Approximating the Heston-Hull-White Model

Riaz Patel

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

September 11, 2019

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

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

Signed by candidate

Riaz Patel

September 11, 2019

Abstract

The hybrid Heston-Hull-White (HHW) model combines the Heston (1993) stochastic volatility and Hull and White (1990) short rate models. Compared to stochastic volatility models, hybrid models improve upon the pricing and hedging of long-dated options and equity-interest rate hybrid claims. When the Heston and Hull-White components are uncorrelated, an exact characteristic function for the HHW model can be derived. In contrast, when the components are correlated, the more useful case for the pricing of hybrid claims, an exact characteristic function cannot be obtained. Grzelak and Oosterlee (2011) developed two approximations for this correlated case, such that the characteristics functions are available. Within this dissertation, the approximations, referred to as the determinist and stochastic approximations, were implemented to price vanilla options. This involved extending the Carr and Madan (1999) method to a stochastic interest rate setting. The approximations were then assessed for accuracy and efficiency. In determining an appropriate benchmark for assessing the accuracy of the approximations, the full truncation Milstein and Quadratic Exponential (QE) schemes, which are popular Monte Carlo discretisation schemes for the Heston model, were extended to the HHW model. These schemes were then compared against the characteristic function for the uncorrelated case, and the QE scheme was found to be more accurate than the Milstein-based scheme. With the differences in performance becoming increasingly noticeable when the Feller (1951) condition was not satisfied and the maturity and volatility of the Hull-White model (η) was large. In assessing the accuracy of the approximations against the QE scheme, both approximations were similarly accurate when η was small. In contrast, when η was large, the stochastic approximation was more accurate than the deterministic approximation. However, the deterministic approximation was significantly faster than the stochastic approximation and the stochastic approximation displayed signs of potential instability. When η is small, the deterministic approximation is therefore recommended for use in applications such as calibration. With its shortcomings, the stochastic approximation could not be recommended. However, it did show promising signs of accuracy that warrants further investigation into its efficiency and stability.

Acknowledgements

I gratefully thank my supervisor, Dr Ralph Rudd for the mentorship and guidance. To the AIFMRM staff for their kind help and co-operation throughout my studies. Lastly, to my family and friends for their support, without which this would not be possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Hybrid stochastic volatility and interest rate models are a class of models that combine a stochastic volatility and stochastic interest rate component in a correlated manner. These models retain the advantages of stochastic volatility, whilst improving on the pricing and hedging of interest rate sensitive products, such as long-dated options and equity-interest rate hybrid claims. However, many of these hybrid models suffer from poor tractability. This is primarily owing to their added complexity

The Heston-Hull-White model is a hybrid model that combines the popular Heston (1993) stochastic volatility and Hull and White (1990) short rate models. When its component models are uncorrelated, an exact characteristic function can be derived. In contrast, when the components are correlated, which is the more useful case for the pricing of hybrid claims, an exact characteristic function cannot be obtained. Grzelak and Oosterlee (2011) developed two approximations for this correlated case, such that the characteristics functions are available. In conjunction with Fourier methods, these characteristic functions can be used to price options that are frequently traded in the market.

The efficient calibration of the Heston-Hull-White model is directly dependent on the accuracy of these prices and the speed at which they can be obtained. The ability to calibrate efficiently is a highly sought after feature that could lead to the Heston-Hull-White model becoming a popular choice for practical use. However, due to its relatively recent introduction, there has not been a great focus on assessing the characteristic functions that result for these approximations. As such, this dissertation aims to address this gap in literature and further explore the applications of these approximations from a practical perspective.

Following on from this chapter, Chapter 2 will introduce the model, along with its key features and properties. Chapter 3 will outline the approximations and derive their resulting characteristic functions. Chapter 4 considers pricing with these characteristic functions and includes a discussion on extending the Fourier based

methods to a stochastic interest rate setting. A Monte Carlo discretisation scheme will be used as a benchmark for assessing the accuracy of the approximations. In Chapter 5, the popular discretisation schemes for the Heston model are extended to the Heston-Hull-White model and are then compared to the prices obtained from the exact characteristic function for the uncorrelated case. Lastly, in Chapter 6, the approximations are assessed and in Chapter 7, conclusions are drawn based on the findings of this dissertation. It must be noted that for the purposes of this dissertation, all of the implementation has been done in MATLAB.

Chapter 2

The Heston-Hull-White Model

In this chapter the Heston and Hull-White models are outlined, and are then combined to form the Heston-Hull-White model. For all the models that are introduced, it is assumed that the relevant filtered probability space, which is denoted by $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{Q})$, consists of an appropriate sample space, sigma-algebra, filtration and risk-neutral measure. Throughout this chapter, properties and notation that will be useful for this dissertation are introduced. One of the features that will be looked at, is the affine nature of the Heston-Hull-White model. It will become apparent in Chapter 3, that this feature is particularly useful when deriving a model's characteristic function.

2.1 The Heston Model

Under the Heston specification, the asset price volatility is indirectly modelled through the variance process, which is assumed to follow the same dynamics as the one-factor short-rate model proposed by Cox *et al.* (2005).

The Heston model is described by the following system of stochastic differential equations (SDEs):

$$\begin{cases} dS_t = rS_t dt + \sqrt{v_t} S_t dW_t^s, & S_0 > 0, \\ dv_t = \kappa(\bar{v} - v_t)dt + \sigma\sqrt{v_t} dW_t^v, & v_0 > 0, \end{cases} \tag{2.1}$$

where $r > 0$ is the risk-free rate and $\rho_{s,v}$ is the correlation between W_t^s and W_t^v. The variance process is mean reverting. This means that the variance will revert to its long-run level of $\bar{v} > 0$ with $\kappa > 0$ controlling the speed at which this reversion occurs. The volatility of the variance process is denoted by $\sigma > 0$.

Under this specification, Feller (1951) showed that the variance process is guaranteed to be non-negative. Furthermore, the process is guaranteed to be positive if

$2\kappa\bar{v} > \sigma^2$. This has become widely known as the Feller condition[1]. In Theorem 2.1 the expectation and variance of v_t is derived.

Theorem 2.1. *For $t > s \geq 0$ and conditional on $\mathcal{F}_s$, the expectation and variance of v_t is given by*

$$\mathbb{E}[v_t \mid \mathcal{F}_s] := m_s = v_s e^{-\kappa(t-s)} + \bar{v}\big(1 - e^{-\kappa(t-s)}\big),$$
$$\mathbb{V}ar[v_t \mid \mathcal{F}_s] := S_s^2 = \frac{v_s\sigma^2}{\kappa} e^{-\kappa(t-s)}\big(1 - e^{-\kappa(t-s)}\big) + \frac{\bar{v}\sigma^2}{2k}\big(1 - e^{-\kappa(t-s)}\big)^2.$$

Proof. Through an application of Itô's lemma, the dynamics of $e^{\kappa t}v_t$ can be shown to be

$$d(e^{\kappa t}v_t) = \kappa\bar{v}e^{\kappa t} + \sigma e^{\kappa t}\sqrt{v_t}dW_t^v.$$

Integrating over $[s, t]$ yields

$$v_t = v_s e^{-\kappa(t-s)} + \bar{v}\big(1 - e^{-\kappa(t-s)}\big) + \int_s^t \sigma e^{\kappa(r-t)}\sqrt{v_r}dW_r^v. \tag{2.2}$$

Along with the Itô isometry and the properties of the Itô integral, (2.2) can be used to obtain the expressions for m_s and S_s^2. □

Note, that if the expectation or variance is conditional on $\mathcal{F}_0$, then the conditional notation is dropped, i.e.,

$$\mathbb{E}[v_t \mid \mathcal{F}_0] = \mathbb{E}[v_t].$$

2.2 The Hull-White Model

The Hull-White short-rate model is described by

$$dr_t = \lambda\big(\theta(t) - r_t\big)dt + \eta dW_t^r, \quad r_0 > 0. \tag{2.3}$$

As is the case with the Heston model's variance process, the Hull-White model reverts to its long-run mean, with $\lambda > 0$ controlling the speed at which this reversion occurs. The function $\theta(t)$, is deterministic and can be specified to fit the current term structure of interest rates exactly. The volatility of the short rate process is denoted by $\eta > 0$. Furthermore, solving the SDE in (2.3), reveals that under this specification, the short rate is normally distributed. This allows for the possibility of negative interest rate.

