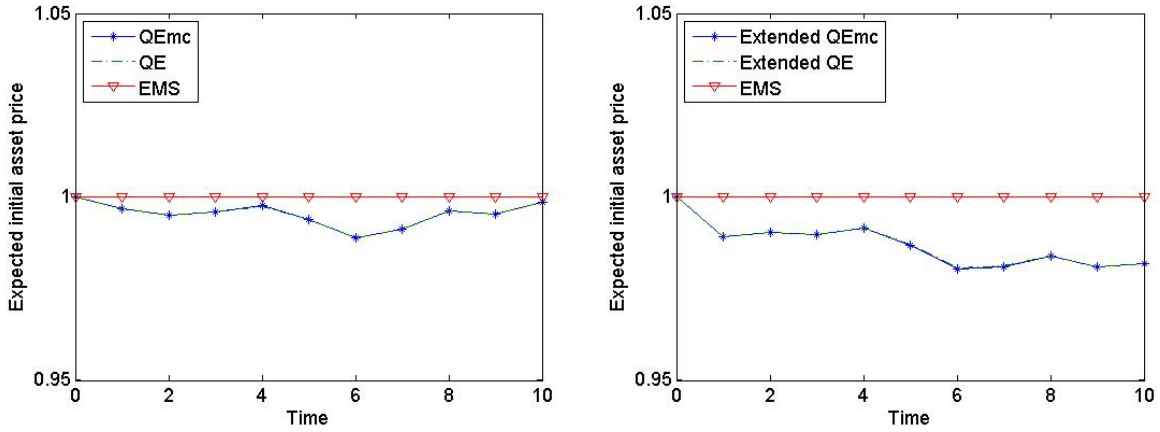


Figure 9 shows that the EMS exactly reproduces the initial asset price, i.e. without a discretization or a statistical error, either we use a fixed interest rate or a stochastic interest rate. Furthermore, we observe that the effect of the martingale correction is minor in this case. We note that when one uses more simulations, more accurate results are obtained using the martinale correction. Similar figures are obtained for Case I, II and IV.

4.7 Convergence

As already remarked, we distinguish between a discretization scheme and an unbiased estimator of the asset price at a certain time. For a discretization scheme the convergence property is a very important concept, i.e. the strong and weak order of convergence[21]. It turns out that the theoretically convergence properties of the QE scheme are very (if not impossible) hard to determine. Additionally, Andersen (2007) notes that an examination of the weak convergence is not always meaningful since higher-order moments do not have to exist. He focuses instead on weak consistency (see Kloeden and Platen (1999)) because according to Kloeden and Platen (1999) there is a direct link between weak consistency and weak convergence and it is theoretically more appealing. In other words, the weak consistency property is often more easily verified than the weak order of convergence. We refer to Andersen (2007) for proofs on weak consistency.

[21] For option pricing, the weak order of convergence is especially important.

The weak order of convergence can, however, be determined numerically. Glasserman and Kim (2008) show for instance that the QE scheme of Andersen (2007) has a weak order of convergence, which is defined as the value α if there exists a positive constant K and a positive constant Δ such that for fixed $T = \Delta t N$:

$$\left| \mathbb{E}\left[S(T)\right] - \mathbb{E}\left[\widetilde{S}(T)\right] \right| < K(\Delta t)^{\alpha},$$

for all $0 < \Delta t < \Delta$, approximately equal to 1.28, 1.88, 1.63 and 1.27 for cases I-IV[22].

Because of the complexity of the proofs on weak convergence and weak consistency, we do not discuss it here for the proposed discretization scheme for the asset price.

[22]Note that this is larger than the weak order of convergence of the Euler scheme, which is typically 1.

5 Monte Carlo and variance reduction techniques

5.1 Monte Carlo

In practice, MC simulation is a valuable tool for the valuation of (complex) options. It is relatively easy to price options using the simulation output of the underlying model, even for options that are complicated or high-dimensional. Another interesting property is that it provides one uniform simulation framework, i.e. one can run a risk-neutral MC simulation and price several (different) options at once.

In order to perform a MC simulation using a discretization scheme, $\widetilde{X}$, which typically denotes the asset price or the interest rate, we draw N_{MC} independent scenarios of $\widetilde{X}_1, \ldots, \widetilde{X}_{N_{MC}}$ using an equidistant time-grid with fixed time step Δt. Then the standard MC estimator is given by

$$\mu \approx \mu_{MC} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} h(\widetilde{X}_i), \tag{40}$$

where h denotes the (discounted) pay-off function of the option and $\mu = E^{\mathbb{Q}}\left[h\left(\widetilde{X}\right)\right]$, where $\mathbb{Q}$ denotes the risk-neutral measure. Note that the pay-off function can also be dependent on several assets.

The MC estimator converges according to the Strong Law of Large Numbers to μ as $N \to \infty$. It can be shown using the Central Limit Theorem that the convergence rate of the MC method is of order $\frac{1}{\sqrt{N}}$, which is the statistical error. So the statistical error is controlled by the number of simulations, i.e. when N tends to infinity then the statistical error tends to zero. Note that the convergence rate of a MC estimate does not depend on the dimension of the problem, which is the reason why the MC method is well suited for high-dimensional problems governed by path-dependent options or basket options.

Another attraction of the MC method is that it provides a confidence interval of level $1-\alpha$ for $\widetilde{\mu}_{MC}$, which is given by

$$\left[\mu_{MC} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\sigma^2_{h(\widetilde{X})}}{N_{MC}}}\,,\, \mu_{MC} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\sigma^2_{h(\widetilde{X})}}{N_{MC}}}\right],$$

where $\sigma^2_{h(\widetilde{X})}$ denotes the variance of $h(\widetilde{X}_i)$ $i = 1, \ldots, N_{MC}$ and Φ^{-1} denotes the inverse distribution function of the normal distribution.

5.2 Variance reduction techniques

A variance reduction technique is often used in practice as a tool to improve the accuracy of the standard MC estimator given in (40). As mentioned before the accuracy of the MC method heavily depends on the number of simulations, N_{MC}. A large number of simulations is usually needed to achieve a reasonable accuracy level. Variance reduction techniques are used to achieve a better accuracy and, thus, more reliable estimates, using (significantly) fewer simulations than the plain MC estimator.

Variance reduction techniques transform the (pay-off) function $h\left(\widetilde{X}\right)$, of which the expectation is to be computed, to another function, $\widehat{h}\left(\widetilde{X}\right)$ with the additional property that

$$\sigma^2_{\widehat{h}(\widetilde{X})} < \sigma^2_{h(\widetilde{X})}.$$

Using this function, $E\left[\widehat{h}\left(\widetilde{X}\right)\right]$ can be estimated using MC simulations. On the one hand, there is a group of variance reduction techniques that reduce the variance of the simulation output and, thus, the variance of the mean. On the other hand, there is group of variance reduction techniques that only reduce the variance of the mean and, as a consequence, not of the simulation output. Although (some) more work has to done to transform $h\left(\widetilde{X}\right)$ to $\widehat{h}\left(\widetilde{X}\right)$, the number of simulations is often decreased significantly, which results in lower computation times. Classical textbooks about variance reduction techniques are, for example, Glasserman (2003) and Jäckel (2002).

Many variance reduction techniques exist, but often they are problem dependent. For example, the importance sampling (IS) method (see Glasserman (2003)) heavily depends on the pay-off function of the option of interest. Especially for high-dimensional problems the application of the IS method is complex because an optimization algorithm has to be performed. Despite Capriotti (2007) constructs an efficient IS sampling method, using the extended Heston model for valuing embedded options, the application of the IS method is rather difficult, due to the complex structure of the pay-off functions. Furthermore, the stratified sampling (SS) method can be used (see Glasserman (2003)). Although the method is rather effective, it suffers the curse of dimensionality in high-dimensional problems. Therefore, we do not discuss the IS and the SS method is this thesis.

We have also investigated the quasi MC (qMC) method. After performing some numerical tests using qMC we found that these methods are not appropriate for valuing embedded options, because horizon of embedded options is often rather large (typically $50 - 100$ years), as a consequence, the application of a dimension reduction technique, such as the principal components method, to reduce the effective dimension is too time-consuming. Therefore, we do not discuss this method in this thesis. More information about qMC methods can be found in Glasserman (2003).

To value the embedded options discussed in Section 2.2 using MC simulations in an effective way, we are especially interested in generally applicable variance reduction techniques, such as antithetic sampling and moment matching. We briefly discuss these techniques as well as the control variate technique in the following chapters. In Chapter 6 we construct an efficient control variate for pricing a call or a put option using the Heston model extended with a one-factor HW interest rate model and in Chapter 7 we compare the effectiveness of the methods. More information about variance reduction techniques can be found in Glasserman (2003).

5.2.1 Antithetic sampling

The antithetic sampling method is one of the most widely used method in practice to reduce the variance of the standard MC method, because it is simple and easy to implement. The antithetic sampling method is based on the observation that if U is (standard) uniformly distributed, then $1-U$ is also (standard) uniformly distributed. This implies that, if one generates

a path using M independent uniform random numbers, $U_1, \ldots, U_M$, a second path is automatically generated by using $1 - U_1, \ldots, 1 - U_M$ without changing the law[23] of the process. By taking the average of the values of the two paths, a relatively large or small output computed from the first path is balanced by the value computed from the 'mirrored' path.

Using the approach discussed above, we can extend it to other distributions as well by using that $F^{-1}(U)$ and $F^{-1}(1-U)$ are both based on distribution function F, which is by definition a monotone function. In particular, when the distribution is symmetric, like the normal distribution, the values $F^{-1}(U)$ and $F^{-1}(1-U)$ have the same value but with opposite signs.

To apply the antithetic sampling method using a discretization scheme, one first generates two paths

$$\widetilde{X}^1 = \widetilde{X}(U) \quad \text{and} \quad \widetilde{X}^2 = \widetilde{X}(1-U),$$

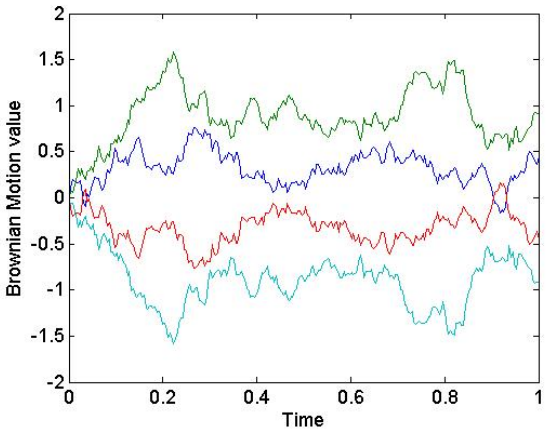
where $\widetilde{X}(U)$ denotes a scenario using $U_1, \ldots, U_M$. To obtain the antithetic path, we average the values of these two paths, i.e. the antithetic path is given by $\widetilde{X}^3 = \frac{1}{2}\left(\widetilde{X}^1 + \widetilde{X}^2\right)$. Doing this for N_{MC} scenarios, we get N_{MC} independent copies of $\widetilde{X}$, $\widetilde{X}_1^3, \ldots, \widetilde{X}_{N_{MC}}^3$. The unbiased antithetic sampling estimator of μ is given by

$$\mu \approx \mu_{AS} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} h\left(\widetilde{X}_i^3\right) = \frac{1}{2}\left(\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} h(\widetilde{X}_i^1) + \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} h(\widetilde{X}_i^2)\right), \tag{41}$$

It turns out according to Glasserman (2003) that the antithetic sampling method reduces the variance of the MC estimator when $Cov\left(h(\widetilde{X}_i^1), h(\widetilde{X}_i^2)\right) < 0$.

The advantage of this variance reduction technique is that it is generally applicable and that a relatively large variance reduction is often achieved for a relative small amount of extra work. As an illustration we show two mirrored Brownian motion paths in Figure 10. One path is obtained by a sequence $(Z_1, \ldots, Z_M)$, where every Z_i $(i = 1, \ldots, M)$ is standard normally distributed and the mirrored path is obtained by using $(-Z_1, \ldots, -Z_M)$.

Figure 10: Two mirrored Brownian motion paths



[23] Technically speaking, the law of a stochastic process is the measure that the process induces on the collection of functions from the index set into the state space.

Using these 4 paths, 2 antithetic paths are obtained by averaging. From Figure 10 it becomes clear that the variance of the antithetic paths is smaller than the variance of the total scenarios at some time t, $0 \leq t \leq 1$.

5.2.2 Control variates

The control variate technique is very popular because a large variance reduction can be obtained in a simple theoretical framework. The method takes advantage of random variables with known expected values and positive correlations with the variable under consideration, $h(\widetilde{X})$.

Let $g(\widetilde{X})$ be a random variable with known mean, which is denoted by μ_g, where g denotes a (discounted) pay-off function dependent on $\widetilde{X}$. Furthermore, let $\widetilde{Y}$ denote a given discretization scheme of a quantity Y and let $h(\widetilde{Y})$ be a random variable with known mean, which is denoted by μ_h, where h denotes a (discounted) pay-off function dependent on $\widetilde{Y}$. Now, for each trial the outcome of $g(\widetilde{X}_i)$ or $h(\widetilde{Y}_i)$ is calculated along with the output $h(\widetilde{X}_i)$. Furthermore, we suppose that the pairs $\left(h(\widetilde{X}_i), g(\widetilde{X}_i)\right)$ and $\left(h(\widetilde{X}_i), h(\widetilde{Y}_i)\right)$, $i = 1, \ldots, N_{MC}$ are identically independently distributed (i.i.d.). The control variate estimator, μ_{CV}, is then given by either

$$\mu \approx \mu_{CV} = \begin{cases} \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left(h(\widetilde{X}_i) - g(\widetilde{X}_i) + \mu_g\right), \\ \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left(h(\widetilde{X}_i) - h(\widetilde{Y}_i) + \mu_h\right). \end{cases} \tag{42}$$

The control variate estimator can be improved by introducing a parameter $\beta \in \mathbb{R}$, which is chosen in such a way that the variance of the control variate estimator μ_{CV} is minimized. The unbiased parameterized control variate estimator is then given by either

$$\mu \approx \mu_{CV}(\beta) = \begin{cases} \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} h(\widetilde{X}_i) - \beta \left(\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} g(\widetilde{X}_i) - \mu_g\right), \\ \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} h(\widetilde{X}_i) - \beta \left(\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} h(\widetilde{Y}_i) - \mu_h\right). \end{cases}$$

The optimal value, β^*, depends on unknown quantities and, thus, one has to perform a pilot study in order to obtain an approximation for it.

Although this technique is very effective in reducing the variance of the MC estimator and relatively easy to implement, the construction of a control variate is rather complex. This is mainly due to the fact that the control variate has to be correlated with the stochastic variable of interest in order to be effective. Effective control variates of the profit sharing embedded option can be found in, for example, Plat and Pelsser (2008). These control variates can also be used as analytical approximations.

5.2.3 Moment matching

The moment matching method relies on the idea that if a sample of, for example, a standard normal distribution is generated then the moments of the sample do not match the analytical moments. In general, to overcome this problem one has to proceed as follows: take a sample $\mathbf{X} = (X_1, \ldots, X_N)$ of length N of a given distribution with mean μ and standard deviation σ then to match the moments of the sample one has to compute:

$$\mathbf{X}' = \left(X_1', \ldots, X_N'\right) = \left((X_1 - \mu_X)\frac{\sigma}{\sigma_X} + \mu, \ldots, (X_N - \mu_X)\frac{\sigma}{\sigma_X} + \mu\right),$$

where μ_X and σ_X are the sample mean and sample standard deviation, respectively.

For example, the moments of a generated sample of the standard normal distribution, $\mathbf{Z} = (Z_1, \ldots, Z_N)$ are matched by constructing the following sequence

$$\mathbf{Z}' = \left(Z_1', \ldots, Z_N'\right) = \left(\frac{Z_1 - \mu_Z}{\sigma_Z}, \ldots, \frac{Z_N - \mu_Z}{\sigma_Z}\right),$$

where μ_Z is the sample mean and σ_Z is the standard deviation of the sample. This sequence has zero mean and variance equal to 1.

Since the distribution of the interest rate and the variance processes are known in advance, we can match the moments of the sample with the true moments. The moment matched algorithm to match the moments of the interest rate is as follows:

- Use the proposed simulation scheme, which is discussed in Chapter 4.2, to generate a sample, $R_1, \ldots, R_{N_{MC}}$, of the interest rate process at a certain time $t \in \mathcal{I}$ of length N;
- Compute the analytic mean and variance of the interest rate at time t conditional on time $t = 0$, μ_r and σ_r^2, respectively, using Result 3.2;
- Create a new sequence by computing $\left((R_1 - \mu_R)\frac{\sigma_r}{\sigma_R} + \mu_r, \ldots, (R_{N_{MC}} - \mu_R)\frac{\sigma_r}{\sigma_R} + \mu_r\right)$, where μ_R and σ_R^2 denote the mean and the variance of the generated sample.

We follow the same approach to match the moments of the variance process. The algorithm then reads,

- Use a simulation scheme, which is discussed in Chapter 4.1 and 4.3, to generate a sample, $V_1, \ldots, V_{N_{MC}}$, of the variance process at a certain time $t \in \mathcal{I}$ of length N;
- Compute the analytic mean and variance of the variance at time t, μ_ν and σ_ν^2, respectively, using (14) and (15), conditional on $t = 0$;
- Create a new sequence by computing $\left((V_1 - \mu_V)\frac{\sigma_\nu}{\sigma_V} + \mu_\nu, \ldots, (V_{N_{MC}} - \mu_V)\frac{\sigma_\nu}{\sigma_V} + \mu_\nu\right)$, where μ_V and σ_V^2 denote the mean and the variance of the generated sample.

Furthermore, we wish to match the mean and variance of the asset price. When the asset price is modeled by a two-factor HWBS model with deterministic variance process the analytic moments can be derived as explained in Chapter 3. However, using the extended Heston model the moments of the asset price are much harder, if not impossible, to obtain analytically. Therefore, we do not proceed in this way.

In case the asset price is modeled by a two-factor HWBS model we obtain the following algorithm for matching the moments

- Use a simulation scheme to generate a sample, $S_1, \ldots, S_{N_{MC}}$, of the asset price process at a certain time $t \in \mathcal{I}$ of length N;
- Compute the analytic mean and variance of the asset price at time t, μ_s and σ_s^2, respectively, by extending the results in Brigo and Mercurio (2006, Appendix B), conditional on $t = 0$;
- Create a new sequence by computing $\left((S_1 - \mu_S)\frac{\sigma_s}{\sigma_S} + \mu_s, \ldots, (S_N - \mu_S)\frac{\sigma_s}{\sigma_S} + \mu_s\right)$, where μ_S and σ_S^2 denote the mean and the variance of the generated sample.

Matching the moments of the interest rate, variance or the asset price process is only meaningful if one is interested in one of these processes. For example, when one wants to value an option written on the interest rate or the asset price.

6 Intermezzo: calibration using Monte Carlo simulations

A calibration procedure is an optimization step to estimate the model parameters in such a way that market prices are replicated as good as possible. Market prices of plain vanilla options are often used for calibration, because much market data is available for these options. The calibration procedure can be viewed as the following optimization problem:

$$\begin{aligned}
&\min\left\{\|C-\widehat{C}\|_p\right\} = \min\left\{\left(\sum_{j=1}^{m}\sum_{k=1}^{n}\left|C(T_j,K_k)-\widehat{C}(T_j,K_k)\right|^p\right)^{\frac{1}{p}}\right\},\\
&s.t. \qquad a,\, b,\, \kappa,\, \sigma_r,\, \sigma_u,\, \sigma_\nu, \xi,\, S(0),\, r(0),\, u(0) \geq 0\\
&\qquad\quad \rho_{S,r},\, \rho_{S,u}\, \rho_{S,\nu}\, \rho_{r,u},\, \rho_{r,\nu},\, \rho_{u,\nu} \in [-1,1],
\end{aligned}$$

where $\|\cdot\|_p$ denotes the p-norm, $C(T_j,K_j)$ and $\widehat{C}(T_j,K_j)$ denotes the market price and its approximation with maturity T_j and strike K_k, with $j=1,\ldots,m$ and $k=1,\ldots,n$, respectively.

This optimization problem is often solved iteratively using a numerical optimization algorithm. Therefore, the pricing of plain vanilla options has to be done as fast as possible. Nowadays, very fast valuation methods are available, see for example Fang and Oosterlee (2008), who use their COS method to obtain (semi) analytic options. The only drawback of these methods is that the characteristic function should exist in closed form, otherwise time consuming MC simulations have to be used.

Typically, using the Heston model extended with a one-factor HW model, which we call the Heston Hull-White (HHW) model, to model the asset price, the characteristic function is only available in special cases. It is proven in Muskulus *et al.* (2007) that the characteristic function is available in closed form when we set some restrictions on parameters such as:

- Interest rate process is independent of asset price and variance process, i.e. $\rho_{r,S}=\rho_{r,\nu}=0$;
- The volatility of variance (σ_ν) is equal to $2\sqrt{\xi\kappa}$;
- $\rho_{S,\nu}=\frac{4\kappa^2+\sigma_\nu^2}{4\kappa\sigma_\nu}$ and $\rho_{\nu,r}=\frac{2\kappa}{\sigma_\nu}\rho_{S,r}$;
- The volatility of the interest rate process is also driven by the same square root process as the asset price, i.e. $dr(t)=(\theta(t)-ar(t))dt+\sigma_r\sqrt{\nu(t)}dW^r(t)$.

In the case one of these restrictions is found to be realistic, one could price plain vanilla options rather fast using the characteristic function in combination with, for example the COS method. In the case these restriction are not realistic one has to use other numerical techniques. To overcome these shortcomings, one could model the variance process by the Schöbel-Zhu model instead of the CIR model. The advantage of this model is that the characteristic function is available in closed form for all parameter settings (see Grzelak and Oosterlee *et al.* (2008)), so that calibration on plain vanilla options can be done very fast using, for example, the COS method. A drawback of this model is according to the literature that it does not fit the volatility smile as good as the Heston model.

Although closed-form approximations of the characteristic function can be found, in the case of the HHW model, MC simulations seem to be an important alternative for calibration since they are easily and generally applicable. Additionally, finding numerical approximations of the characteristic function are difficult due to the Feller condition (16), which makes numerical

approximations rather unstable for certain parameter settings. As already mentioned, these MC simulations are time-consuming (especially when a high dimensional model is used). Consequently, the calibration procedure is time-consuming. Using an appropriate discretization scheme, as discussed in the previous section, scenarios of the asset price can be generated relatively cheaply. For fixed maturity options an exact simulation, such as the scheme of Glasserman and Kim (2008), may be preferable because only a sample of the asset price is needed at maturity time. In combination with variance reduction techniques very accurate results can be obtained for a given computational budget. In the following new Result we propose an efficient control variate to improve the MC estimator.

Result 6.1. *Using MC simulations for calibration using the HHW model, one could make use of the HHW model, with independent interest rate, i.e.* $\rho_{S,r} = 0$ *and* $\rho_{\nu,r} = 0$*, as a control variate to reduce the variance of the MC estimator. For this particular model, the characteristic function is known in closed-form and therefore, fast valuation techniques, can be used to obtain (semi) analytic prices. The MC algorithm using a control variate to obtain the price of a plain vanilla option is then as follows:*

1. *Set the initial parameter settings of the model;*
2. *Compute the exact price,* μ_{exact}*, of the plain vanilla option of the HHW model with independent interest rate using, for example, the COS method (see Fang and Oosterlee (2008));*
3. *Draw* N_{MC} *independent scenarios of* $\widetilde{X}_1(T), \ldots, \widetilde{X}_{N_{MC}}(T)$ *of the HHW model using an appropriate simulation scheme and draw* N_{MC} *independent scenarios of* $\widetilde{Y}_1(T), \ldots, \widetilde{Y}_{N_{MC}}(T)$ *using the HHW model with independent interest rate using the same simulation scheme, where* $\widetilde{X}$ *and* $\widetilde{Y}$ *denote discretizations of the asset price (using different models);*
4. *Compute the MC estimator* $\frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} \left(h\left(\widetilde{X}_k(T)\right) - h\left(\widetilde{Y}_k(T)\right) + \mu_{exact} \right)$ *using* $h\left(\widetilde{Y}(T)\right)$ *as a control variate*[24]*.*

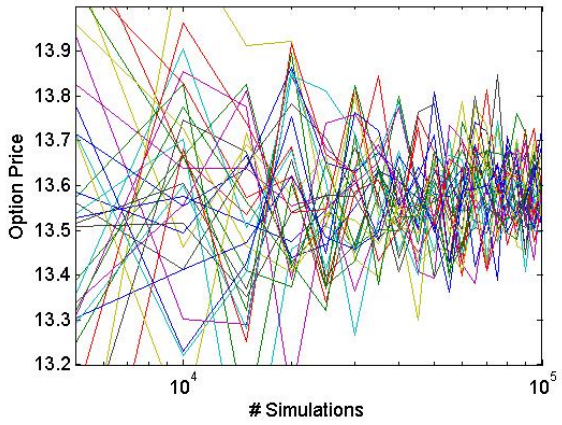
Since the correlation between $h(\widetilde{X})$ *and* $h(\widetilde{Y})$ *is high (because there is only a restriction on two correlation parameters), the variance reduction is high. This will result in a huge variance reduction.*

To show the effectiveness of this approach, we value an at-the-money call option using plain MC and MC with the control variate approach discussed above. We model the interest rate by a one-factor HW model with a fixed value of $\theta(t)$. More specifically, we assume $\theta(t) = 0.04$, $a = 1$, $\kappa = 6$, $\xi = 0.09$, $\sigma_r = 0.2$, $r(0) = 0.04$, $\nu(0) = 0.09$, $S(0) = 100$, $\rho_{S,r} = -0.1$, $\rho_{S,\nu} = 0.3$ and $\rho_{r,\nu} = 0.05$. We then get the following convergence in Figure 11:

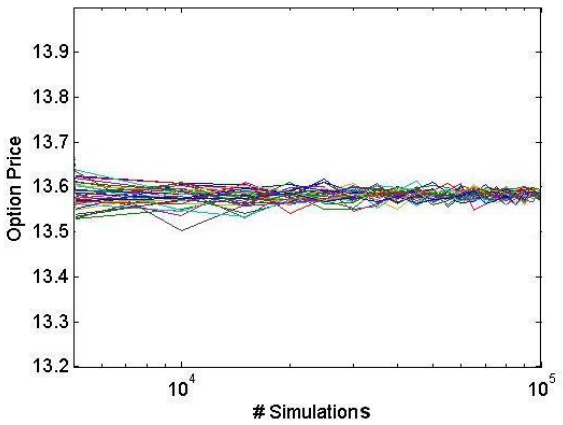
[24]Note that other variance reduction techniques can be used as well.

Figure 11: Convergence figures for the HHW model

(a) Plain Monte Carlo



(b) Control variate approach



The exact price is approximately equal to 13.58. Figure 11 shows that the control variate approach is very effective in reducing the variance of the plain MC estimator. Whereas the MC method needs more than $100,000$ simulations in order to obtain an accurate price of the option, the control variate approach already achieves accurate results after less than $10,000$ simulations.

7 Numerical results

In this section we show the results of numerical experiments. We compare the proposed adapted NCI scheme with the QE scheme of Andersen (2007) by pricing a call option using the Heston model. We test variance reduction techniques on reducing the variance of the MC estimator. The proposed simulation of the interest rate is compared with the Euler scheme by approximating the analytical statistics of the interest rate. Furthermore, the QE scheme is compared with the Euler scheme by approximating the analytical statistics of the variance process. Last but not least, we value the profit sharing embedded option, which is discussed in Chapter 2.2, by using the proposed simulation scheme of the interest rate and the Euler discretization scheme.

7.1 Valuing a call option using the Heston model

To test the adapted NCI scheme, we value a European call option, which is written on one asset modeled by the Heston model. This is a standard test case, because prices can be computed very accurately using for example the approach of Fang and Oosterlee (2008). Since these prices can be obtained very fast and accurately, we have reliable reference values.

We consider a call option C with maturity T and strike level K and we denote the exact initial call price by $C(0)$. Since we use a discretization scheme, we approximate the asset price at time T, $S(T)$, by $\widetilde{S}(T)$ and an approximation, $\widetilde{C}(0)$, of the call option is given by computing the expectation with respect to the risk-neutral measure, $\mathbb{Q}$,

$$\widetilde{C}(0) = \mathbb{E}^{\mathbb{Q}}\left[e^{-r\Delta t}\max\left(\widetilde{S}(T) - K, 0\right)\right].$$

Because of the fact that we use a discretization scheme, the value $\widetilde{C}(0)$ is generally not equal to $C(0)$ and therefore, we define the bias of a discretization scheme as follows

$$e = C(0) - \widetilde{C}(0).$$

Obviously, e is a function of the time step Δt which is used in the discretization scheme. We are especially interested in determining the function $e(\Delta t)$ in a numerical way, for different values of Δt for the discretization schemes considered in this chapter.

To estimate $\widetilde{C}(0)$ we use the MC method, which is discussed in Section 5 and is given by

$$\widetilde{C}(0) \approx \frac{1}{N_{MC}}\sum_{i=1}^{N_{MC}} e^{-r\Delta t}\max\left(\widetilde{S}_i(T) - K, 0\right).$$

Choosing N_{MC} sufficiently high, the variance of the MC estimator can be kept relatively low and an accurate approximation of $\widetilde{C}(0)$ is obtained[25].