[1] Note that when calibrating the Heston model, the Feller condition is often violated (Andersen, 2007).

2.3 The Heston-Hull-White Model

The Heston-Hull-White (HHW) model is formed by combining (2.1) and (2.3) in a correlated manner. This results in the following system of SDEs:

$$\begin{cases} dS_t = r_t S_t dt + \sqrt{v_t} S_t dW_t^s, & S_0 > 0, \\ dv_t = \kappa(\bar{v} - v_t)dt + \sigma\sqrt{v_t} dW_t^v, & v_0 > 0, \\ dr_t = \lambda\big(\theta(t) - r_t\big)dt + \eta dW_t^r, & r_0 > 0, \end{cases}$$

where $\rho_{s,v}$, $\rho_{s,r}$ and $\rho_{v,r}$ denote the correlations between W_t^s, W_t^r and W_t^v.

Since the the dynamics have been modelled under a risk-neutral measure, the price of a claim can be obtained by risk-neutral valuation, i.e., by computing the expectation of the discounted payoff. However, due to the stochastic interest rate, the discount factor cannot be disentangled from the expectation. With a payoff of $\nu(S_T, v_T, r_T)$ at time T, the time-t price of a claim is given by

$$\mathbb{E}\big[N_T \nu(S_T, v_T, r_T) \mid \mathcal{F}_t\big], \tag{2.4}$$

where $N_T := e^{-Y_t} = e^{-\int_0^T r_s ds}$. In (2.4), N_T is referred to as the discount factor process and $Y_T = \int_0^T r_s ds$ is the integrated short rate process. Notice that the $\mathbb{Q}$ superscript that indicates a risk-neutral expectation, is dropped for notational convenience.

For the rest of this dissertation, it is assumed that $\theta(t) = \theta$. The Hull-White model is therefore reduced to the Vasicek (1977) model. Under the Vasicek model, the mean-reverting and normally distributed behaviour of the short rate is retained. However, the current term structure of interest rates cannot be fitted exactly. This assumption is often made when studying the HHW model. It simplifies the mathematics that will follow without detracting from the goals of this dissertation.

2.3.1 The HHW Log Dynamics

As is the case with the Heston model, it will often be simpler to work with the log asset price process. The resulting dynamics are referred to as the HHW log dynamics and are summarised in Theorem 2.2.

Theorem 2.2. *With $x_t = \log S_t$, the HHW log dynamics are described by the following system of SDEs:*

$$\begin{cases} dx_t = \big(r_t - \frac{1}{2}v_t\big)dt + \sqrt{v_t} dW_t^s, & x_0 = \log S_0, \\ dv_t = \kappa(\bar{v} - v_t)dt + \sigma\sqrt{v_t} dW_t^v, & v_0 > 0, \\ dr_t = \lambda\big(\theta(t) - r_t\big)dt + \eta dW_t^r, & r_0 > 0. \end{cases}$$

Proof. This result follows from a straightforward application of Itô's lemma. □

The HHW log dynamics have been expressed in terms of correlated Brownian motions. In Theorem 2.3, these dynamics are alternatively represented in terms of independent Brownian motions.

Theorem 2.3. *The HHW log dynamics can be expressed as*

$$d\mathbf{X}_t = \mu(\mathbf{X}_t)dt + \sigma(\mathbf{X}_t)d\mathbf{B}_t,$$

where

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ v_t \\ r_t \end{bmatrix},$$

$$\mu(\mathbf{X}_t) = \begin{bmatrix} r_t - \frac{1}{2}v_t \\ \kappa(\bar{v} - v_t) \\ \lambda(\theta - r_t) \end{bmatrix},$$

$$\sigma(\mathbf{X}_t) = \begin{bmatrix} \sqrt{v_t} & 0 & 0 \\ \rho_{s,v}\sigma\sqrt{v_t} & \sqrt{1-\rho_{s,v}^2}\sigma\sqrt{v_t} & 0 \\ \rho_{s,r}\eta & \frac{\rho_{v,r}-\rho_{s,r}\rho_{s,v}}{\sqrt{1-\rho_{s,v}^2}}\eta & \sqrt{1-\rho_{s,r}^2-\left(\frac{\rho_{v,r}-\rho_{s,r}\rho_{s,v}}{\sqrt{1-\rho_{s,v}^2}}\right)^2}\eta \end{bmatrix}$$

and $\mathbf{B}_t$ *is a standard 3-dimensional Brownian motion.*

Proof. Let $\mathbf{B}_t$ be a standard n-dimensional Brownian motion, $\mathbf{W}_t$ be a correlated n-dimensional Brownian motion and $\mathbf{p}$ be the corresponding symmetric correlation matrix of $\mathbf{W}_t$. With $\mathbf{L}$ as the lower triangular matrix from the Cholesky decomposition of $\mathbf{p}$, it can be easily verified that

$$\mathbf{W}_t = \mathbf{L}\mathbf{B}_t. \tag{2.5}$$

The result in (2.5) yields a method for constructing a correlated n-dimensional Brownian motion from a standard n-dimensional Brownian motion and can be used to represent the HHW log dynamics in terms of independent Brownian motions.

For the HHW log dynamics in Theorem 2.2, define $\mathbf{W}_t = \begin{bmatrix} W_t^s & W_t^v & W_t^r \end{bmatrix}^\top$ and $\mathbf{p}$ as

$$\mathbf{p} = \begin{bmatrix} 1 & \rho_{s,v} & \rho_{s,r} \\ * & 1 & \rho_{v,r} \\ * & * & 1 \end{bmatrix}$$

With $\mathbf{B}_t$ as a standard 3-dimensional Brownian motion, computing $\mathbf{W}_t = \mathbf{L}\mathbf{B}_t$ yields

$$\begin{aligned} W_t^s &= B_t^s, \\ W_t^v &= \rho_{s,v} B_t^s + \bar{A} B_t^v, \\ W_t^r &= \rho_{s,r} B_t^s + \bar{B} B_t^v + \bar{C} B_t^r, \end{aligned}$$

where

$$\bar{A} = \sqrt{1-\rho_{s,v}^2}, \qquad \bar{B} = \frac{\rho_{v,r} - \rho_{s,r}\rho_{s,v}}{\sqrt{1-\rho_{s,v}^2}}, \qquad \bar{C} = \sqrt{1-\rho_{s,r}^2 - B^2}.$$

Replacing the correlated Brownian motions in the HHW log dynamics with the independent ones and expressing the results in terms of matrix notion completes the proof. □

In terms of this independent Brownian motion representation, $\mathbf{X}_t$ is known as the state vector. The drift and volatility matrices are denoted by $\mu(\mathbf{X}_t)$ and $\sigma(\mathbf{X}_t)$. The symmetric matrix $\Sigma(\mathbf{X}_t) = \sigma(\mathbf{X}_t)\sigma(\mathbf{X}_t)^\top$ is the covariance matrix and for the HHW model this is given by

$$\Sigma(\mathbf{X}_t) = \begin{bmatrix} v_t & \rho_{s,v}\sigma v_t & \rho_{s,r}\eta\sqrt{v_t} \\ * & \sigma^2 v_t & \rho_{v,r}\eta\sqrt{v_t} \\ * & * & \eta^2 \end{bmatrix}. \tag{2.6}$$

With the following section in mind, Table 2.1 specifies three different scenarios for the correlations between the state variables. These three scenarios are labelled as the HHW-1, HHW-2 and HHW-3 models respectively. For each of these scenarios, the correlation between the stock price and variance process is unimportant, i.e., $\rho_{s,v}$ can be zero or non-zero. Since the covariance matrix is dependent on the correlation structure, it is also provided in the table.