As a test case we use the parameter settings of Case IV in Table 1 together with $\nu(0) = \xi$, $S(0) = 100$, $r = 0$ and maturity time $T = 1$. In our numerical results, we use the adapted NCI scheme and the QE scheme (QE) of Andersen (2007). The reason why we only compare the adapted NCI scheme to the QE scheme is that in the literature already many numerical experiments are performed on different discretization schemes, including the Euler discretization scheme, which is our starting point, and it turns out that the QE scheme is one of the best schemes in terms of efficiency. For the QE scheme we use the same parameter settings as used

[25]To measure the accuracy of the MC estimator one can compute confidence intervals.

in Andersen (2007), i.e. $\gamma_1 = \gamma_2 = \frac{1}{2}$ and $\psi_c = 1.5$. As discussed in Andersen (2007), the QE scheme can be improved by a martingale correction term. However, we only discuss the case without a martingale correction term. To apply the adapted NCI scheme, we have specified the grid $\mathcal{U}$, which is discussed in Chapter 4.1.4, by trial and error.

In Table 3 we report the estimated bias and the corresponding standard errors[26] in the parenthesis. The results are obtained using the computation program Matlab. We use $N_{MC} = 100,000$ simulations to value 50 options and we vary Δt from 1/32 year to 1/4 year, the exact price is given by $C(0) = 5.266$.

Table 3: Estimated bias (e) in test Case IV with standard errors in parenthesis

Δt	QE	Adapted NCI
1/4	0.5696 (0.0037)	0.6441 (0.0041)
1/8	0.2891 (0.0032)	0.3680 (0.0040)
1/16	0.1431 (0.0029)	0.2291 (0.0036)
1/32	0.0665 (0.0027)	0.1590 (0.0034)

The results in Table 3 show that the QE scheme is more accurate for all time steps than the adapted NCI scheme. We observe that the bias of both schemes converges very rapidly as the time step decreases.

In comparing the numerical efficiency of the discretization schemes, we need to consider both the bias of the schemes, as well as the time to compute the price of a call option using MC simulations. We therefore measure the computation time which is needed to price a call option, using $N = 100,000$ simulations. We get the results in Table 4:

Table 4: Computation times

Δt	QE	Adapted NCI
1/4	50.8	6.1
1/8	110.4	13.5
1/16	236.8	28.4
1/32	511.0	57.5

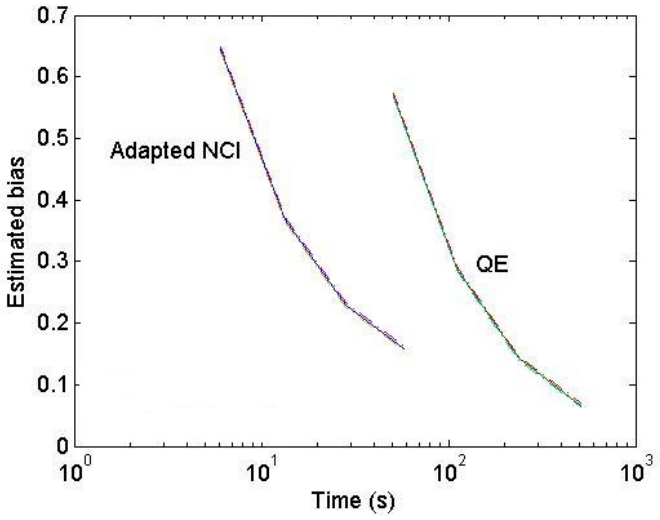
From Table 4 we conclude that the adapted NCI scheme is significantly faster than the QE scheme. Obviously, this is due to the fact that only a linear interpolation is performed in combination with the generation of uniform random numbers which can be done rather efficient.

Whereas the QE scheme is more accurate than the adapted NCI scheme for a given time step, the adapted NCI scheme is more efficient. For example, approximately the same accuracy is obtained using the QE scheme with a time step $\Delta t = 1/8$ and the adapted NCI scheme with $\Delta t = 1/12$, however the adapted NCI scheme is still approximately a factor 5 faster than the QE scheme, which is significant.

When we visualize the efficiency of the adapted NCI and the QE scheme we get the results in Figure 12.

[26]The standard error is defined by the standard deviation in the sample mean divided by the square root of the number of samples.

Figure 12: Measuring the efficiency with corresponding confidence bounds



From Figure 12 we conclude that the adapted NCI outperforms the QE scheme for this particular test case.

Remark. *We note that these experiments are performed in Matlab, so for simplicity we have used build-in functions to simulate the asset price using the adapted NCI and the QE scheme. We found that we were very limited by using Matlab for performing numerical experiments based on Monte Carlo. Performing experiments in Matlab is rather time-consuming and the memory of Matlab is limited, which is the reason why we choose* $100,000$ *simulations and* 50 *options, which might not be enough for fair comparisons. Additionally, by using build-in functions, the computation time of pricing a call option using the adapted NCI and the QE scheme, is affected. Therefore, a more fair comparison of the methods is obtained in, for example, C++. However, this is left for future research.*

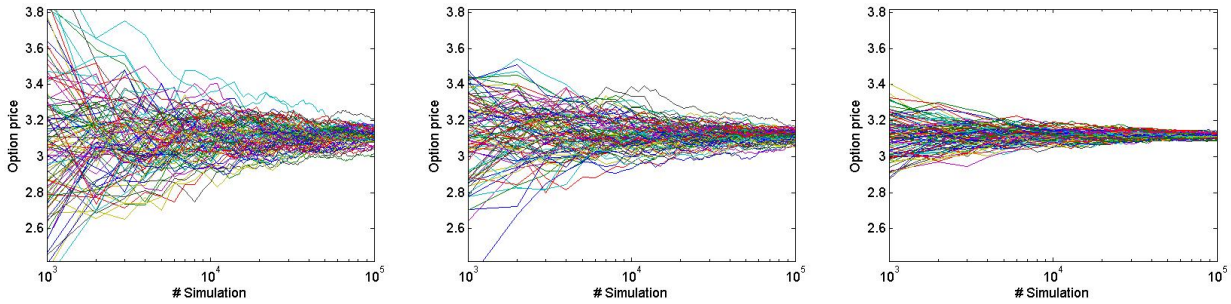
7.2 Variance reduction techniques

In Section 5 we have discussed several variance reduction techniques. To value embedded options or complex options using a high-dimensional underlying model it is convenient to apply generally applicable variance reduction techniques, for example the antithetic sampling method and the moment matching technique. To show the effectiveness of these methods we value an out-of-the-money call option using the standard BS model with $S(0) = 100$, $r = 0.05$, $\sigma = 0.3$ and strike level $K = 140$. We apply the standard MC method, the antithetic sampling method and the moment matching technique. The moment matching technique is applied by matching the moments of the generated sample with the analytic moments of the BS model.

After performing some numerical tests using the importance sampling method, the stratified sampling method and the control variate method we find that major variance reductions are obtained using this simple experiment, however since the antithetic sampling method and the moment matching technique are (often) generally applicable, we only apply these methods. We get the convergence behavior in Figure 13:

Figure 13: Convergence figures of several estimators

(a) Plain Monte Carlo (b) Antithetic sampling (c) Moment matching



These figures show that the variance of the MC estimator is reduced significantly by using the antithetic sampling method and the moment matching technique. Using these variance reduction techniques leads to much more reliable estimates.

7.3 Approximation and comparison of statistics

In Chapter 1.2 we discussed that our starting point is the Euler discretization scheme. In this chapter we want to compare the proposed simulation schemes with the Euler scheme by first approximating the analytical statistics. In Chapter 4.2 we discussed the simulation of the interest rate model. Although the two-factor Gaussian model has a more simpler form than the two-factor HW model, which is convenient for simulating the asset price, we use the second algorithm of Chapter 4.2.1 for simulation, which is based on the two-factor HW model. The reason is that using this simulation method the full risk-neutral model, which is discussed in Appendix A, can be simulated more efficiently. Furthermore, although we have proposed the adapted NCI scheme for simulating the variance process, we consider in this chapter the QE scheme of Andersen (2007), because this simulation method is already extensively tested in the literature whereas the adapted NCI scheme is not.

We compare the mean and the standard deviation of the proposed simulation method of the interest rate, given by the second algorithm of Chapter 4.2.1, at every year with the mean and standard deviation of the Euler discretization scheme, and with the analytic mean and standard deviation of the interest rate. We use the two-factor HW model with $T = 10$, $a = 0.028$, $b = 0.1$, $\sigma_r = 0.005$, $\sigma_u = 0.001$, $r(0) = 0.043$, $\rho_{r,u} = -0.1$ and with an appropriate function of the instantaneous forward rates, $f(0,t)$. Furthermore, we use $5,000$ simulations in combination with antithetic sampling and for every method the same random seed is used and, therefore we have not computed confidence intervals. We get the following results in Table 5, where the mean is given on the left-hand side and the standard deviation on the right-hand side:

Table 5: Comparison of mean and standard deviation of the interest rate

Time	Analytic	Proposed ($\Delta t = 0.01$)	Proposed ($\Delta t = 1$)	Euler ($\Delta t = 0.01$)
1	$5.81 \cdot 10^{-2}$,$5.20 \cdot 10^{-3}$	$5.81 \cdot 10^{-2}$,$5.14 \cdot 10^{-3}$	$5.81 \cdot 10^{-2}$,$5.14 \cdot 10^{-3}$	$5.81 \cdot 10^{-2}$,$5.30 \cdot 10^{-3}$
2	$5.05 \cdot 10^{-2}$,$7.20 \cdot 10^{-3}$	$5.05 \cdot 10^{-2}$,$7.18 \cdot 10^{-3}$	$5.05 \cdot 10^{-2}$,$7.20 \cdot 10^{-3}$	$5.05 \cdot 10^{-2}$,$7.20 \cdot 10^{-3}$
3	$5.15 \cdot 10^{-2}$,$8.90 \cdot 10^{-3}$	$5.15 \cdot 10^{-2}$,$8.84 \cdot 10^{-3}$	$5.14 \cdot 10^{-2}$,$8.90 \cdot 10^{-3}$	$5.14 \cdot 10^{-2}$,$8.80 \cdot 10^{-3}$
4	$4.87 \cdot 10^{-2}$,$1.04 \cdot 10^{-2}$	$4.86 \cdot 10^{-2}$,$1.02 \cdot 10^{-2}$	$4.87 \cdot 10^{-2}$,$1.02 \cdot 10^{-2}$	$4.87 \cdot 10^{-2}$,$1.02 \cdot 10^{-2}$
5	$4.85 \cdot 10^{-2}$,$1.17 \cdot 10^{-2}$	$4.85 \cdot 10^{-2}$,$1.17 \cdot 10^{-2}$	$4.85 \cdot 10^{-2}$,$1.15 \cdot 10^{-2}$	$4.85 \cdot 10^{-2}$,$1.16 \cdot 10^{-2}$
6	$4.89 \cdot 10^{-2}$,$1.31 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$,$1.31 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$,$1.28 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$,$1.29 \cdot 10^{-2}$
7	$4.98 \cdot 10^{-2}$,$1.44 \cdot 10^{-2}$	$4.98 \cdot 10^{-2}$,$1.44 \cdot 10^{-2}$	$4.97 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$	$4.97 \cdot 10^{-2}$,$1.44 \cdot 10^{-2}$
8	$5.12 \cdot 10^{-2}$,$1.57 \cdot 10^{-2}$	$5.11 \cdot 10^{-2}$,$1.57 \cdot 10^{-2}$	$5.11 \cdot 10^{-2}$,$1.55 \cdot 10^{-2}$	$5.11 \cdot 10^{-2}$,$1.57 \cdot 10^{-2}$
9	$5.22 \cdot 10^{-2}$,$1.69 \cdot 10^{-2}$	$5.22 \cdot 10^{-2}$,$1.70 \cdot 10^{-2}$	$5.22 \cdot 10^{-2}$,$1.67 \cdot 10^{-2}$	$5.22 \cdot 10^{-2}$,$1.70 \cdot 10^{-2}$
10	$5.36 \cdot 10^{-2}$,$1.82 \cdot 10^{-2}$	$5.36 \cdot 10^{-2}$,$1.82 \cdot 10^{-2}$	$5.36 \cdot 10^{-2}$,$1.79 \cdot 10^{-2}$	$5.36 \cdot 10^{-2}$,$1.83 \cdot 10^{-2}$

This table shows that when using $\Delta t = 1$ using the exact simulation scheme comparable results are obtained when using the Euler scheme using $\Delta t = 100$. Obviously, this results in major computation time reductions. When using a larger sample, the statistics will be even more comparable to the analytical statistics.

Next, we use the FT discretization scheme of Lord and Koekkoek (2008) and compare the mean and standard deviation with the QE scheme of Andersen (2007) and with the analytic mean and standard deviation. We use $\nu(0) = 0.04$, $\xi = 0.04$, $\kappa = 1$ and $\sigma_\nu = 0.1$, we then get the following results in Table 6, where the mean is given on the left-hand side and the standard deviation on the right-hand side:

Table 6: Comparison of mean and standard deviation of the variance process

Time	Analytic	QE ($\Delta t = 0.01$)	QE ($\Delta t = 0.1$)	FT ($\Delta t = 0.01$)
1	$4.00 \cdot 10^{-2}$,$1.32 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.31 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.30 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.30 \cdot 10^{-2}$
2	$4.00 \cdot 10^{-2}$,$1.40 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.40 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.39 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$
3	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$,$1.37 \cdot 10^{-2}$	$4.01 \cdot 10^{-2}$,$1.44 \cdot 10^{-2}$
4	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.45 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.43 \cdot 10^{-2}$
5	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$,$1.39 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.43 \cdot 10^{-2}$
6	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$,$1.40 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$,$1.39 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.44 \cdot 10^{-2}$
7	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.43 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$
8	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$
9	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.43 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$
10	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.42 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$,$1.41 \cdot 10^{-2}$	$4.01 \cdot 10^{-2}$,$1.46 \cdot 10^{-2}$

The results in Table 6 show that the analytic mean and variance are approximated very well by the QE scheme and the FT scheme. Note that the QE scheme with $\Delta t = 0.1$ is also very accurate.

Furthermore, we compare the mean and the variance of the current discretization scheme of the asset price, thus, using a deterministic variance process, with the mean and variance using the exact simulation scheme in combination with a central discretization to approximate the integrated interest rate process. We use the same parameters for the interest rate as before together with $S(0) = 1$, $\sigma_S = 0.1$, $\rho_{S,r} = 0.05$, $\rho_{S,u} = -0.05$ and a horizon of five years. We use $10,000$ simulations using antithetic sampling and for every method the same random seed is used and, therefore we have not computed confidence intervals. We get the following results

in Table 7:

Table 7: Comparison of the mean and standard deviation of the return on asset price

T	Analytical	Euler ($\Delta t = 0.01$)	proposed ($\Delta t = 0.02$)	proposed ($\Delta t = 0.1$)
1	0.0539 , 0.1058	0.0538 , 0.1059	0.0537 , 0.1053	0.0538 , 0.1065
2	0.0535 , 0.1059	0.0536 , 0.1069	0.0535 , 0.1063	0.0534 , 0.1051
3	0.0516 , 0.1058	0.0518 , 0.1063	0.0516 , 0.1065	0.0516 , 0.1059
4	0.0498 , 0.1058	0.0497 , 0.1049	0.0498 , 0.1063	0.0498 , 0.1056
5	0.0485 , 0.1058	0.0485 , 0.1057	0.0486 , 0.1072	0.0485 , 0.1058

We remark that the statistical fluctuations are due to the relatively low number of simulations. These results show that the proposed simulation of the interest rate in combination with the Euler discretization of the log asset price using a central discretization for the integrated interest rate process is more accurate than the standard Euler discretization scheme using smaller time steps.

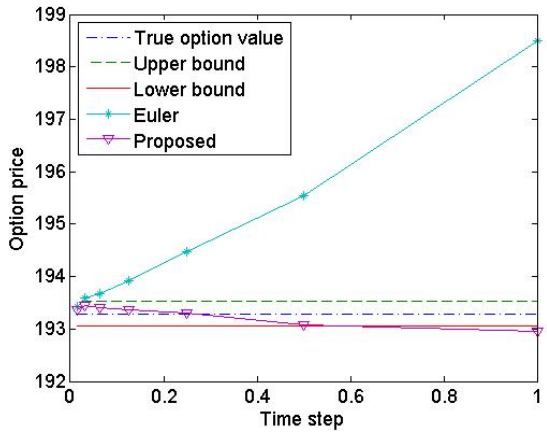
7.4 Valuation of an embedded insurance options using the extended Heston model

Next, we value the profit sharing embedded option using a representative set of parameters. We compare the proposed simulation scheme of the interest rate process, which is the second algorithm in Chapter 4.2.1, with the standard Euler scheme, because that is our starting point. As reference values we use the Euler scheme using a very fine grid, such that the discretization is extremely small, and $100,000$ simulations in order to reduce the statistical error of the MC estimator, where we use the antithetic sampling method as a variance reduction technique. Since there is still a statistical error involved we have computed confidence intervals of the 'assumed' true option value.

For comparison we vary Δt between $1/64$ and 1 and compute the option price using 50.000 simulations in combination with antithetic sampling. To make a fair comparison between the methods, we use the same random seed for both methods. In Figure 14 we show the true option price with corresponding confidence bounds, the option prices computed by the Euler and the proposed scheme as a function of Δt and of the computation time.

Figure 14: Comparison of the Euler discretization method and the proposed simulation method

(a) Comparison of accuracy



(b) Comparison of computation time

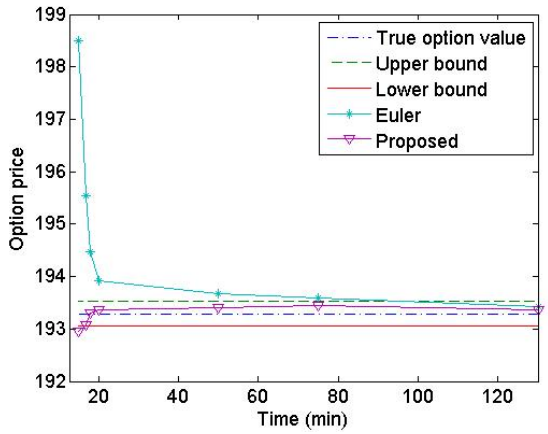


Figure 14 shows that the proposed simulation scheme is rather accurate for almost all values of Δt. We note that there is still some uncertainty around the approximated values of the proposed simulation and the Euler simulation method due to the fact we used $50,000$ simulations. We haven't computed confidence intervals around these value because otherwise the pictures would be unclear. We again note that we have used the same random seed such that comparisons are done in a fair way.

When measuring the computation time for both methods to obtain an option price, we conclude that approximately the same work is needed. As a consequence, using the figure above on the right hand side we conclude that the same accuracy of the Euler scheme using 64 time steps is obtained when using the proposed simulation scheme with only 2 or 4 time steps. This is a computation time reduction of approximately a factor 8, which is significant.

7.5 Conclusions

Our conclusions are the following:

- The proposed adapted NCI scheme is much more efficient than the QE scheme in valuing a call option. For example, approximately the same accuracy is obtained using the QE scheme with a time step $\Delta t = 1/8$ and the adapted NCI scheme with $\Delta t = 1/12$, however the adapted NCI scheme is still approximately a factor 5 faster than the QE scheme, which is significant[27];

- Using the antithetic sampling method and the moment matching technique, the variance of the MC estimator is reduced significantly such that more reliable estimates can be obtained;

- The proposed simulation scheme of the interest rate, which is the second algorithm discussed in Chapter 4.2.1, and the QE scheme are more accurate than the Euler discretization scheme to approximate the analytic statistics for a given time step. This implies that computation time can be saved by using a larger time step;

[27]We again note that more fair comparisons are done in, for example, C++.

- Valuing the profit sharing embedded option using the proposed simulation scheme of the interest rate process is done rather accurately compared to the Euler discretization. The computation time is reduced by approximately a factor 8.

8 Conclusions

8.1 Summary and conclusion

In this thesis the simulation of the extended Heston model and the efficient valuation of embedded options is discussed. We started with a general description of the field and the objectives of this thesis were discussed. The thesis is especially interesting for companies who deal with the risk-neutral valuation of embedded options, which are typically insurance companies. Embedded options are option type of products embedded in the liabilities of an (life) insurer.

Our starting point in this thesis is the two-factor Hull-White Black-Scholes risk-neutral model, where the variance process is deterministic. We have extended this model by modeling the variance by a so-called Cox-Ingersoll-Ross process. We have chosen this particular model because it is of importance in the insurance industry. We considered the description of the risk-neutral model and distributional properties of the model are discussed because these are of high importance as output. Several analytic expressions are derived so that the implementation of these statistics can be done efficiently, i.e. no numerical techniques have to be used to approximate the statistics.

Using the distributional properties discussed in Chapter 3, an efficient simulation scheme for the interest rate, the variance and the asset price process is constructed in Chapter 4. Since the interest rate process is normally distributed, we opt for simulating directly from the normal distribution. Because of the fact that the generation of normally distributed random numbers can be done rather fast, the generation of the interest rate can be done in a fast way. Comparing the Euler scheme with the true dynamics of the interest rate, we noticed that the Euler scheme needs a relative small time step in order to approximate the true underlying dynamics. Additionally, approximately the same computation time is needed for the Euler scheme and the proposed simulation scheme for a given time step, which implies that computation time is saved significantly. So, the exact simulation of the interest rate should be the default choice.

Recently, many papers emerged about the discretization of the Heston model. Several algorithms are proposed to simulate the Heston model and discussed briefly in Chapter 4. We have derived the adapted NCI scheme to simulate the variance process, which is especially efficient in the special case $d > 1$, where $d = \frac{4\kappa\xi}{\sigma_\nu^2}$. This proposed discretization scheme is based on analysis of representations of the Cox-Ingersoll-Ross process. By using a non-equidistant grid (which can be prestored once and for all) in combination with a linear interpolation method, the generation of the variance process can be done rather fast. Valuing a call option using the Heston model shows that the accuracy and computation time are preferable compared to the QE scheme of Andersen (2007), which is considered as the most efficient simulation scheme.

We derived a discretization scheme to simulate the asset price, when it is modeled by the extended Heston model. This discretization scheme is also based on analysis of the underlying distribution. Since we are interested in a discretization scheme for the asset price, we use a drift interpolation for approximating the integrated interest rate and the integrated variance process. Although its simplicity, the drift interpolation method is sufficiently accurate for a relative small time step. A martingale correction for the discretized asset price is rather hard to derive due to the correlated random variables. However, one could still use existing martingale corrections, based on the simulation scheme of the variance process, to correct the discretized asset price. By using the empirical martingale simulation method the martingale condition is fulfilled exactly (without any statistical error). Due to the limitations of Matlab, we have not

performed numerical test on our proposed simulation scheme of the asset price. Although the derivation of the scheme is theoretically appealing, numerical tests, preferable performed in C++, on the model are left for future research.

Since the convergence rate of the Monte Carlo estimator is relatively low, improvements are preferable to obtain more accurate results when pricing (embedded) options. Therefore, we discussed variance reduction techniques in order to reduce the variance of the Monte Carlo estimator. Since we are especially interested in generally applicable variance reduction techniques, because of the complex structure of the pay-off function of embedded options, we mainly focused on the antithetic sampling method and the moment matching technique. We have shown by using a simple example that the antithetic sampling method and the moment matching technique reduce the variance of the MC estimator significantly. This results in more reliable estimates.

Although the control variate approach is rather effective in reducing the variance of the MC estimator, it is not generally applicable. We have, however, constructed a control variate when one is interested in valuing plain vanilla options, which is typically the case in the calibration procedure. We have shown that this control variate provides significant variance reductions which implies that computation time can be saved. Because of the fact that the calibration of the extended Heston model is (still) a problem, because the characteristic function of the pay-off function cannot be obtained in closed form, the control variate approach could serve as an alternative.

Last but not least, after performing several numerical experiments, we have shown that the valuation of the profit sharing embedded options can be done much more efficient, by using the proposed simulation of the interest rate process instead of the Euler discretization scheme. Computation time is saved by approximately a factor 8.

8.2 Future research

Obviously, there are still many ways for improvements. In the following we list some topics for future research:

- Test the simulation scheme of the asset price, modeled by the extended Heston model by performing numerical tests;
- Investigate the application of the proposed scheme of the asset price in multi-asset applications that involve several correlated processes;
- Investigate the effect of including jumps to the asset price and possibly also to the interest rate process on valuing embedded options using simulation methods.

References

1. Abramowitz, M. and Stegun, I.A. (1972), *"Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables"*, National Bureau of Standards, Washington D.C.

2. Andersen, L. (2007), "Simple and Efficient Simulation of the Heston Stochastic Volatility Model", Working Paper, Bank of America Securities, New York.

3. Ahrens, J.H. and Dieter, U. (1982), "Computer generation of poisson deviates from modified normal distributions", *ACM Transactions on Mathematical Software*, **8**(2), 163179.

4. Bakker, B. (2006), "Calibration of the Hull-White Model", Master thesis, Technical University of Delft.

5. Beasley and Springer (1977), Algorithm AS 111, "The Percentage Points of the Normal Distribution", *Applied Statistics*, **26**, 118-121.

6. Björk, T. (2004), *"Arbitrage Theory in Continuous Time"*, Oxford University Press, Oxford.

7. Brigo, D. and Mercurio, F. (2006), *"Interest Rate Models Theory and Practice, with Smile, Inflation and Credit"*, Springer-Verslag, Heidelberg.

8. Broadie M. and Kaya, Ö. (2006), "Exact Simulation of Stochastic Volatility and other Affine Jump Processes", *Operations Research*, **54**(2), 217-231.

9. Cairns, A.J.G., (2004), *"Interest Rate Models"*, Princeton University Press, Princeton.

10. Capriotti, L, (2007), "Least Squares Importance Sampling for Monte Carlo Security Pricing", Available at SSRN: http://ssrn.com/abstract=975134.

11. Dominici, D. (2003), "Nested Derivatives: a Simple Method for Computing Series Expansions", *International Journal of Mathematics and Mathematical Sciences*, (58), 3699-3715.

12. Drăgulescu, A.A. and Yakovenko, V.M. (2002), "Probability distribution of returns in the Heston model with stochastic volatility", *Quantitative Finance*, **2**, 443-453.

13. Duan, J. C. and Simonato, J. G. (1998), "Empirical Martingale Simulation for Asset Prices", *Management Science*, **44**(9).

14. L'Ecuyer, P. and Panneton, F. (2007), "F_2-Linear Random Number Generators", GERAD Report 2007-21, to appear with minor revisions in *Advancing the Frontiers of Simulation*: A Festschrift in Honor of George S. Fishman.

15. Fang, F. and Oosterlee, C. W. (2008), "A Novel Pricing Method for European options Based on Fourier-Cosine Series Expansion", *SIAM Journal of Scientific Computing*, **31**(2), 826-848.

16. Glasserman, P. (2003), *"Monte Carlo Methods in Financial Engineering"*, Springer-Verlag, New York.

17. Glasserman, P. and Kim, K. (2008), "Gamma Expansion of the Heston Stochastic Volatility Model", Available at SSRN: http://ssrn.com/abstract=1279850.

18. Grzelak,L.A. and Oosterlee, C.W., van Weeren, S. (2008), "Extension of Stochastic Volatility Models with Hull-White Interest Rate Process", Report 08-04, Delft University of Technology.

19. Heston, S.L. (1993), "A closed-form solution for options with stochastic volatility with applications to bond and currency options", *Review of Financial Studies*, **13**, 585-625.

20. Hoevenaars, R.P.M.M. (2008), *"Strategic Asset Allocation Asset Liability Management"*.

21. Jäckel, P. (2002), *"Monte Carlo Methods in Finance"*, John Wiley and Sons, West Sussex.

22. Kahl, C. and Jäckel, P. (2006), "Fast strong approximation Monte-Carlo schemes for stochastic volatility models", *Quantitative Finance*, **6**, 513-536.

23. Kloeden, P. E., and E. Platen. (1999), *"Numerical Solution of Stochastic Differential Equations"*, Springer-Verlag, Berlin.

24. Kloeden, P. E., E. Platen. and Schurz, H. (1994), *"Numerical Solution of SDE Through Computer Experiments"*, Springer-Verlag, Berlin.

25. Knuth, D.E. (1981), "The Art of Computer Programming", **2**, *Seminumerical Algorithms*, Addison-Wesley, Reading, Mass.

26. Kocken, T. (2006), "Curious Contracts. Pension Fund Redesign for the Future", Proefschrift Vrije Universiteit, Tutein Nolthenius.

27. Lord, R, Koekkoek, R. and Van Dijk, D. (2008), "A comparison of biased simulation schemes for stochastic volatility models", Discussion Paper TI 2006-046/4, Tinbergen Institute, Erasmus University, Rotterdam, Netherlands.

28. Matsumoto, M. and Saito, M. (2006), "SIMD-oriented Fast Mersenne Twister: a 128-bit Pseudorandom Number Generator", *Monte Carlo and Quasi-Monte Carlo Methods 2006*, Springer, 607-622.

29. Matsumoto, M. and Nishimura, T. (1998), "Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator", *ACM Transactions on Modeling and Computer Simulation*, **8**(3),3-30.

30. Marsaglia, G. (2003), "Random number generators", *Journal of Modern Applied Statistical Methods*, **2**.

31. Marsaglia, G., (2005), "On the randomness of Pi and other decimal expansions", *Interstat*, **5**.

32. M. Matsumoto and T. Nishimura (1998), "Mersenne twister: a 623-dimensionally equidistributed uniform pseudorandom number generator", *ACM Transactions on Modeling and Computer Simulation*, **8**(1), 3-30 .