	HHW-1	HHW-2	HHW-3
$\rho_{s,r}$	zero	non-zero	non-zero
$\rho_{v,r}$	zero	zero	non-zero
Σ	$\begin{bmatrix} v_t & p_{s,v}\sigma v_t & 0 \\ * & \sigma^2 v_t & 0 \\ * & * & \eta^2 \end{bmatrix}$	$\begin{bmatrix} v_t & p_{s,v}\sigma v_t & \rho_{s,r}\eta\sqrt{v_t} \\ * & \sigma^2 v_t & 0 \\ * & * & \eta^2 \end{bmatrix}$	$\begin{bmatrix} v_t & \rho_{s,v}\sigma v_t & \rho_{s,r}\eta\sqrt{v_t} \\ * & \sigma^2 v_t & \rho_{v,r}\sigma\eta\sqrt{v_t} \\ * & * & \eta^2 \end{bmatrix}$

Tab. 2.1: The covariance matrices for the different correlation scenarios.

2.4 The Affine Nature of the HHW Model

Definition 2.4 formalises what it means for a model to be affine. This definition is adapted from the one used by Duffie *et al.* (2000) [2].

Definition 2.4. A model, with a n-dimensional state vector $\mathbf{X}_t$, is said to be affine if

$$\mu(\mathbf{X}_t) = a_0 + a_1\mathbf{X}_t \tag{2.7}$$

$$\big(\Sigma(\mathbf{X}_t)\big)_{i,j} = (c_0)_{i,j} + (c_1)_{i,j,\bullet}\mathbf{X}_t \tag{2.8}$$

$$r(\mathbf{X}_t) = r_0 + r_1^{\intercal}\mathbf{X}_t \tag{2.9}$$

where $r_0 \in \mathbb{R}$, $(a_0, r_1) \in \mathbb{R}^n$, $(a_1, c_0) \in \mathbb{R}^{n\times n}$ and $c_1 \in \mathbb{R}^{n\times n\times n}$.

The conditions in (2.7) and (2.8) imply that for a model to be affine, each element of its drift and covariance matrices must be a linear function of the state variables. In (2.9), $r(\mathbf{X}_t)$ is referred to as the model's interest rate matrix, which, on an element-wise basis, must also must also be a linear function of the state variables. This therefore includes a constant interest rate or a stochastic interest rate that is driven by one or more of the state variables.

Since the HHW model's interest rate component is only dependent on the short rate, its interest rate matrix is given by:

$$r(\mathbf{X}_t) = \begin{bmatrix} 0 \\ 0 \\ r_t \end{bmatrix}.$$

With a_0, a_1, r_0 and r_1 defined in the following way, it is straightforward to see that the HHW drift and interest rate matrices satisfy (2.7) and (2.9):

$$a_0 = \begin{bmatrix} 0 \\ k\bar{v} \\ \lambda\theta \end{bmatrix}, \quad a_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 1 \\ 0 & -k & 0 \\ 0 & 0 & -\lambda \end{bmatrix}, \quad r_0 = 0, \quad r_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Determining whether or not the HHW covariance matrix satisfies (2.8) is more difficult. In sections 2.4.1 and 2.4.2, the covariance matrices for the each of scenarios in Table 2.1 are separately considered.

[2] Duffie *et al.* (2000) considered models that could include jumps. This is not applicable to the HHW model and hence the definition could be simplified somewhat. Furthermore, Duffie *et al.* (2000) use the dot product and matrix multiplication in their definition. In Definition 2.4, only matrix multiplication is used.

2.4.1 The Affine Nature of the HHW-1 Model

The matrices, c_0 and c_1, are found on an element-wise basis. The decomposition of each element of the HHW-1 covariance matrix is summarised in Table 2.2.

{i,j}	$(\Sigma_{\text{HHW-1}})_{i,j}$	$(c_0)_{i,j}$	$(c_1)_{i,j,\bullet}$
{1,1}	v_t	0	$\begin{bmatrix}0 & 1 & 0\end{bmatrix}$
{1,2}	$\rho_{s,v}\sigma v_t$	0	$\begin{bmatrix}0 & \rho_{s,v}\sigma & 0\end{bmatrix}$
{1,3}	0	0	$\begin{bmatrix}0 & 0 & 0\end{bmatrix}$
{2,2}	$\sigma^2 v_t$	0	$\begin{bmatrix}0 & \sigma^2 & 0\end{bmatrix}$
{2,3}	0	0	$\begin{bmatrix}0 & 0 & 0\end{bmatrix}$
{3,3}	η^2	η^2	$\begin{bmatrix}0 & 0 & 0\end{bmatrix}$

Tab. 2.2: Decomposition of the HHW-1 covariance matrix.

Recall, that the covariance matrix is symmetric and thus the decomposition of the remaining elements can be obtained from Table 2.2. Column three of the table can be combined to form c_0, i.e.,

$$c_0 = \begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & \eta^2\end{bmatrix}.$$

The i_{th} element of each of the vectors in column four are combined to form $(c_1)_{\bullet,\bullet,i}$, i.e.,

$$(c_1)_{\bullet,\bullet,1} = \begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}, \quad (c_1)_{\bullet,\bullet,2} = \begin{bmatrix}1 & \rho_{s,v}\sigma & 0\\ \rho_{s,v}\sigma & \sigma^2 & 0\\ 0 & 0 & 0\end{bmatrix}, \quad (c_1)_{\bullet,\bullet,3} = \begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

This can then be written in the following compact form:

$$c_1 = \begin{bmatrix}(0,1,0) & (0,\rho_{s,v}\sigma,0) & (0,0,0)\\ (0,\rho_{s,v}\sigma,0) & (0,\sigma^2,0) & (0,0,0)\\ (0,0,0) & (0,0,0) & (0,0,0)\end{bmatrix}.$$

For the HHW-1 model, the conditions in Definition 2.4 are satisfied and the model is affine. Note that by inspection, it is clear that each element of the HHW-1 covariance matrix is a linear function of the state variables. Hence, the decomposition that was given above, is not required when determining whether or not a model is affine. However, these matrices will be important when the characteristic function is derived.

2.4.2 The Affine Nature of the HHW-2 and HHW-3 Models

The presence of $\sqrt{v_t}$ in the HHW-2 and HHW-3 covariance matrices means that each element of their respective covariance matrices are not linear functions of the state variables and thus these models are not affine.

Chapter 3

The HHW Characteristic Function

Duffie *et al.* (2000) derived a method for determining the characteristic function of an affine model. Applying this method to the affine HHW-1 model, results in a closed-form expression for its characteristic function. On the other hand, the HHW-2 and HHW-3 models are not affine and the method of Duffie *et al.* (2000) cannot be used to determine their characteristic functions. In response to this, Grzelak and Oosterlee (2011) developed two different techniques for approximating the HHW-2 and HHW-3 models with affine forms. The characteristic functions for these approximated models can then be obtained through the result of Duffie *et al.* (2000).

In section 3.1 and 3.2, the approximations which are referred to as the deterministic and stochastic approximations respectively, are presented. In section 3.3, the exact characteristic function of the HHW-1 model and the two characteristic functions for the affine approximations of the HHW-2 model are derived. Compared to the HHW-2 case, the affine approximations of the HHW-3 model bring greater complexity. In terms of assessing the deterministic and stochastic approximations, this complexity does not add any value and hence the HHW-3 case is not considered in this chapter.

3.1 The Deterministic Approximation

Under the deterministic approximation, the terms in the HHW-2 covariance matrix that contain $\sqrt{v_t}$ are replaced by their expectation. The resulting approximated model is referred to as the HHW-2D model and its affine covariance matrix is given by

$$\Sigma_{\text{HHW-2D}} = \begin{bmatrix} v_t & \rho_{s,v}\sigma v_t & \rho_{s,r}\eta\mathbb{E}\left[\sqrt{v_t}\right] \\ * & \sigma^2 v_t & 0 \\ * & * & \eta^2 \end{bmatrix}$$

In order to determine the HHW-2D characteristic function, an expression for the expectation of $\sqrt{v_t}$ is needed. This is presented in Theorem 3.1.