33. Mehta, N., Wu, J., Molisch, A. and Zhang, J. (2007), "Approximating a Sum of Random Variables with a Lognormal", *Wireless Communications, IEEE Transactions on*, **6**(7), 2690-2699.

34. Moro, B. (1995), "The full monte", *Risk*, **8**, 57-58.

35. Muskulus, M. (2007), "Three approaches to extend the Heston model", *"Proceedings of the Fifty-Eighth European Study Group Mathematics with Industry"*, Utrecht.

36. Nelson, C. R. and A. F. Siegel (1987). "Parsimonious modeling of yield curves", *The Journal of Business*, **60**(4), 473-489.

37. Nemes, G. (2007), "New asymptotic expansion for the $\Gamma(z)$ function", *Stans Library*, Volume II.

38. Ninomiya, S. and Victoir, N. (2004), "Weak approximation of stochastic differential equations and application to derivative pricing"', working paper, Tokyo Institute of Technology and Oxford University.

39. Plat, R., Pelsser, A. (2008), "Analytical Approximations for Prices of Swap Rate Dependent Embedded Options in Insurance Products", Discussion Paper 2008-008, Netspar.

40. Press,W.H., Flannery, B.P.,Teukolski, S.A. and Vetterling, W.T. (1988), *"Numerical Recipes in C"*, Cambridge University Press, Cambridge.

41. Shreve, S. (2006), "Stochastic Calculus for Finance II: Continuous-Time Models", Springer-Verslag, New-York.

42. Seydel, R (2003), *"Tools for Computational Finance"*, Springer-Verslag, Berlin.

43. Smith, R.D. (2008), "An almost exact simulation method for the heston model", *Journal of Cumputational Finance*, **11**(1),115125.

44. Van Bragt, D. and Steehouwer, H. (2007), *"Recent Trends in Asset and Liability Modeling for Life Insurers*", OCFR Methodological Paper No. 2007-01.

45. Van Haastrecht, A. and Pelsser, A.A.J. (2008), *"Efficient, almost exact simulation of the Heston stochastic volatility model"*, Available at SSRN: http://ssrn.com/abstract=1131137.

46. Zenios, S.A. and Ziemba, W. (2006), *"Handbook of Asset and Liability Management Volume I"*, Elsevier B.V., Amsterdam.

47. Zenios, S.A. and Ziemba, W. (2007), *"Handbook of Asset and Liability Management Volume II"*, Elsevier B.V., Amsterdam.

48. Zhu, J. (2008), "A Simple and Exact Simulation Approach to Heston Model", Available at SSRN: http://ssrn.com/abstract=1153950.

Appendix A

The dynamics for the asset price and inflation in the full risk-neutral model are given by

$$\begin{aligned}
dS(t) &= (r_N(t) - \delta)\, S(t)dt + \sqrt{\nu_S(t)}S(t)dW^S(t) && S(0) \geq 0 \\
dI(t) &= (r_N(t) - r_R(t))\, I(t)dt + \sqrt{\nu_I(t)}I(t)dW^I(t) && I(0) \geq 0 \\
d\nu_s(t) &= \kappa_S\left(\xi_S - \nu_S(t)\right)dt + \sigma_{\nu_S}\sqrt{\nu_S(t)}dW^{\nu_S}(t) && \nu_S(0) \geq 0 \\
d\nu_I(t) &= \kappa_I\left(\xi_I - \nu_I(t)\right)dt + \sigma_{\nu_I}\sqrt{\nu_I}(t)dW^{\nu_I}(t) && \nu_I(0) \geq 0 \\
dr_N(t) &= \left(\theta_N(t) + u_N(t) - a_N r_N(t)\right)dt + \sigma_{rN}dW^{rN}(t) && r_N(0) \geq 0 \\
du_N(t) &= -b_N u_N(t)dt + \sigma_{uN}dW^{uN}(t) && u_N(0) = 0 \\
dr_R(t) &= \left(\theta_R(t) - \rho_{I,rR}\sigma_{rR}\sigma_I + u_R(t) - a_R r_R(t)\right)dt + \sigma_{rR}dW^{rR}(t) && r_R(0) \geq 0 \\
du_R(t) &= -\left(\rho_{I,uR}\sigma_{uR}\sigma_I + b_R u_R(t)\right)dt + \sigma_{uR}dW^{uR}(t) && u_R(0) = 0,
\end{aligned}$$

where at time t, $S(t)$ denotes the asset price, $I(t)$ denotes the inflation, $r_N(t)$ denotes the nominal short interest rate, $u_N(t)$ denotes the second factor of the nominal short interest rate, $r_R(t)$ denotes the real short interest rate and $u_R(t)$ denotes the second factor of the real short interest rate, $\nu_S(t)$ denotes the variance process of the asset price and $\nu_I(t)$ denotes the variance of the inflation, respectively. Furthermore, a_R, a_N, b_R, b_N, κ_S and κ_I are mean reversion parameters of the real and nominal interest rates, the second factor of the real and nominal interest rates, the variance of the asset price and the variance of the inflation, respectively and ξ_S and ξ_I denotes the long term mean of the variance process of the asset price and the inflation process, respectively. The time-dependent parameters $\theta_N(t)$ and $\theta_R(t)$ fit the nominal and real short interest rate to the belonging inital term structure. Next, σ_{rN}, σ_{uN}, σ_{rR}, σ_{uR}, σ_{ν_S} and σ_{ν_I} are volatility parameters of the nominal short interest rate, the second factor of the nominal short interest rate, real short interest rate, the second factor of the real short interest rate, the variance process of the asset price and the variance of the inflation process, respectively. Finally, $\left(dW^S(t), dW^I(t), dW^{\nu_S}(t), dW^{\nu_I}(t), dW^{r_N}(t), dW^{u_N}(t), dW^{r_R}(t), dW^{u_R}(t)\right)$ represent 8 correlated Wiener processes with correlation $\rho_{i,j}$, where $i, j \in \{S, I, \nu_S, \nu_I, r_N, u_N, r_R, u_R\}$ and $i = j$ implies $\rho_{i,j} = 1$. The correlation matrix is a symmetric matrix and it is positive semi-definite matrix, i.e. for all $\mathbf{x} \in \mathbb{R}^8$ the correlation matrix Σ has to satisfy $\mathbf{x}^{tr}\Sigma\mathbf{x} \geq 0$.

Appendix B

Generation of pseudo-random numbers

To perform a MC simulation for valuing options based on the extended Heston model, uniformly distributed random numbers have to be generated and transformed to other distributions, of which the normal distribution is the most important. To obtain accurate results of the MC simulation one needs uniform random numbers with good uniform properties. So, the generation of 'high quality' uniform random numbers is of importance when performing a MC simulation. For the generation of uniform random numbers various generation methods are available which, however, differ in quality. Although in recent years the quasi-random numbers have gained more attention, we discuss in this chapter the generation of pseudo-random numbers. More information about quasi-random numbers can be found in Glasserman (2003).

According to Chapter 3 the asset price and interest rate are simulated using normally distributed random numbers and for the simulation of the variance process we need the non-central chi-squared distribution, which can be represented as a chi-squared random variable with stochastic degrees of freedom generated by a Poisson distribution and a combination of a standard normal random variable and a chi-squared random variable according to (17). To transform uniformly distributed random numbers to these particular distributions we need transformation methods. In the case of quasi-random numbers, a direct inversion of the cumulative distribution function is necessary, because otherwise the structure of the quasi-random numbers is unsatisfactory. Although these methods are slower, they represent a direct (accurate) approximation of the quantile function (inverse distribution function).

In this chapter we deal with the generation of normally distributed random numbers and by observing representation (17) for generating the variance process, we are also interested in the generation of chi-squared random variables, which are a special case of gamma random variables. In particular, we derive a generation method for the normal and the gamma distributions based on a careful analysis of the quantile functions. Because the direct generation of non-central chi-squared distributed random numbers is rather complex and time-consuming we discuss alternative generation methods for this distribution in Chapter 4 and Appendix E.

Generation of uniform random numbers

Many methods to generate uniform random numbers are available but they differ in quality. Several methods are discussed in Press and Flannery *et al.* (1994) and more recently in L'Ecuyer and Panneton (2007). Popular generation methods are, for example, the Mersenne Twister (MT) and RNG Streams. Although they are relatively slow, their uniform property is superior to other methods.

The MT generation method is commonly used in standard computation programs such as R and Matlab. It was developed by Matsumoto and Nishimura (1998) and it provides for fast generation and good quality of random numbers. An improvement of the standard MT method is the Single Input Multiple Data (SIMD) oriented Fast Mersenne Twister (STMF). This method is approximately twice as fast and provides for better uniform properties. For more information, one can consult Matsumoto and Saito (2006).

Another improvement of the MT method is the Multiply-With-Carry (MWC) method. The

main advantages of the MWC method are that a simple computer integer arithmetic and leads to very fast generation of sequences of random numbers. More information can be found in Marsaglia (2003) and Marsaglia (2005).

Transformation to standard normal random numbers

Below we have listed some transformation methods to obtain standard normally distributed random numbers. Since the normal distribution is a very popular distribution in practice, these algorithms are also very popular and well-known (see Glasserman (2003)). Using quasi-random numbers to obtain normally distributed random numbers one typically wants to use an approximation of the quantile function (inverse distribution function) that the uniform property of the quasi-random numbers preserves. Therefore, we discuss both approximations of the quantile function as well as other generation methods.

Box-Muller transformation

If U_1 and U_2 are standard uniform random numbers, then

$$Z_1 = \sqrt{-2\log(U_1)}\cos(2\pi U_2) \qquad Z_2 = \sqrt{-2\log(U_1)}\sin(2\pi U_2)$$

are two independent standard normal random numbers. This forms the basis of the Box-Muller transformation method.

Marsaglia polar method

The Marsaglia polar method is an optimization of the Box-Muller method and is based on an acceptance-rejection sampling algorithm. One generates U_1 and U_2, which are standard uniform random numbers, until $W = V_1^2 + V_2^2 < 1$, where $V_1 = 2U_1 - 1$ and $V_2 = 2U_2 - 1$. Then $Z_1 = V_1\sqrt{-\frac{2\log(W)}{W}}$ and $Z_2 = V_2\sqrt{-\frac{2\log(W)}{W}}$ are two independent standard normal random numbers. According to Seydel (2004) the probability that $W < 1$ is equal to $\frac{\pi}{4}$. So, in approximately $1 - \frac{\pi}{4}$ of all uniform drawings the (V_1, V_2)-tuple is rejected because $W \geq 1$. However, according to Seydel (2004) the Marsaglia polar method is more efficient than the Box-Muller method.

Moro's inversion method

Moro (1995) approximates the cumulative distribution function by a series representation. Moro divides the interval in a central region plus the tails and models these intervals separately. The central region is approximated by the Beasley and Springer inversion method (see Beasley and Springer (1977)) and the tails are approximated by a truncated Chebyschev series. More information on the algorithm can be found in Moro (1995) and pseudo code can be found in Glasserman (2003).

The method of Moro is well suited for transforming quasi-random numbers to normally distributed random numbers, because it is a direct approximation of the inverse distribution and, thus, the structure of the quasi-random input is preserved.

Acklams method

The method of Acklam is, just as the inversion method of Moro, a direct approximation of the quantile function. Although the method is extremely accurate, it is slower than the inversion

method of Moro. The method is based on an initial approximation which is improved by one step of Halley's root finding method, which is explained in the next chapter.

A known fact of the cumulative distribution function is that $F_{norm}^{-1}\left(p+\frac{1}{2}\right)$, for $p \in [0,1]$, is an odd function. The algorithm employs this fact to construct an accurate initial approximation. The algorithm uses two separate rational minimax approximations. One rational approximation is used for the central region and another is used for the tails. The central region is described by $0.02425 \geq p \geq 0.97575$. In this region a rational minimax approximation is used directly. In the tails, when $p < 0.02425$ or $p > 0.97575$, the argument is first passed through a non-linear transformation, to linearize the function, before a rational minimax approximation is applied.

Taylor reversion method

For the generation of standard normal random numbers we explain another generation method. This method is based on a Taylor reversion expansion. When we construct a Taylor series expansion of the cumulative distribution function, this expansion has a unique inverse which is also represented by a series expansion. This inverse series can be used to transform uniform random numbers to normally distributed random numbers. Note that this approach can also be applied to transform quasi-random numbers to standard normally distributed random numbers.

The cumulative distribution of the standard normal distribution is given by[28]

$$F_{norm}(x) = \frac{1}{2}\left(1 + Erf\left(\frac{x}{\sqrt{2}}\right)\right),$$

where $Erf(\cdot)$ denotes the error function. To solve $F_{norm}(x) = U$ for x, where U is a (standard) uniform random number, we get $x = \sqrt{2}E\left(2U-1\right)$, where E denotes the inverse error function. Thus, we need the inverse of the error function to obtain a value of the inverse cumulative distribution function.

In Dominici (2003) a series representation of the inverse error function around zero is given by the following:

$$E \approx \sum_{n\geq 0} A_n \left(\frac{\sqrt{\pi}}{2}\right)^{2n+1} \frac{x^{2n+1}}{(2n+1)!},$$

where $A_0 = 1$, $A_1 = 2$, $A_2 = 28$, $A_3 = 1016$, $A_4 = 69904$, $A_5 = 7796768$, $A_6 = 1282366912$, $A_7 = 291885678464, \ldots$. Such an approximation can also be constructed around other values than zero, however in that case the error function has to be evaluated which requires more computation time.

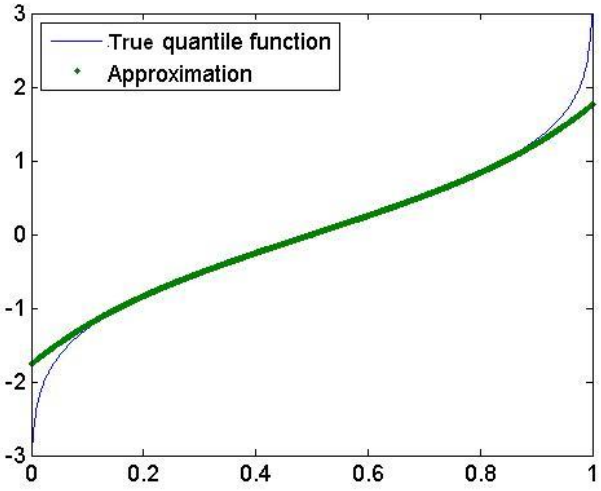
The inverse distribution function is now approximated by

$$F_{norm}^{-1}(x) \approx \sqrt{2}E(2x-1) \qquad (x \in [0,1]).$$

This approximation accounts for the center part of the total distribution. In Figure 15 we show the initial approximation with 5 terms.

[28] Remark that for generating a sample of a normal distribution with arbitrary mean and variance, μ and σ, we only need to sample from the standard normal distribution because, $N(\mu,\sigma^2) = \mu + \sigma N(0,1)$.

Figure 15: True quantile function versus approximation



We observe that the tails of the distribution are difficult to approximate by this approach.

The approximation is improved by one step of a root finding algorithm, such as Newton's or Halley's method. The Newton and Halley methods are second and third order, respectively, root finding procedures. Higher order accuracy can be obtained by a higher order root finding algorithm using Householders root finding algorithm. The Householders inversion solves x from $F(x) - U = 0$ iteratively by performing the following iterations

$$x_{k+1} = x_k + n \left. \frac{\frac{\partial^{n-1}}{\partial x^{n-1}} \left(\frac{1}{F(x)} \right)}{\frac{\partial^n}{\partial x^n} \left(\frac{1}{F(x)} \right)} \right|_{x=x_k} \qquad (k = 0, 1, \ldots), \tag{43}$$

where $F(x) = F_{norm}(x) - U$; x_0 is the initial approximation using the above Taylor reversion method and the convergence rate is $n + 1$.

Using the first step of Halley's root finding method, we have the following iteration

$$x_{k+1} = x_k - \frac{F_{norm}(x_k)}{F'_{norm}(x_k)} \left(1 - \frac{1}{2} \frac{F_{norm}(x_k) F''_{norm}(x_k)}{F'_{norm}(x_k)^2} \right)^{-1} = x_k - \frac{F_{norm}(x_k)}{F'_{norm}(x_k)} \left(1 + \frac{x}{2} \frac{F_{norm}(x_k)}{F'_{norm}(x_k)} \right)^{-1},$$

with $(k = 0, 1, \ldots)$ and where we have used

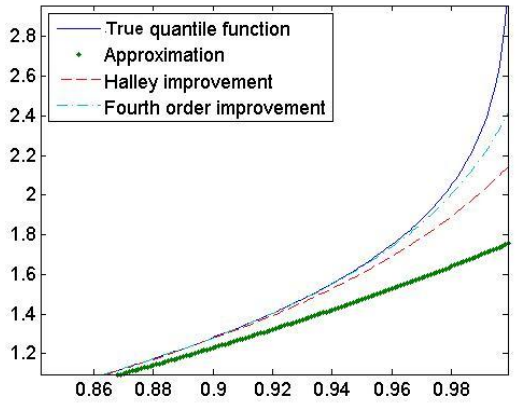
$$F'_{norm}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad F''_{norm}(x) = -\frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = -x f'_{norm}(x).$$

For implementation, one can best first evaluate $z = \frac{F_{norm}(x_k)}{F'_{norm}(x_k)}$ and then evaluate $x_k = x_k - z \left(1 + \frac{x_k}{2} z \right)^{-1}$.

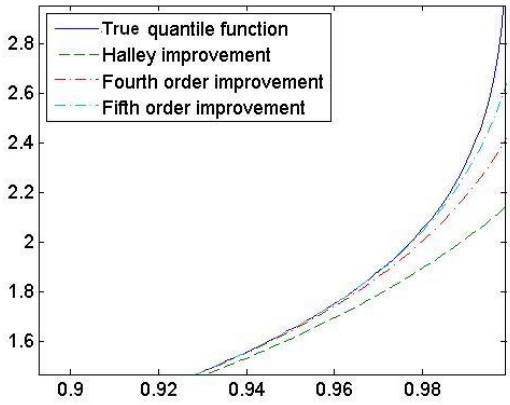
Using this iteration we get the picture in Figure 16:

Figure 16: True quantile function versus approximation

(a) Right tail of quantile function



(b) Right tail of quantile function



This figure shows that the accuracy is drastically improved by using higher order root finding algorithms.

Transformation to gamma random numbers

Acceptance-rejection algorithm

For the generation of gamma random numbers, the application of an acceptance-rejection algorithm can be used. Marsaglia and Tsang (2000) propose a fast and simple acceptance-rejection method. More information can be found in Press and Flannery *et al.* and in Glasserman (2003).

Taylor reversion method

To avoid the acceptance-rejection method to generate gamma random numbers, we consider again the Taylor reversion expansion method. The distribution function of the gamma distribution is given by

$$F_{gam}(x;k_1,k_2) = \frac{\Gamma\left(k_1,0,\frac{x}{k_2}\right)}{\Gamma(k_1)} \qquad (k_1>0,k_2>0),$$

where $\Gamma(k_1,0,x)$ denotes the incomplete gamma function and $\Gamma(x)$ the standard gamma function. Solving x from $F_{gam}(x)=U$, where U is uniformly distributed, we get

$$x = \Gamma^{-1}(k_1,0,U\,\Gamma(k_1))k_2. \tag{44}$$

Thus, in order to obtain a value of the inverse distribution function we need an approximation of the inverse of the incomplete gamma function.

Using Dominici (2003) we can construct a series expansion of the inverse of the incomplete gamma function around an appropriate value.

Result 8.1. *The series expansion for the inverse incomplete gamma function is constructed around* $c(k_1)=\Gamma(k_1,0,1)$ *and is given by*

$$\Gamma^{-1}(k_1,0,x) \approx 1+\sum_{n\geq 0} e^n Q_{n-1}\frac{(x-c(k_1))^n}{n!},$$

where $Q_1 = 2 - k_1$, $Q_2 = 7 - 7k_1 + 2k_1^2$, $Q_3 = 36 - 53k_1 + 29k_1^2 - 6k_1^3$, $Q_4 = 245 - 474k_1 + 375k_1^2 - 146k_1^3 + 24k_1^4$, $Q_5 = 2076 - 4967k_1 + 5104k_1^2 - 2847k_1^3 + 874k_1^4 - 120k_1^5$. *To obtain a value of the inverse distribution function we evaluate (44) where we use the above approximation to evaluate the inverse incomplete gamma function.*

This initial approximation can be improved by one step of Halley's method. Using a first step of Halley's root finding method, we get the following iteration

$$\begin{aligned} x_{k+1} &= x_k - \frac{F_{gam}(x_k)}{F'_{gam}(x_k)} \left(1 - \frac{1}{2} \frac{F_{gam}(x_k) F''_{gam}(x_k)}{F'_{gam}(x_k)^2}\right)^{-1} \\ &= = x_k - \frac{F_{gam}(x_k)}{F'_{gam}(x_k)} \left(1 - \frac{\left(\frac{k_1 - 1}{x} - \frac{1}{k_2}\right)}{2} \frac{F_{gam}(x_k)}{F'_{gam}(x_k)^2}\right)^{-1}, \end{aligned}$$

with $(k = 0, 1, \ldots)$ and where we have used

$$F'_{gam}(x) = \frac{\left(\frac{x}{k_2}\right)^{k_1} e^{-\frac{x}{k_2}}}{x\,\Gamma(k_1)} \qquad F''_{gam}(x) = \left(\frac{k_1 - 1}{x} - \frac{1}{k_2}\right) F'_{gam}(x).$$

In Figure 17 we show the approximation of the quantile function using $k_1 = 1.5$ and $k_2 = 2$:

Figure 17: True quantile function versus approximation

(a) Complete quantile function

(b) Right tail of quantile function

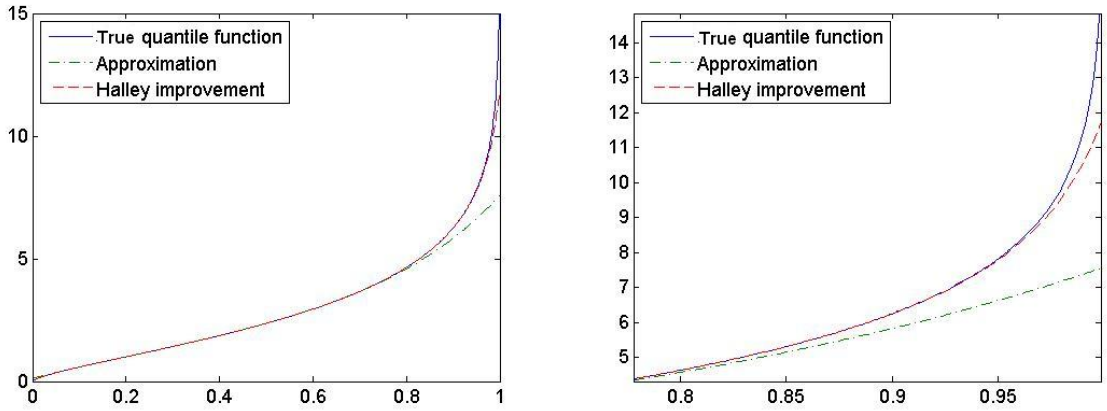


Figure 17 shows that one step of Halley's method improves the initial approximation significantly. However, for certain parameters settings, the root finding procedure can break down due to a division by zero. To overcome this one could set zero values to a (very) small value and then apply the root finding algorithm.

Conclusion

We have discussed the generation of uniformly distributed random numbers and the transformation to other distributed random numbers. To obtain accurate and reliable results for a MC simulation one needs a high quality uniform random number generator. The MT generation method can be improved by other generation methods which are faster and have better uniformly properties. According to the literature, the STMF generation method is an improvement

of the MT method. It is faster than the MT method and provides for better uniform properties.

To transform uniform random numbers to the normal distribution, the Marsaglia polar method is according to Seydel (2004) very efficient. This method, however, is not applicable when one wishes to use quasi-random numbers since it is not a direct approximation of the quantile function. In the latter case it is best to use the Moro inversion method, which is according to Glasserman (2003) fast and accurate. The Taylor reversion method we explained is not as accurate as the inversion of Moro, additionally, the method is time-consuming due to the application of root finding algorithms. To generate gamma random numbers it is best to use the acceptance-rejection method if one wants a fast generation, because the Taylor reversion method is too time-consuming.

Appendix C

In case of the one-factor HW interest rate model we have that, conditional on $\mathcal{F}_s$, with $(0 \leq s < t \leq T)$, the integrated interest rate process, $\int_s^t r(u)du$, is normally distributed with mean and variance given by

$$\begin{aligned}
\mathbb{E}\left[\int_s^t r(u)du \,|\, \mathcal{F}_s\right] &= B(s,t)\left(r(s) - \alpha(s)\right) + \log\left(\frac{P(0,s)}{P(0,t)}\right) + \frac{1}{2}\left(V(0,t) - V(0,s)\right), \\
Var\left(\int_s^t r(u)du \,|\, \mathcal{F}_s\right) &= V(s,t),
\end{aligned}$$

where $B(s,t) = \frac{1}{a}\left(1 - e^{-a(t-s)}\right)$, $V(s,t) = \frac{\sigma^2}{a^2}\left(t - s + \frac{2}{a}e^{-a(t-s)} - \frac{1}{2a}e^{-2a(t-s)} - \frac{3}{2a}\right)$ and $P\left(0,t\right)$ denotes the initial value of a zero coupon bond with maturity t. For the derivation of these moments we refer to Brigo and Mercurio (2006, Chapter 3).

When the interest rate is modeled by a two-factor HW model the distributional properties of the integrated interest rate are determined by using the two-factor Gaussian model. Conditional on $\mathcal{F}_s$, with $(0 \leq s < t \leq T)$, the integrated interest rate process, $\int_s^t r(u)du = \int_s^t x(u) + y(u)du + \int_s^t \varphi(u)$, is normally distributed with mean and variance given by (for the derivation of the moments we refer to Brigo and Mercurio (2006, Chapter 4))

$$\begin{aligned}
\mathbb{E}\left[\int_s^t r(u)du \,|\, \mathcal{F}_s\right] &= \frac{1}{b_x}\left(1 - e^{-b_x(t-s)}\right)x(s) + \frac{1}{b_y}\left(1 - e^{-b_y(t-s)}\right)y(s) + \int_s^t \varphi(u)du, \\
Var\left(\int_s^t r(u)du \,|\, \mathcal{F}_s\right) &= V(s,t),
\end{aligned}$$

where

$$\begin{aligned}
V(s,t) &= \frac{\sigma_x^2}{b_x^2}\left(t - s + \frac{2}{b_x}e^{-b_x(t-s)} - \frac{1}{2b_x}e^{-2b_x(t-s)} - \frac{3}{2b_x}\right) \\
&+ \frac{\sigma_y^2}{b_y^2}\left(t - s + \frac{2}{b_y}e^{-b_y(t-s)} - \frac{1}{2b_y}e^{-2b_y(t-s)} - \frac{3}{2b_y}\right) \\
&+ 2\rho_{x,y}\frac{\sigma_x\sigma_y}{b_x b_y}\left(t - s - \frac{1}{b_x}\left(1 - e^{-b_x(t-s)}\right) - \frac{1}{b_y}\left(1 - e^{-b_y(t-s)}\right)\right. \\
&+ \left.\frac{1}{b_x + b_y}\left(1 - e^{-(b_x+b_y)(t-s)}\right)\right).
\end{aligned}$$

Next, we want to derive an expression of $\int_s^t \varphi(u)du = \int_s^t f(0,u) + \phi(0,u)du$. Since the annually compounded zero coupon curve is often modeled by a spline function, the integral $\int_s^t f(0,u)du$ needs to be approximated using numerical integration techniques such as the trapezoidal method. Furthermore, $\int_s^t \phi(0,u)du$ depends on the kind of factor model is used. Nevertheless, it can be derived analytically. For example for the one-factor Hull-White model, we have

$$\int_s^t \phi(0,u)du = \frac{\left(2a(t-s) + e^{-2as} - e^{-2at} + 4e^{-at} - 4e^{-as}\right)\,{\sigma_r}^2}{4\,a^3}.$$

Expressions of $\int_s^t \phi(0,u)du$ for the two-factor HW model can easily be derived by using, for example, Mathematica.

Appendix D

Approximation of modified Bessel function of the first kind

In the evaluation of the characteristic function, which is given by (23), of the integrated variance process a modified Bessel function of the first kind has to be evaluated. Since this is a time-consuming process, we propose an approximation of the modified Bessel function of the first kind which can be evaluated much faster. The modified Bessel function of the first kind can be represented by the following series

$$B(v,z) = \left(\frac{1}{2}z\right)^v \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k!\Gamma\left(v+k+1\right)} \tag{45}$$

Result 8.2. *This series representation contains a gamma function,* $\Gamma(z)$*. According to Nemes (2007) the gamma function function can be approximated by*

$$\Gamma(z) \approx \sqrt{\frac{2\pi}{z}}\left(\frac{z}{e}\left(1+\frac{10z}{120z^2-1}\right)\right)^z.$$

Next, we compare the accuracy of the series expansion for different values of a truncation parameter, N_{Bessel}. To make a comparison, we vary v and z between $[-0.5, 2]$ and $[0, 0.5]$ with step size 0.01, respectively, and measure the maximum relative error. The computation time needed to compute 12801 Bessel functions is approximately 1.43 seconds. We get the results in Table 8:

Table 8: Maximum error of recovering $B(v,z)$ using the derived approximation

N_{Bessel}	0	1	2	3	4
error (%)	10.84	$5.31 \cdot 10^{-1}$	$5.31 \cdot 10^{-1}$	$5.31 \cdot 10^{-1}$	$5.31 \cdot 10^{-1}$
time (s)	$2.23 \cdot 10^{-2}$	$4.56 \cdot 10^{-2}$	$6.72 \cdot 10^{-2}$	$9.89 \cdot 10^{-2}$	$1.27 \cdot 10^{-1}$

We conclude from Table 8 that the series representation of the modified Bessel function of the first kind with an approximation for the gamma function converges very fast. The resulting error of $\approx 5.31 \cdot 10^{-1}\%$ is due to the approximation of the gamma function. The speed-up is enormous, the first order approximation is a factor ≈ 64 faster than the exact evaluation.