Theorem 3.1. *For $t > 0$, the expectation and variance of $\sqrt{v_t}$ is given by*

$$\mathbb{E}\big[\sqrt{v_t}\big] = \sqrt{2c(t)}e^{-\frac{\lambda(t)}{2}}\frac{\Gamma\left(\frac{1+d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}{}_1F_1\left(\frac{d+1}{2}, \frac{d}{2}, \frac{\lambda(t)}{2}\right),$$

$$\mathbb{V}ar\big[\sqrt{v_t}\big] = c(t)\Big(d + \lambda(t)\Big) - 2c(t)e^{-\lambda(t)}\left(\frac{\Gamma\left(\frac{1+d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}{}_1F_1\left(\frac{d+1}{2}, \frac{d}{2}, \frac{\lambda(t)}{2}\right)\right)^2,$$

where

$$c(t) = \frac{1}{4k}\sigma^2(1 - e^{-kt}), \qquad d = \frac{4k\bar{v}}{\sigma^2}, \qquad \lambda(t) = \frac{4kv_0e^{-kt}}{\sigma^2(1 - e^{-kt})}.$$

$\Gamma(x)$ is the gamma function and ${}_1F_1\,(r, s, z)$ is the confluent hypergeometric function.

Proof. The proof of the $\mathbb{E}\big[\sqrt{v_t}\big]$ can be found in Dufresne (2001). With $\mathbb{E}\big[\sqrt{v_t}\big]$ and $\mathbb{E}\big[v_t\big]$ from Theorem 2.1, the $\mathbb{V}ar\big[\sqrt{v_t}\big]$ can be computed as

$$\mathbb{V}ar\big[\sqrt{v_t}\big] = \mathbb{E}\big[v_t\big] - (\mathbb{E}\big[\sqrt{v_t}\big])^2. \tag{3.1}$$

□

In Grzelak and Oosterlee (2011), the hypergeometric function in the expression for the expectation and variance of $\sqrt{v_t}$ was expressed in terms of its infinite sum representation. This allowed for further simplifications to be made. For this dissertation, the result is left unchanged because in this form it is easier to implement in MATLAB.

It will be shown in section 3.3 that in order to evaluate the HHW-2D characteristic function, an integral that involves $\mathbb{E}\big[\sqrt{v_t}\big]$ must be computed. The expression for $\mathbb{E}\big[\sqrt{v_t}\big]$ is complicated and hence this integral is computationally expensive to evaluate. Grzelak and Oosterlee (2011) therefore suggest approximating $\mathbb{E}\big[\sqrt{v_t}\big]$ by a simpler expression. The details are given in Proposition 3.2.

Proposition 3.2. *For $t > 0$, $\mathbb{E}\big[\sqrt{v_t}\big]$ can be approximated by*

$$\Lambda(t) = \sqrt{c(t)(\lambda(t) - 1) + c(t)d + \frac{c(t)d}{2(d + \lambda(t))}}.$$

where $c(t)$, d and $\lambda(t)$ are given in Theorem 3.1.

Proof. Consider a random variable, X, whereby its first two moments are finite and a differentiable function of this random variable, $\psi(X)$. This function can be approximated by the first two terms of its Taylor series about $X = \mathbb{E}[X]$, i.e.,

$$\psi(X) \approx \psi\big(\mathbb{E}[X]\big) + \big(X - \mathbb{E}[X]\big)\frac{d\psi}{dX}\big(\mathbb{E}[X]\big).$$

The variance of $\psi(X)$ can then be approximated by

$$\begin{aligned}\mathbb{V}ar\big[\psi(X)\big] &\approx \mathbb{V}ar\Big[\psi\big(\mathbb{E}[X]\big) + \big(X - \mathbb{E}[X]\big)\frac{d\psi}{dX}\big(\mathbb{E}[X]\big)\Big] \\ &= \Big(\frac{d\psi}{dX}\big(\mathbb{E}[X]\big)\Big)^2 \mathbb{V}ar[X].\end{aligned}$$

The second line follows from the fact that $\psi(X)$ and its derivative evaluated at $\mathbb{E}[X]$ are constants.

Setting $\psi(X) = \sqrt{v_t}$, evaluating the required derivative and simplifying yields the following approximation for the variance of $\sqrt{v_t}$:

$$\mathbb{V}ar\big[\sqrt{v_t}\big] \approx \frac{1}{4}\frac{\mathbb{V}ar[v_t]}{\mathbb{E}[v_t]}.$$

Replacing the left-hand side with (3.1) yields

$$\mathbb{E}\big[v_t\big] - \big(\mathbb{E}\big[\sqrt{v_t}\big]\big)^2 \approx \frac{1}{4}\frac{\mathbb{V}ar[v_t]}{\mathbb{E}[v_t]}.$$

Rearranging results in the following approximation for the mean of $\sqrt{v_t}$:

$$\mathbb{E}\big[\sqrt{v_t}\big] \approx \sqrt{\mathbb{E}[v_t] - \frac{1}{4}\frac{\mathbb{V}ar[v_t]}{\mathbb{E}[v_t]}}.$$

Substituting the expressions for the mean and variance of v_t, from Theorem 2.1, into the approximation above and simplifying completes the proof. □

Grzelak and Oosterlee (2011) show that the expression under the square root sign of $\Lambda(t)$ is always non-negative when the Feller condition is satisfied. For parameters that do not satisfy the Feller condition, the expression is non-negative if $\frac{8k\bar{v}}{\sigma^2} > 1$. $\Lambda(t)$ is only valid if either of these conditions are met. If they are not met, the exact expression for $\mathbb{E}\big[\sqrt{v_t}\big]$ must be used.

$\Lambda(t)$ is simpler to compute than the exact expression for $\mathbb{E}\big[\sqrt{v_t}\big]$. However, it still does not result in a closed-form expression for the integral that must be computed when evaluating the HHW-2D characteristic function. With this in mind, Grzelak and Oosterlee (2011) further approximate $\Lambda(t)$ with a simpler form. The details are given in Proposition 3.3.

Proposition 3.3. *For $t > 0$, $\Lambda(t)$ can be approximated by*

$$\hat{\Lambda}(t) = a + be^{-ct} \tag{3.2}$$

where

$$a = \sqrt{\bar{v} - \frac{\sigma^2}{8k}}, \qquad b = \sqrt{v_0} - a, \qquad c = -\log\big(b^{-1}(\Lambda(1) - a)\big).$$

	$\Lambda(t)$	$\hat{\Lambda}(t)$
$t \to \infty$	$\sqrt{\bar{v} - \frac{\sigma^2}{8k}}$	a
$t \to 0$	$\sqrt{v_0}$	$a + b$
$t \to 1$	$\Lambda(1)$	$a + be^{-c}$

Tab. 3.1: $\Lambda(t)$ and $\hat{\Lambda}(t)$ as $t \to 0$, $t \to 1$ and $t \to \infty$.

Proof. Grzelak and Oosterlee (2011) propose that $\Lambda(t)$ be approximated by (3.2), where a, b and c are determined by matching $\Lambda(t)$ and $\hat{\Lambda}(t)$ as $t \to 0$, $t \to 1$ and $t \to \infty$. The limits for these cases are summarised in Table 3.1.
Equating columns 1 and 2 of Table 3.1 and solving for a, b and c completes the proof. □

$\hat{\Lambda}(t)$ is valid when $\Lambda(t)$ is valid. However, care must be taken to ensure that $\bar{v} > \frac{\sigma^2}{8k}$ and hence that a is well defined. If this condition is not satisfied, then $\Lambda(t)$ should be used.

In Figure 3.1 the accuracy of approximations 1 and 2, Propositions 3.2 and 3.3 respectively, are assessed against the true value of the expectation of $\sqrt{v_t}$. Various times ranging from 0 to 10 were considered and two different parameter sets were used. For parameter set 1, $v_0 = 0.04$, $\kappa = 1.2$, $\sigma = 0.1$ and $\bar{v} = 0.03$. For parameter set 2, $v_0 = 0.06$, $\kappa = 2.5$, $\sigma = 0.5$ and $\bar{v} = 0.06$.

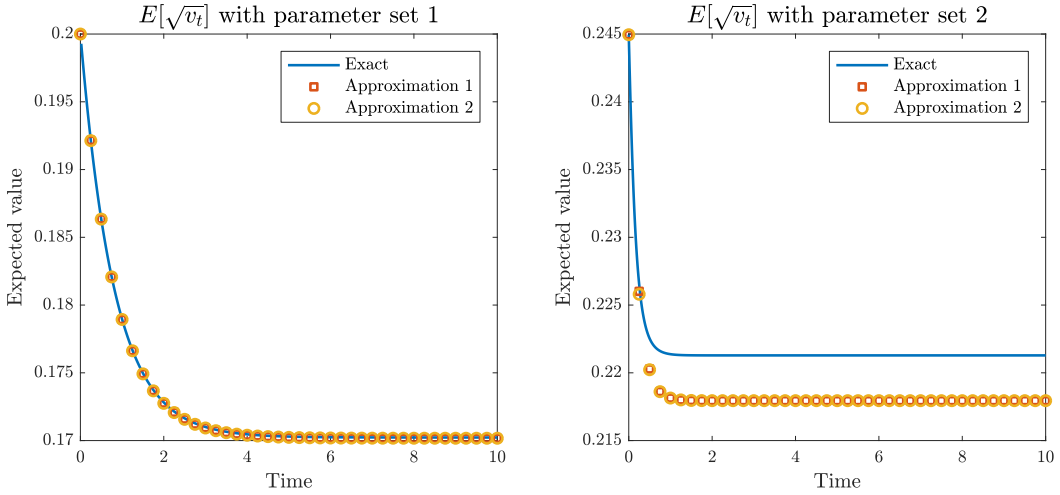


Fig. 3.1: The accuracy of the two approximations for two different parameter sets.

The figure shows that for both parameter sets, the two approximations follow each other closely. For parameter set 1, the approximations are also really accurate when compared to the true value. However, this is not the case for the second pa-

rameter set, with a clear bias arising. The accuracy of the approximations are therefore parameter dependent and this should be checked when considering whether or not these approximations should be used.

3.2 The Stochastic Approximation

The stochastic approximation involves adding $\sqrt{v_t}$ to the system of state variables that are modelled. The result of doing this is that the presence of $\sqrt{v_t}$ in the HHW-2 covariance matrix no longer nullifies its affine nature. In order to implement the stochastic approximation, the dynamics of $\sqrt{v_t}$ need to be found. Since $\sqrt{v_t}$ is not twice differentiable at the origin, Itô's lemma cannot be applied to find its exact dynamics (Jäckel, 2004). To this end, Theorem 3.4 shows how the dynamics of $\sqrt{v_t}$ can be approximated.

Theorem 3.4. *The dynamics of $\sqrt{v_t}$ can be approximated by*

$$d\xi_t = \mu^{\xi}(t)dt + \psi^{\xi}(t)dW_t^v, \qquad \xi_0 = \sqrt{v_0},$$

where

$$\begin{aligned}\mu^{\xi}(t) =& \frac{1}{2\sqrt{2}}\frac{\Gamma(\frac{1+d}{2})}{\sqrt{c(t)}}\left[{}_1F_1\left(-\frac{1}{2},\frac{d}{2},-\frac{\lambda(t)}{2}\right)\Gamma\left(\frac{d}{2}\right)^{-1}\frac{\sigma^2 e^{-kt}}{2}\right.\\ &\left.+\,{}_1F_1\left(\frac{1}{2},\frac{2+d}{2},-\frac{\lambda(t)}{2}\right)\Gamma\left(\frac{2+d}{2}\right)^{-1}\frac{v_0 k}{1-e^{kt}}\right],\end{aligned}$$

$$\psi^{\xi}(t) = \left(k(\bar{v}-v_0)e^{kt} - 2\mathbb{E}\left[\sqrt{v_t}\right]\mu^{\xi}(t)\right)^{\frac{1}{2}}.$$

The expressions for $c(t)$, d and $\lambda(t)$ are as in Theorem 3.1. $\Gamma(x)$ is the gamma function and ${}_1F_1\left(r,s,z\right)$ is the confluent hypergeometric function.

Proof. Cox *et al.* (2005) showed that v_t follows a noncentral chi-squared distribution which, by Patnaik (1949), can be approximated by a central Chi-Squared distribution. The result of Fisher (1928) can then be used to approximate the central Chi-Squared distribution by a normal distribution. The dynamics of ξ_t is constructed to be normally distributed with the required mean and variance. A more in-depth discussion surrounding this result can be found in Grzelak and Oosterlee (2011). □

Adding $d\xi_t$ to the log-HHW dynamics yields the following dynamics:

$$\begin{cases} dS_t = r_t S_t dt + \sqrt{v_t}S_t dW_t^s, & S_0 > 0,\\ dv_t = k(\bar{v}-v_t)dt + \sigma\sqrt{v_t}dW_t^v, & v_0 > 0,\\ dr_t = \lambda(\theta - r_t)dt + \eta dW_t^r, & r_0 > 0,\\ d\xi_t = \mu^{\xi}(t)dt + \psi^{\xi}(t)dW_t^v, & \xi_0 = \sqrt{v_0}. \end{cases} \tag{3.3}$$

The stochastic approximation is obtained by imposing the HHW-2 correlation structure on these dynamics. The resulting approximated model is referred to as the HHW-2S model. With $\sqrt{v_t} \approx \xi_t$, these dynamics can be expressed in terms of an independent system with the following drift and covariance matrices[1]

$$\mu_{\text{HHW-2S}} = \begin{bmatrix} r_t - \frac{1}{2}v_t \\ k(\bar{v} - v_t) \\ \lambda(\theta - r_t) \\ \mu^{\xi}(t) \end{bmatrix}, \quad \Sigma_{\text{HHW-2S}} \approx \begin{bmatrix} v_t & p_{s,v}\sigma v_t & \rho_{s,r}\eta\xi_t & \rho_{s,v}\psi^{\xi}(t)\xi_t \\ * & \sigma^2 v_t & 0 & \sigma\psi^{\xi}(t)\xi_t \\ * & * & \eta^2 & 0 \\ * & * & * & \left(\psi^{\xi}(t)\right)^2 \end{bmatrix}.$$

These matrices are clearly linear functions of the state variables. Furthermore, the interest component of the HHW-2S model remains unchanged and hence the model is affine.

3.3 The Characteristic Functions

In Theorem 3.5, the method of Duffie *et al.* (2000) for determining the characteristic function of an affine model is presented [2].

Theorem 3.5. *For an affine model, $\mathbf{X}_t$ with n state variables, the discounted joint characteristic function is given by*

$$\bar{\phi}(\mathbf{u}, \mathbf{X}_t, t, T) = \mathbb{E}\left[N_T e^{i\mathbf{u}\mathbf{X}_T} \mid \mathcal{F}_t\right] = \exp\Big(A(\mathbf{u}, \tau) + \mathbf{B}^{\top}(\mathbf{u}, \tau)\mathbf{X}_t\Big). \tag{3.4}$$

where N_T is the discount factor process, $\mathbf{u}$ is a n-element row vector and $\tau = T - t$. The function $A(\mathbf{u}, \tau)$ and the n-dimensional vector-valued function $\mathbf{B}(\mathbf{u}, \tau)$ satisfy the following system of complex-valued ordinary differential equations (ODEs):

$$\begin{cases} \frac{d}{d\tau}\mathbf{B}(\mathbf{u}, \tau) = -r_1 + a_1^{\top}\mathbf{B}(\mathbf{u}, \tau) + \frac{1}{2}\mathbf{B}^{\top}(\mathbf{u}, \tau)c_1\mathbf{B}(\mathbf{u}, \tau), & \mathbf{B}^{\top}(\mathbf{u}, \tau) = i\mathbf{u}, \\ \frac{d}{d\tau}A(\mathbf{u}, \tau) = -r_0 + \mathbf{B}^{\top}(\mathbf{u}, \tau)a_0 + \frac{1}{2}\mathbf{B}^{\top}(\mathbf{u}, \tau)c_0\mathbf{B}(\mathbf{u}, \tau), & A(\mathbf{u}, 0) = 0, \end{cases} \tag{3.5}$$

where a_0, a_1, c_0, c_1, r_0 and r_1 are the matrices given in definition 2.4. $\mathbf{B}^{\top}(\mathbf{u}, \tau)c_1\mathbf{B}(\mathbf{u}, \tau)$ is a n-dimensional vector where the k_{th} element is equal to $\mathbf{B}^{\top}(\mathbf{u}, \tau)(c_1)_{\bullet,\bullet,k}\mathbf{B}(\mathbf{u}, \tau)$.