Appendix E

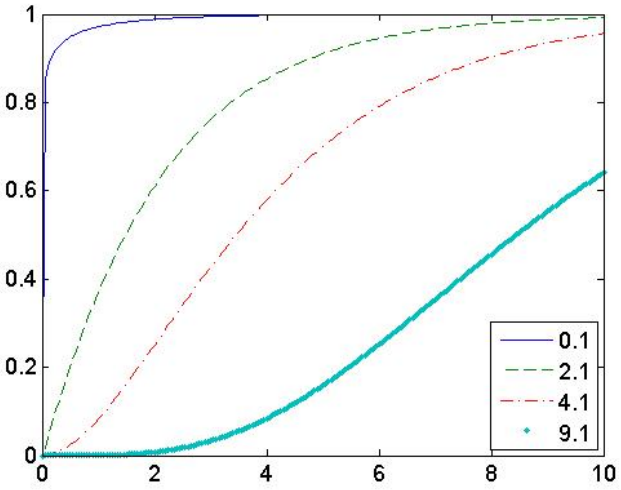
Alternative derived methods to generate the variance process

In this Appendix we explain new simulation methods for the variance process, which, however, do not meet the performance of existing simulation schemes.

Improving the NCI algorithm

A way to improve the NCI scheme is by using an interpolation technique to cache the chi-squared distributions. We do not use the whole grid, $\mathcal{N}$, but only two values, N_1 and N_2 and interpolate other Poisson values between these two values. We choose these values in such a way that $\mathbb{P}(N < N_1) < \epsilon$ and $\mathbb{P}(N > N_2) < \epsilon$, where ϵ is chosen to be sufficiently small. This can be accomplished by, for example, setting $N_1 = (1-c)\frac{1}{2}\lambda$ and $N_2 = (1+c)\frac{1}{2}\lambda$, where $\frac{1}{2}\lambda$ is the mean of the Poisson random variable and $c \in \mathbb{N}$ has to be chosen. We use the fact that $F_{chis}(x; d_1) \geq F_{chis}(x; d_2)$, with $0 < d_1 < d_2$, for every x, which indicates that the distribution functions are strictly monotone in the parameter. As an illustration, we present in Figure 18 four chi-squared distribution functions with different parameters:

Figure 18: Four chi-squared distribution functions with different parameters



To cache the inverse chi-squared distributions we prestore two inverse distribution functions, $F^{-1}_{chi}(U; d+2N_1)$ and $F^{-1}_{chi}(U; d+2N_2)$ conditional on N_1 and N_2, respectively, where $U \in \mathcal{U}$[29]. Based on a generated Poisson random number, N^P, and N_1 and N_2 we cache the inverse chi-squared distribution function, $F^{-1}_{chi}(U; d+2N^P)$, by means of interpolation using $F^{-1}_{chi}(U; d+2N_1)$ and $F^{-1}_{chi}(U; d+2N_2)$. If $N^P < N_1$ or $N^P > N_2$, where $i = 1, \ldots, N$, we use an extrapolation technique. The resulting algorithm to simulate the variance process using the second representation of (17), introduced here, then reads,

1. Precompute $N_1 = (1-c)\frac{1}{2}\lambda$ and $N_2 = (1+c)\frac{1}{2}\lambda$, based on a chosen value of c. The constant λ can be found in Chapter 3.7;

2. Precompute two inverse chi-squared distribution functions $F^{-1}_{chi}(U; d+2N_1)$ and $F^{-1}_{chi}(U; d+2N_2)$ conditional on N_1 and N_2, respectively, where $U \in \mathcal{U}$;

[29] The grid $\mathcal{U}$ is a discretization of the uniform interval.

3. Generate a Poisson random number, N^P, with mean $\frac{1}{2}\lambda$;

4. In case the Poisson random number is within $[N_1, N_2]$ compute either $D_1 = \frac{N^P - N_1}{N_2 - N_1}$ or $D_2 = \frac{N_2 - N^P}{N_2 - N_1}$. In case the Poisson random number is not in the interval $[N_1, N_2]$, then a linear (or a more advanced) extrapolation method can be used;

5. These weights can now be used to interpolate between the precomputed inverse distribution functions, $F_{chi}^{-1}(U; d+2N_1)$ and $F_{chi}^{-1}(U; d+2N_2)$, to obtain the approximated inverse chi-squared distribution function, $\widetilde{F}_{chi}^{-1}(U; d+2N^P)$, where $U \in \mathcal{U}$. More specifically, the approximated inverse distribution function, $\widetilde{F}_{chi}^{-1}(U; d+2N^P)$, belonging to the Poisson random number N^P is given by

$$\begin{aligned} \widetilde{F}_{chi}^{-1}(U; d+2N^P) &= \left(F_{chi}^{-1}(U; d+2N_2) - F_{chi}^{-1}(U; d+2N_1)\right) D_1 + F_{chi}^{-1}(U; d+2N_1) \text{ or} \\ \widetilde{F}_{chi}^{-1}(U; d+2N^P) &= F_{chi}^{-1}(U; d+2N_1) - \left(F_{chi}^{-1}(U; d+2N_2) - F_{chi}^{-1}(U; d+2N_1)\right) D_2, \end{aligned}$$

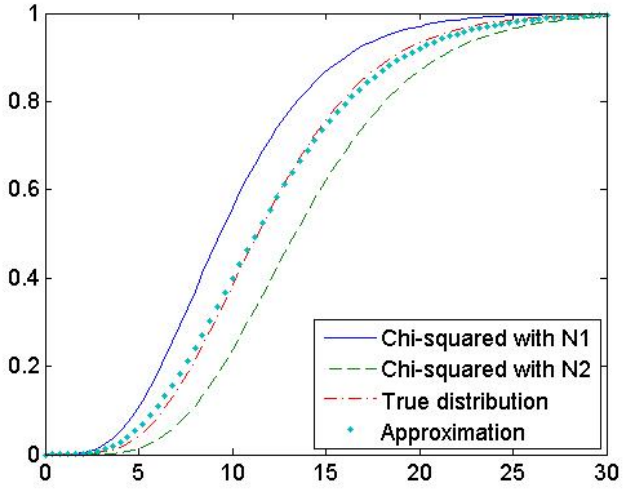
   where $U \in \mathcal{U}$.

6. Conditional on the generated Poisson random number, we can generate a sample, U_ν of the uniform distribution and use $\widetilde{F}_{chi}^{-1}(U_\nu; d+2N^P)$ to obtain a sample of the chi-squared distribution by means of interpolation;

7. A sample of $\widetilde{\nu}(t+\Delta t)$, conditional on $\widetilde{\nu}(t)$, is then generated by computing $C\,\widetilde{F}_{chi}^{-1}(U_\nu)$, where the constant C can be found in Chapter 3.7.

After performing some numerical experiments, we conclude that the proposed method is especially accurate when the parameter λ is relative large as can be seen in Figure 19, where we determine the distribution function of χ^2_{12} given χ^2_{10} and χ^2_{14} by means of interpolation using $N_1 = 10$, $N_2 = 14$, $d = 0$:

Figure 19: Approximation of chi-squared distribution by interpolation



As remarked by Haastrecht and Pelsser (2008), the parameter λ is often relatively low, which results in Poisson values concentrated around zero and, hence, precomputing the chi-squared distributions does note take much effort, because only a few inverse chi-squared distributions have to be precomputed (typically 2 or 3). Thus, the interpolation approach discussed above will only be accurate in unrealistic situations. We therefore do not proceed in this way.

Moment matched gamma distribution

According to Chapter 3 the stationary distribution of the CIR process is a gamma distribution. In other words, when the time horizon becomes large, the variance process tends to a gamma distribution with scale and shape parameters equal to $\frac{2\xi}{d}$ and $\frac{d}{2}$, respectively. The gamma distribution has just as the non-central chi-squared distribution, a positive support. We will investigate whether the gamma distribution, $G_{var}(x; v_1, v_2)$, is a good substitute for the non-central chi-squared distribution, because the generation of gamma random numbers can be done significantly faster than the generation of non-central chi-squared random numbers.

The parameters of the gamma distribution can be computed in several ways, for example, by

(i) Using a moment matching technique;

(ii) Fitting two values of the distribution, probability density, characteristic or moment generating function, so that the two parameters can be solved from the resulting two equations (in case of using the moment generating function, see Mehta and Wu *et al.* (2007) for an application);

In case of approach (ii) numerical tools have to be used to obtain the parameters v_1 and v_2, because of the highly non-linear shape of the distribution, probability density, characteristic or moment generating function. Thus, this approach is too time-consuming in general. However, the moment matching technique approach in (i), can be performed analytically. The mean and variance of the gamma distribution are $v_1 v_2$ and $v_1 v_2^2$, respectively, so by setting these equal to the mean and variance of the variance process, the parameters v_1 and v_2 can easily be obtained by solving two equations[30].

Result 8.3. *Using the mean and variance of the variance process given in (14) and (15), and solving*

$$v_1 v_2 = C\left(d + \lambda\right), \qquad v_1 v_2^2 = C^2\left(2d + 4\lambda\right)$$

results in the following shape and scale parameters of the gamma distribution:

$$v_1 = \frac{(d+\lambda)^2}{2(d+2\lambda)} \qquad v_2 = \frac{2C(d+2\lambda)}{d+\lambda}. \tag{46}$$

The gamma distribution function, which approximates the distribution function of the variance process is then given by $G_{var}(x; v_1, v_2)$ where v_1 and v_2 are given by (46).

However, using the fact that the variance is distributed as a constant times a non-central chi-squared distribution we can also match the moments of the corresponding non-central chi-squared distribution with the moments of a gamma distribution, $\widetilde{G}_{var}(x; \widetilde{v}_1, \widetilde{v}_2)$, and then multiply with the constant C[31]. Using this approach we can investigate the accuracy of the moment matched gamma distribution in terms of the parameters. Using the mean and variance of the non-central chi-squared distribution, which are $d + \lambda$ and $2d + 4\lambda$, respectively, gives,

$$\widetilde{v}_1 \widetilde{v}_2 = d + \lambda, \qquad \widetilde{v}_1 \widetilde{v}_2^2 = 2d + 4\lambda.$$

[30] One could also apply a moment matching technique to approximate the distribution function of the volatility given in Chapter 3. However, after performing some numerical experiments, it turns out that this approach is not accurate for low values of d. Additionally, the moments of the volatility model are complex, which implies that the moment matching technique is more time-consuming. Therefore we only discuss the moment matching technique for the variance process.

[31] Note that this approach results in the same distribution as the first approach.

This results in the following shape and scale parameters of the gamma distribution:

$$\widetilde{v}_1 = \frac{(d+\lambda)^2}{2(d+2\lambda)} \qquad \widetilde{v}_2 = 4 - \frac{2d}{d+\lambda}. \tag{47}$$

Then the variance process is approximated by a moment matched gamma distribution with parameters given by (47) times a constant C.

Result 8.4. *The characteristic functions of the non-central chi-squared distribution and the gamma distribution are given by*

$$\frac{e^{\frac{\lambda ti}{1-2ti}}}{(1-2ti)^{\frac{d}{2}}} \qquad \text{and} \qquad \frac{1}{(1-\widetilde{v}_2 ti)^{\widetilde{v}_1}} = \frac{1}{(1-(4-\frac{2d}{d+\lambda})ti)^{\frac{(d+\lambda)^2}{2(d+2\lambda)}}},$$

respectively. We note that when λ tends to zero, the approximating gamma distribution is exact, because then the characteristic functions are the same.

From parameter λ, we see that for high time horizons λ tends to zero. Thus, when the time horizon tends to infinity, the moment matched gamma distribution is exact. When the time horizon is relative small and the other parameters are such that λ is relatively small, the moment matched gamma distribution is also accurate.

The approximating gamma distribution is, however, not always sufficiently accurate. When the non-centrality parameter is relatively low or when the number of degrees of freedom is relatively high (independent of each other)[32], then the moment matched gamma distribution is accurate. When the fit is not good enough, we adjust the gamma distribution in the following way:

Result 8.5. *We use $\widetilde{G}\left(\left(\frac{x}{C}\right)^{1-\frac{1}{(a(x+b))^c}}; \widetilde{v}_1, \widetilde{v}_2\right)$, where $\widetilde{G}$ denotes a gamma distribution function with parameters $\widetilde{v}_1$ and $\widetilde{v}_2$; $a, b, c \in \mathbb{R}$ have to be determined so that the distribution function of the variance is approximated sufficiently well. As noted before, the right tail is approximated well by the gamma distribution, so $1 - \frac{1}{(a(x+b))^c}$ should converge to 1 for x increasing. On the other hand, when x is small, $1 - \frac{1}{(a(x+b))^c}$ should be below 1.*

For the cases I-IV we have worked out numerically the values for a, b and c:

- Case I ($\nu(0) = \xi, T = 1$): $a = 2.3$, $b = 1$ and $c = 2$;
- Case II ($\nu(0) = \xi, T = 1$): $a = 1.9$, $b = 1$ and $c = 2$;
- Case III ($\nu(0) = \xi, T = 1$): $a = 3.3$, $b = 1$ and $c = 2$;
- Case IV ($\nu(0) = \xi, T = 1$): no transformation is needed, use $G_{var}(x; v_1, v_2)$, where the parameters v_1 and v_2 are obtained using moment matching.

Observe that we fixed $b = 1$ and $c = 2$ so that we only have to vary a. Note also that there might be other values. Small perturbation in the parameter settings does not change the fit very much.