Proof. The proof of this result can be found in Duffie *et al.* (2000). □

It must be stressed that Theorem 3.5 yields what Duffie *et al.* (2000) refer to as the discounted joint characteristic function. The definition of the discounted joint

[1] The proof of this result is almost identical to the one for Theorem 2.3.

[2] As was the case with Definition 2.4, the method of Duffie *et al.* (2000) could be simplified due to the fact that the HHW model does not contain jumps.

characteristic function is given in (3.4). The inclusion of the discount factor is what differentiates it from the usual definition of the joint characteristic function[3].

Theorem 3.5 is relatively straightforward to apply. The functional form of the relevant model's joint characteristic function is given in (3.4) . The explicit solution can then be found by solving the system of ODEs that result from performing the matrix multiplication in (3.5). In the following sections, the theorem will be applied to determine the HHW-1, HHW-2D and HHW-2S characteristic functions for the log asset price, i.e., $x_t = \log S_t$. The characteristic functions for individual state variables can be found by setting the appropriate terms in $\mathbf{u}$ to zero. Specifically, to determine the characteristic function for x_t, only the first element of $\mathbf{u}$ must be nonzero.

3.3.1 The HHW-1 Characteristic Function

Theorem 3.6. *The HHW-1 discounted characteristic function is given by*

$$\bar{\phi}_{\text{HHW-1}}(u, \mathbf{X}_t, t, T) = \exp\Big(A(u,\tau) + B(u,\tau)x_t + C(u,\tau)r_t + D(u,\tau)v_t\Big), \tag{3.6}$$

where

$$\begin{aligned}
B(u,\tau) &= iu,\\
C(u,\tau) &= \frac{1}{\lambda}\Big(iu-1\Big)\Big(1-e^{-\lambda\tau}\Big),\\
D(u,\tau) &= \frac{1-e^{-D_1\tau}}{\sigma^2\big(1-ge^{-D_1\tau}\big)}\Big(\kappa-\sigma\rho_{s,v}iu-D_1\Big),\\
A(u,\tau) &= \lambda\theta I_1(u,\tau) + \kappa\bar{v}I_2(u,\tau) + \frac{1}{2}\eta^2 I_3(u,\tau),
\end{aligned}$$

with

$$\begin{aligned}
D_1 &= \sqrt{(\sigma\rho_{s,v}iu-\kappa)^2 - \sigma^2 iu(iu-1)},\\
g &= \frac{\kappa-\sigma\rho_{s,v}iu-D_1}{\kappa-\sigma\rho_{s,v}iu+D_1},\\
I_1(u,\tau) &= \frac{1}{\lambda}\Big(iu-1\Big)\Big(\tau+\frac{1}{\lambda}\big(e^{-\lambda\tau}-1\big)\Big),\\
I_2(u,\tau) &= \frac{\tau}{\sigma^2}\Big(\kappa-\sigma\rho_{s,v}iu-D_1\Big) - \frac{2}{\sigma^2}\log\left(\frac{1-ge^{-D_1\tau}}{1-g}\right),\\
I_3(u,\tau) &= \frac{1}{2\lambda^3}\Big(i+u\Big)^2\Big(3+e^{-2\lambda\tau}-4e^{-\lambda\tau}-2\lambda\tau\Big).
\end{aligned}$$

[3] Going forward, unless otherwise stated, "characteristic function" should be read as "discounted characteristic function".

Proof. Recall from section 2.4.1 that the HHW-1 affine decomposition is given by

$$a_0 = \begin{bmatrix} 0 \\ \kappa\bar{v} \\ \lambda\theta \end{bmatrix}, \quad a_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 1 \\ 0 & -\kappa & 0 \\ 0 & 0 & -\lambda \end{bmatrix}, \quad c_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \eta^2 \end{bmatrix} \quad r_0 = 0, \quad r_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

$$c_1 = \begin{bmatrix} (0,1,0) & (0,\rho_{s,v}\sigma,0) & (0,0,0) \\ (0,\rho_{s,v}\sigma,0) & (0,\sigma^2,0) & (0,0,0) \\ (0,0,0) & (0,0,0) & (0,0,0) \end{bmatrix}.$$

Applying Theorem 3.5 with $\mathbf{B}^\top(u,\tau) = \begin{bmatrix} B(u,\tau) & D(u,\tau) & C(u,\tau) \end{bmatrix}$ yields the form of the characteristic function given in (3.6). The functions $A(u,\tau)$, $B(u,\tau)$, $C(u,\tau)$ and $D(u,\tau)$ are determined by solving the following system of ODEs:

$$\begin{cases} \frac{d}{d\tau}B(\tau) = 0, & B(u,0) = iu, \\ \frac{d}{d\tau}C(\tau) = -1 - \lambda C(\tau) + B(\tau), & C(u,0) = 0, \\ \frac{d}{d\tau}D(\tau) = \frac{1}{2}\big(B(\tau)-1\big) + \big(\sigma\rho_{s,v}B(\tau) - k\big)D(\tau) + \frac{1}{2}\sigma^2 D^2(\tau), & D(u,0) = 0, \\ \frac{d}{d\tau}A(\tau) = \lambda\theta C(\tau) + k\bar{v}D(\tau) + \frac{1}{2}\eta^2 C^2(\tau), & A(u,0) = 0. \end{cases}$$

$B(\tau)$ can be solved by direct integration. With the solution of $B(\tau)$, $C(\tau)$ is a first-order linear ODE and can be solved using the integrating factor technique. The ODE for $D(\tau)$ is the same as the Heston Ricatti equation and a well-laid-out derivation of its the solution can be found in Rouah (2013, p. 13-14) . With the solutions for $C(\tau)$ and $D(\tau)$, $A(\tau)$ can be determined by direct integration. □

3.3.2 The HHW-2D Characteristic Function

Theorem 3.7. *The HHW-2D discounted characteristic function is given by*

$$\bar{\phi}_{\text{HHW-2D}}(u, \mathbf{X}_t, t, T) = \exp\Big(\widetilde{A}(u,\tau) + B(u,\tau)x_0 + D(u,\tau)r_0 + D(u,\tau)v_0\Big), \tag{3.7}$$

where $B(u,\tau)$, $C(u,\tau)$ and $D(u,\tau)$ are as they were for the HHW-1 characteristic function (Theorem 3.6). $\widetilde{A}(u,\tau)$, which is expressed in terms of $A(u,\tau)$ from the HHW-1 characteristic function, is given by

$$\begin{aligned} \widetilde{A}(u,\tau) &= A(u,\tau) + \rho_{s,r}\eta \int_0^\tau \mathbb{E}\big[\sqrt{v_{T-s}}\big] B(u,s)C(u,s)ds \\ &= A(u,\tau) + \rho_{s,r}\eta I_4(u,\tau). \end{aligned} \tag{3.8}$$

If the second approximation for $\mathbb{E}\big[\sqrt{v_t}\big]$ is used (i.e., $\hat{\Lambda}(t)$ from Proposition 3.3), $I_4(u,\tau)$ can be simplified to the following closed-form expression:

$$I_4(u,\tau) \approx -\frac{1}{\lambda}\big(iu+u^2\big)\left[\frac{b}{c}\big(e^{-ct}-e^{-cT}\big) + a\tau + \frac{a}{\lambda}\big(e^{-\lambda\tau}-1\big) + \frac{b}{c-\lambda}e^{-cT}\big(1-e^{-\tau(\lambda-c)}\big)\right],$$

where a, b and c are given in Proposition 3.3.

Proof. Except for c_0, the HHW-2D affine decomposition is the same as the HHW-1 model's. It is relatively straightforward to see that c_0 is given by

$$c_0 = \begin{bmatrix} 0 & 0 & \rho_{s,r}\eta\mathbb{E}\big[\sqrt{v_t}\big] \\ 0 & 0 & 0 \\ \rho_{s,r}\eta\mathbb{E}\big[\sqrt{v_t}\big] & 0 & \eta^2 \end{bmatrix}.$$

The characteristic function is found by applying Theorem 3.5. The system of ODEs for $\mathbf{B}(u,\tau)$ does not depend on c_0 and hence the expressions for $B(u,\tau)$, $C(u,\tau)$ and $D(u,\tau)$ are the same as in the HHW-1 characteristic function. $\widetilde{A}(u,\tau)$ is determined by solving the following ODE:

$$\frac{d}{d\tau}\widetilde{A}(\tau) = \lambda\theta C(\tau) + k\bar{v}D(\tau) + \frac{1}{2}\eta^2 C^2(\tau) + \rho_{s,r}\eta\mathbb{E}\big[\sqrt{v_t}\big]B(\tau)C(\tau), \qquad \widetilde{A}(u,0) = 0.$$

Directly integrating this ODE and recognising the first three terms as $A(u,\tau)$ yields (3.8).

Substituting $B(u,s)$, $C(u,s)$ and $\hat{\Lambda}(t)$ into the expression for $I_4(u,\tau)$ yields

$$I_4(u,\tau) \approx -\frac{1}{\lambda}(iu+u^2)\int_0^\tau \Big(a + be^{-c(T-s)}\Big)\Big(1-e^{-\lambda s}\Big)ds. \tag{3.9}$$

Evaluating the integral completes the proof. □

Note that if the exact value or the first approximation for the expectation of $\sqrt{v_t}$ is used, $I_4(u,\tau)$ in (3.8) will need to be approximated by a numerical integration method. For this dissertation, the trapezoidal integration rule, which is outlined in the next chapter, was used.

3.3.3 The HHW-2S Characteristic Function

Theorem 3.8. *The HHW-2S discounted characteristic function is given by*

$$\bar{\phi}_{\text{HHW-2S}}(u,\mathbf{X}_t,t,T) = \exp\Big(\widehat{A}(u,\tau) + B(u,\tau)x_t + C(u,\tau)r_t + D(u,\tau)v_t + E(u,\tau)\xi_t\Big), \tag{3.10}$$

where $B(u,\tau)$, $C(u,\tau)$ and $D(u,\tau)$ are as they were for the HHW-1 characteristic function (Theorem 3.6). $E(u,\tau)$ and $\widehat{A}(u,\tau)$ satisfy the following system of ODEs:

$$\begin{cases} \frac{d}{d\tau}E(\tau) = \rho_{s,r}\eta B(\tau)C(\tau) + \big(\rho_{s,v}B(\tau) + \sigma D(\tau)\big)\psi^\xi(t)E(\tau), & E(u,0) = 0, \\ \frac{d}{d\tau}\widehat{A}(\tau) = \kappa\bar{v}D(\tau) + \lambda\theta C(\tau) + \mu^\xi(t)E(\tau) + \frac{1}{2}\eta^2C^2(\tau) + \frac{1}{2}\big(\psi^\xi(t)\big)^2E^2(\tau), & \widehat{A}(u,0) = 0. \end{cases}$$

Proof. Following a similar process to the one outlined in section 2.4.1, the HHW-2S affine decomposition can be shown to be

$$a_0 = \begin{bmatrix} 0 \\ \kappa\bar{v} \\ \lambda\theta \\ \mu^\xi(t) \end{bmatrix}, \quad a_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 1 & 0 \\ 0 & -\kappa & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad c_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \eta^2 & 0 \\ 0 & 0 & 0 & \big(\psi^\xi(t)\big)^2 \end{bmatrix} \quad r_0 = 0, \quad r_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

$$c_1 = \begin{bmatrix} (0,1,0,0) & (0,\rho_{s,v}\sigma,0,0) & (0,0,0,\rho_{s,r}\eta) & (0,0,0,\rho_{s,v}\psi^{\xi}(t)) \\ (0,\rho_{s,v}\sigma,0,0) & (0,\sigma^2,0,0) & (0,0,0,0) & (0,0,0,\sigma\psi^{\xi}(t)) \\ (0,0,0,\rho_{s,r}\eta) & (0,0,0,0) & (0,0,0,0) & (0,0,0,0) \\ (0,0,0\rho_{s,v}\psi^{\xi}(t)) & (0,0,0,\sigma\psi^{\xi}(t)) & (0,0,0,0) & (0,0,0,0) \end{bmatrix}.$$

Applying Theorem 3.5 with $\mathbf{B}^{\top}(u,\tau) = \begin{bmatrix} B(u,\tau) & D(u,\tau) & C(u,\tau) & E(u,\tau) \end{bmatrix}$ yields the following system of ODEs:

$$\begin{cases} \frac{d}{d\tau}B(\tau) = 0, & B(u,0) = iu, \\ \frac{d}{d\tau}C(\tau) = -1 - \lambda C(\tau) + B(\tau), & C(u,0) = 0, \\ \frac{d}{d\tau}D(\tau) = \frac{1}{2}\big(B(\tau)-1\big) + \big(\sigma\rho_{s,v}B(\tau) - \kappa\big)D(\tau) + \frac{1}{2}\sigma^2 D^2(\tau), & D(u,0) = 0, \\ \frac{d}{d\tau}E(\tau) = \rho_{s,r}\eta B(\tau)C(\tau) + \big(\rho_{s,v}B(\tau) + \sigma D(\tau)\big)\psi^{\xi}(t)E(\tau), & E(u,0) = 0, \\ \frac{d}{d\tau}\widehat{A}(\tau) = \kappa\bar{v}D(\tau) + \lambda\theta(\tau) + \mu^{\xi}(t)E(\tau) + \frac{1}{2}\eta^2 C^2(\tau) + \frac{1}{2}\big(\psi^{\xi}(t)\big)^2 E^2(\tau), & \widehat{A}(u,0) = 0. \end{cases}$$

The ODEs and hence the expressions for $B(\tau)$, $C(\tau)$ and $D(\tau)$ are the same as for the HHW-1 and HHW-2D characteristic functions. In the current literature, there is no closed-form solution for $E(\tau)$ and $A(\tau)$. A numerical method, such as the Runge-Kutta family of methods, is employed to approximately solve these two ODEs. □

MATLAB has many different options for solving ODEs. In terms of solving the system in Theorem 3.8, `ode23` was found to be the most stable and hence it was used for this dissertation. Furthermore, in solving this system, the hypergeometric function, from Theorem 3.4, must be evaluated repeatedly. MATLAB's built-in hypergeometric function, which is part of the Symbolic Math Toolbox, was found to be extremely inefficient. This significantly slowed down the evaluation of the HHW-2S characteristic function. To address this problem, an efficient hypergeometric function implementation by Brookes (2016) was used. Note that this implementation was also used for Theorem 3.1.

In the derivations of the characteristic functions, the components of $\mathbf{B}^{\top}(u,\tau)$ were defined in an unusual order. This was done so that the resulting characteristic functions were consistent with those presented in Grzelak and Oosterlee (2011). The differences were due to the fact that Grzelak and Oosterlee (2011) preferred to change the order of the variables within the HHW state vector. For reference, they refer to the HHW-2D and HHW-2S characteristic functions as the H1-HW and H2-HW characteristic functions. Different labels were used because the HHW-1 case is not presented in Grzelak and Oosterlee (2011).

3.3.4 Extensions to the HHW-3 Model

A brief discussion on the extension to the HHW-3 case is considered here. In extending the deterministic approximation to the HHW-3 case, the same method, of

replacing the terms that contain $\sqrt{v_t}$ with their expectations, is applied. However, instead of two terms, four terms in the HHW-3 covariance matrix will be approximated. Similarly, for the stochastic approximation, the same dynamics from Theorem 3.4 is added to the state vector but the HHW-3 correlation structure is imposed on the resulting independent Brownian motion representation in (3.3). Due to the additional approximations, the resulting characteristic functions are more complicated and could potentially result in lower levels of accuracy than for the HHW-2 case.