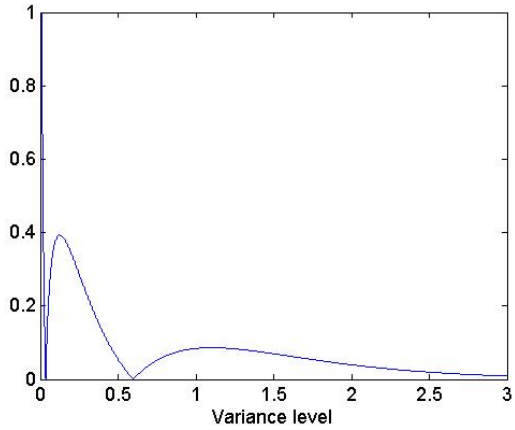
Remark. *Using the adapted moment matched gamma distribution we need to find the inverse of $\left(\frac{x}{C}\right)^{1-\frac{1}{(a(x+b))^c}}$ in order to obtain a random number of this particular distribution. This function does not have an analytic inverse and, for that reason, approximations have to be used or a direct inversion by using a root finding algorithm.*

Comparing the moment matched gamma distribution with the real distribution for the cases I-IV gives Figure 20, where we have measured the relative error (in %):

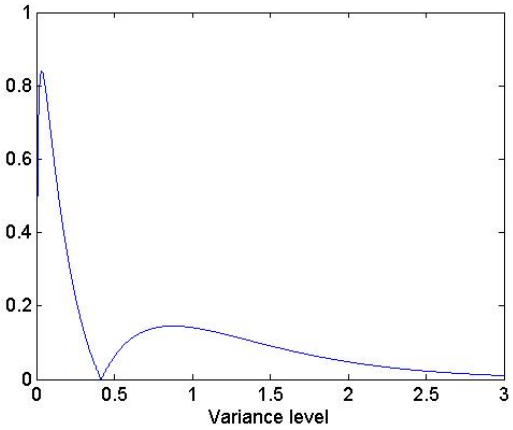
[32] After running some numerical experiments, we found that when d is large, the moment matched gamma distribution is accurate.

Figure 20: Relative error (%)

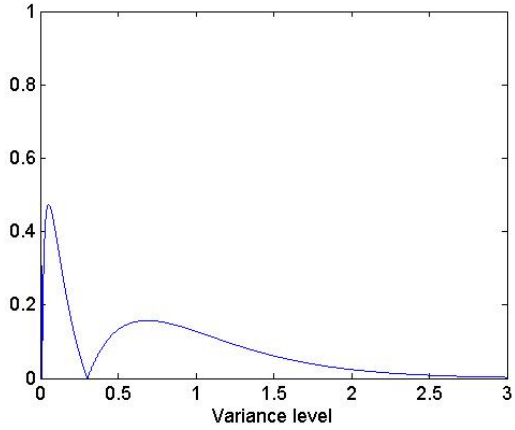
(a) Case I: $\nu(0) = \xi$ and $T = 1$



(b) Case II: $\nu(0) = \xi$ and $T = 1$



(c) Case III: $\nu(0) = \xi$ and $T = 1$



(d) Case IV: $\nu(0) = \xi$ and $T = 1$

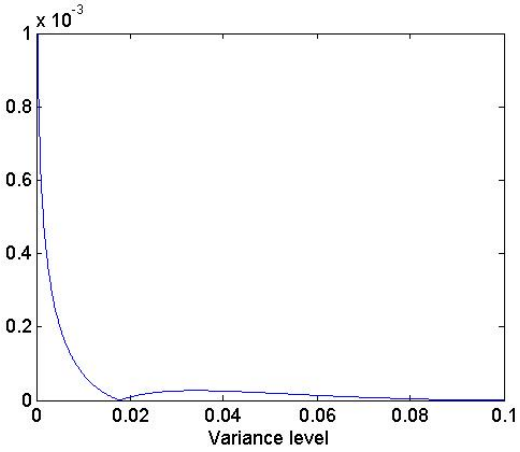


Figure 20 shows that the relative error is never larger than 1%, which implies that the distribution of the variance can be approximated very well.

Discussion Although the generation of gamma random numbers can be done faster than the generation of non-central chi-squared random numbers, it turns out, after performing several numerical experiments that this approach does not meet the performance of, for example, the QE scheme of Andersen (2007). This is mainly due to the fact that the generation of normally distributed random numbers, which are needed for the QE scheme, can be done significantly faster than the generation of gamma random numbers. Although, the generation of gamma distributed random numbers can be speeded up by using interpolation techniques, we do not proceed in this way, because the method is not sufficiently general.

Appendix F

Alternative derived simulation methods of the integrated variance process

Using an exact simulation scheme of the asset price process, the simulation of the integrated variance process is of high importance. In Chapter 4.1.1 we have discussed an exact generation method, which is constructed by Broady and Kaya (2006). Although Glasserman and Kim (2008) constructed an alternative simulation method of the integrated variance process, which is much more efficient, we do not doubt that alternative methods can be constructed which are even more efficient. Therefore, we discuss in this chapter some alternative new simulation methods of the integrated variance process. However, after performing some numerical experiments, we found that the alternative simulation methods, which are introduced here, do not meet the performance of a simple drift interpolation method in combination with a discretization method.

Exact simulation of the integrated variance rate process

We intended to obtain an analytical expression of the distribution function using complex integration theory, however, due to the presence of branch cuts it is, to our point of view nearly impossible to derive such an expression. Therefore we do not proceed in this direction.

To perform an exact simulation of the integrated variance process, one needs to use numerical tools in order to obtain a value of the inverse distribution function. Since the characteristic function is rather complex due to the presence of the modified Bessel function of the first kind, this process is time-consuming. As already explained, Broady and Kaya (2006) perform an exact simulation based on a numerical integration and inversion algorithm. To improve this approach a first step towards an improvement is to adjust the numerical integration scheme so that the number of iterations can be reduced drastically.

Fang and Oosterlee (2008) showed that their COS method is very efficient in obtaining the density function using a series expansion based on the characteristic function. Performing an additional integration step, one can obtain a value of the distribution function. The advantage of this approach is that the additional integration can be done analytically and, as a consequence, the exponential convergence is preserved. Whereas Fang and Oosterlee (2008) apply the COS method to price plain vanilla options, which has the extra feature that the pay-off function decays to zero very fast, in our context there is typically no (extra) factor in front of the integrand which rapidly decays to zero. As a result, the COS method is not very efficient in obtaining a value of the distribution function. After performing some numerical experiments using different kinds of numerical integration methods, it turns out that the trapezoidal method is preferred.

To obtain accurate results, the number of iterations in the numerical integration algorithm, N_{int}, should be chosen relatively large (especially when the function is oscillatory). To reduce N_{int}, we derive an alternative method. We wish to construct an approximation of the characteristic function (23), with the additional property that it has an analytic expression of the distribution function. This is a so-called control variate approach.

In other words,

$$F(x) - G(x) = \frac{2}{\pi} \int_0^\infty \frac{\sin(x\omega)}{\omega} \left(\Re\left(\Psi(\omega)\right) - \Re\left(\widetilde{\Psi}(\omega)\right) \right) d\omega, \tag{48}$$

where $\widetilde{\Psi}(t)$ is the approximated characteristic function with corresponding distribution $G(x)$. If a good control variate can be constructed, the numerical integration can be done much more effective, i.e. N_{int} can be reduced.

We propose the gamma distribution as a control variate. When we consider the gamma distribution with shape and scale parameters k_1 and k_2, respectively, the characteristic function is given by

$$\Psi_{gam}(x) = \frac{1}{(1 - ixk_2)^{k_1}}.$$

Since we have an exact representation of the characteristic function of the integrated variance process (23), we can directly compare the characteristic function of the gamma distribution with given in (23). We obtain the parameters of the gamma distribution by a moment matching technique so that the first two moments of the gamma distribution coincide with the first two moments of the true distribution, i.e.

$$k_1 = \frac{\mu_{Iv}^2}{\sigma_{Iv}^2} \qquad k_2 = \frac{\sigma_{Iv}^2}{\mu_{Iv}}.$$

The moments of the true distribution are derived by computing

$$\mathbb{E}\left[Iv(s,t)^n\right] = i^{-n} \left(\frac{\partial^n \Psi_{Iv}(x)}{\partial x^n} \right)_{x=0},$$

using, for example, Mathematica. To illustrate the effectiveness of the control variate, we show Figure 21, where we have plotted the real part of the characteristic function on an appropriate interval.

Figure 21: Moment matched gamma characteristic function versus true characteristic function

(a) Case III, $\nu(0) = 0.09$, $\nu(1) = 0.05$ (b) Case IV, $\nu(0) = 0.02$, $\nu(3) = 0.01$

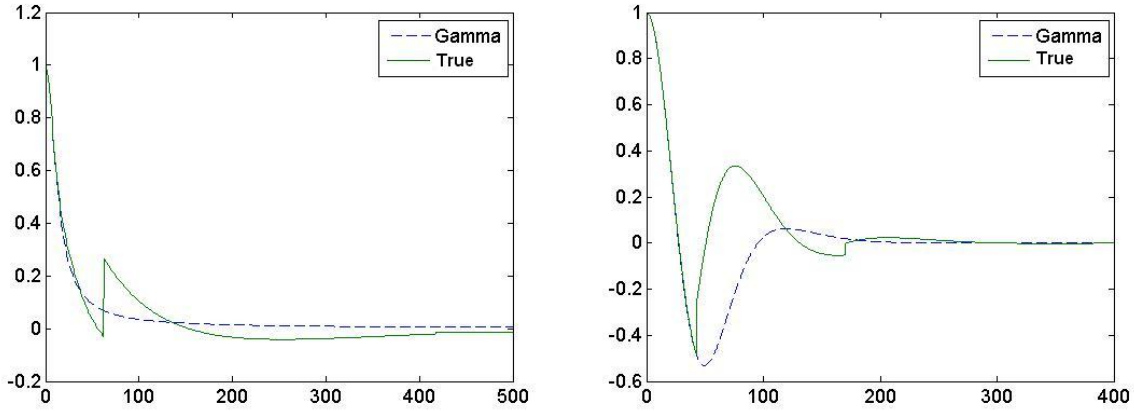


Figure 21 shows that the effectiveness of the control variate is depending on the parameter settings.

A time-consuming step to obtain the moment matched gamma distribution is the computation of the expectation and the variance of the true distribution because these expressions contain multiple modified Bessel functions of the first kind. In Appendix D we propose an efficient approximation of these modified Bessel functions of the first kind. To obtain a value of the inverse distribution function, using this control variate approach, a time-consuming numerical inversion still has to be performed and therefore we do not proceed in this way.

Approximation using moment matching

In the construction of the control variate we have seen that the gamma function is accurate in approximating the true distribution function using a moment matching technique. Whereas many other distributions are not appropriate for approximating the non-central chi-squared distribution, because they neither have a positive support nor the moment matching technique can be performed analytically, this is not the case for the log normal distribution.

Result 8.6. *As already mentioned in the previous chapter, the parameters, k_1 and k_2, of the moment matched gamma distribution are given by*

$$k_1 = \frac{\mu_{Iv}^2}{\sigma_{Iv}^2}, \qquad k_2 = \frac{\sigma_{Iv}^2}{\mu_{Iv}}.$$

To obtain the parameters, μ_{log} and σ_{Iv}^2, of the log normal distribution we solve

$$e^{\mu_{log}+\frac{1}{2}\sigma_{log}^2} = \mu_{Iv} \quad \text{and} \quad \left(e^{\sigma_{log}^2}-1\right) e^{2\mu_{log}+\sigma_{log}^2} = \sigma_{Iv}^2,$$

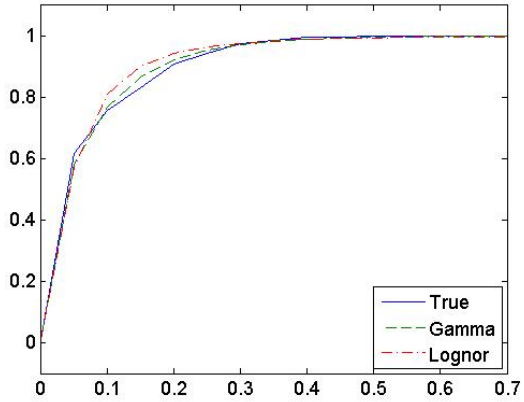
to obtain

$$\mu_{log} = \log\left(\mu_{Iv}\right) - \frac{1}{2}\log\left(\frac{\sigma_{Iv}^2+\mu_{Iv}^2}{\mu_{Iv}^2}\right) \quad \text{and} \quad \sigma_{log}^2 = \log\left(\frac{\sigma_{Iv}^2+\mu_{Iv}^2}{\mu_{Iv}^2}\right).$$

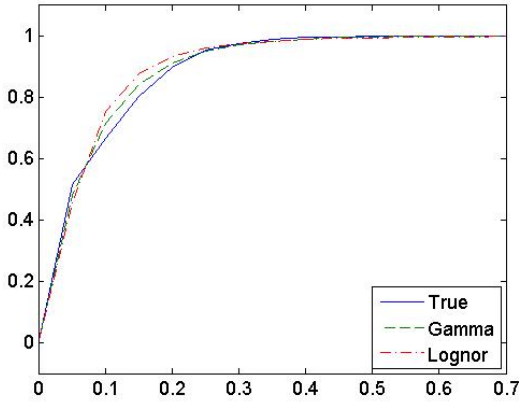
In Figure 22 we have plotted the moment matched gamma (Gamma) as well as the moment matched log normal distribution (lognor) with the true distribution.

Figure 22: True distribution versus moment matched gamma and log normal distribution

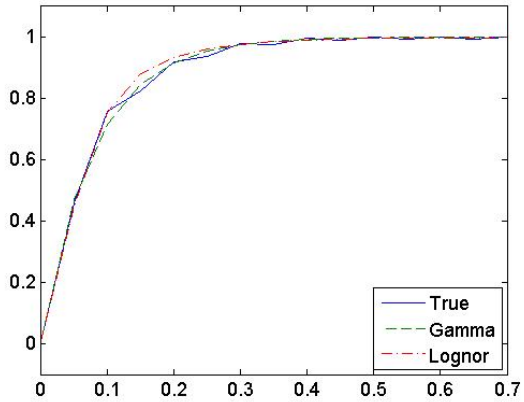
(a) Case I, $\nu(0) = \xi$, $\nu(1) = 0.06$



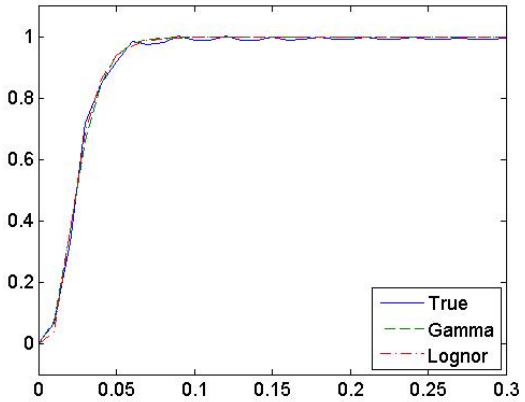
(b) Case II, $\nu(0) = \xi$, $\nu(1) = 0.06$



(c) Case III, $\nu(0) = \xi$, $\nu(1) = 0.06$



(d) Case IV, $\nu(0) = \xi$, $\nu(1) = 0.06$



The maximum absolute error for case I is equal to 0.089 and 0.23, for case II 0.11 and 0.24, for case III 0.043 and 0.16, for case IV 0.055 and 0.038, for the moment matched gamma and log normal distribution, respectively. The figures in Figure 22 show that the true distribution function can be approximated very well by the moment matched gamma and log normal distributions.

After performing some numerical tests, we found that this approach is still too time-consuming in simulating the asset price process. For that reason, we do not proceed using this approach of simulating the integrated variance process.