Grzelak and Oosterlee (2011) derive the characteristic function that results from applying the deterministic approximation to the HHW-3 case. They refer to this as the H3-HW characteristic function. Their derivation shows that regardless of how the expectation of $\sqrt{v_t}$ is evaluated, there is no closed-form solution for the H3-HW characteristic function. Furthermore, the integral that must be numerically computed to evaluate the characteristic function, is more complicated than the one in (3.8).

Similar comments can be made with regards to relaxing the assumption from section 2.3. With $\theta(t)$ as a time dependent function, it will be more challenging to obtain closed-form solutions for the ODEs that result from the method of Duffie *et al.* (2000).

Chapter 4

Fourier Methods

In conjunction with the characteristic function of a model, the family of Fourier methods allow for fast and accurate pricing of some of the most frequently traded options. This is especially useful when calibrating a model. The most commonly used Fourier methods are the Carr and Madan (1999) method and the COS method derived by Fang and Oosterlee (2008). While each method has its own advantages and disadvantages, they have both been shown to yield similarly accurate results. For this dissertation, the Carr-Madan method was used. In section 4.1, the Carr-Madan method for pricing a European call option is presented and in section 4.2, the implementation of the method is discussed.

4.1 The Carr-Madan Method

The Carr-Madan method is centered around the idea that if the Fourier transform of the option price is known, option prices can be recovered through the inverse Fourier transform. However, since the call option price as a function of strike is not square-integrable, its Fourier transform cannot be determined. Through the introduction of a dampening factor, Carr and Madan (1999) show that the call option price can be adjusted to ensure that its Fourier transform can be determined.

In the original paper, the Carr-Madan method is derived in a setting where interest rates are deterministic. This is not the case with the HHW model, and so Theorem 4.1 revises the method to include a stochastic interest rate.

Theorem 4.1. *The initial price of a European call option with strike K and expiry T is given by*

$$C(k) = \frac{e^{-\alpha k}}{\pi} Re\left\{ \int_0^\infty e^{ivk} \psi(v) dv \right\}, \tag{4.1}$$

where $k = log(K)$, $\alpha > 0$,

$$\psi(v) = \frac{\bar{\phi}_{x_T}(v - (\alpha+1)i)}{(\alpha + 1 + iv)(\alpha + iv)}$$

and $Re\{z\}$ denotes the real part of the complex number z.

Proof. By risk-neutral valuation (2.4), the initial price of a European call option with strike K and expiry T is given by

$$C(k) = \mathbb{E}\Big[N_T\big(e^{x_T} - e^k\big)\mathbb{1}_{\{x_T>k\}}\Big]. \tag{4.2}$$

Recall that changing the numeraire to the T-maturity zero coupon bond, which is denoted by $P(0,T)$, results in a change to the T-forward measure. The expectation in (4.2) can be expressed, through the relevant Radon-Nikodym derivative[1], in terms of the T-forward measure, which yields

$$C(k) = P(0,T)\mathbb{E}^T\Big[\big(e^{x_T} - e^k\big)\mathbb{1}_{\{x_T>k\}}\Big].$$

This expectation can then be written in terms of the following integral:

$$C(k) = P(0,T)\int_k^\infty (e^x - e^k)q_{x_T}^T(x)dx,$$

where $q_{x_T}^T$ is the probability density function of x_T under the T-forward measure. Notice that $C(k) \to P(0,T)\mathbb{E}^T\big[S_T\big] > 0$ as $k \to -\infty$ and hence, $C(k)$ is not square-integrable. Carr and Madan (1999) introduce a dampening factor of $e^{\alpha k}$. The dampened call price, which is clearly square-integrable, is given by

$$c(k) = e^{\alpha k}C(k) = e^{\alpha k}P(0,T)\int_k^\infty (e^x - e^k)q_{x_T}^T(x)dx. \tag{4.3}$$

The Fourier transform of the dampened call price is defined as

$$\psi(v) = \int_{-\infty}^\infty e^{ivk}c(k)dk.$$

Substituting (4.3) into $\psi(v)$ and subsequently changing the order of integration yields

$$\psi(v) = P(0,T)\int_{-\infty}^\infty q_{x_T}(x)\int_{-\infty}^x e^{x+(\alpha+iv)k} - e^{(1+\alpha+iv)k}dkdx.$$

By integrating with respect to k, $\psi(v)$ can be simplified to the following expression:

$$\psi(v) = \frac{P(0,T)}{(\alpha+1+iv)(\alpha+iv)}\int_{-\infty}^\infty e^{i(v-(\alpha+1))x}q_{x_T}^T(x)dx.$$

[1] The change between the risk-neutral and T-forward measures is defined by the following Radon-Nikodyn derivative

$$\frac{dQ^T}{dQ} = \frac{N_T}{P(0,T)}$$

.

Recognising the integral in $\psi(v)$ as the definition of the characteristic function of x_T under the T-forward measure yields

$$\psi(v) = P(0,T)\frac{\phi_{x_T}^T(v-(\alpha+1)i)}{(\alpha+1+iv)(\alpha+iv)}. \tag{4.4}$$

The characteristic function under the T-forward measure and the discounted characteristic function are related in the following way:

$$\phi_{x_T}^T(u) = \mathbb{E}^T\left[e^{iux_T}\right] = \frac{1}{P(0,T)}\mathbb{E}\left[N_T e^{iux_T}\right] = \frac{1}{P(0,T)}\bar{\phi}_{x_T}(u). \tag{4.5}$$

Substituting the result from (4.5) into (4.4) yields the expression for $\psi(v)$ in theorem 4.1. Finally, the call price can be recovered by taking the inverse Fourier transform of $\psi(v)$ and undoing the dampening, i.e.,

$$C(k) = \frac{e^{-\alpha k}}{2\pi}\int_{-\infty}^{\infty} e^{-ivk}\psi(v)dv = \frac{e^{-\alpha k}}{\pi}Re\left\{\int_0^{\infty} e^{-ivk}\psi(v)dv\right\}.$$

The last equality follows from the fact that since $C(k)$ is real, the integrand must be odd in its imaginary part and even in its real part. □

Comparing Theorem 4.1 to the result presented in the original paper, reveals that the presence of a stochastic interest rate does not change the Car-Madan method much. In fact, all that changes from the original result is that the discount factor in the expression for $\psi(v)$ is dropped and the characteristic function is replaced by its discounted version.

4.2 Implementing the Carr-Madan Method

In order to implement the Carr-Madan method, an appropriate dampening factor must be chosen and the integral in (4.1) must be evaluated.

4.2.1 Determining the Dampening Factor

For the HHW model and its affine approximations, the literature has made no mention of what an appropriate value for α is. Simons *et al.* (2004) proposed that $\alpha = 0.75$ is sufficient for most models, including the Heston model. This was therefore used as a starting point. During testing, the prices of call options were not sensitive to values of α that were around 0.75. For some parameter sets, the call option prices did start to oscillate for $\alpha < 0.1$ and $\alpha > 6$, indicating that care is needed when choosing the value for α. For all the parameter sets that were considered in this dissertation, $\alpha = 0.75$ was used without issue.

4.2.2 Evaluating the Integral

With the characteristic function of a model, the expression for $\psi(v)$ and hence the integrand in (4.1) is easily determined. However, there is usually no analytical expression for this integral. Evaluating it, involves truncating the infinite upper bound to a finite value, v_{max}. A numerical integration method is then used to approximate the truncated integral. Multiplying by the factor before the integral in (4.1), results in an approximation for the call option price.

The approximation of the truncated integral is the only source of error for the Carr-Madan method. This integral can usually be approximated to a machine-precision degree of accuracy. Carr-Madan, and more generally Fourier, prices are therefore considered to be an accurate reflection of a model's true prices.

For this dissertation, the trapezoidal integration rule was used to approximate the truncated integral. In MATLAB, this was implemented using the built-in function, `trapz`. The function requires two inputs. The first being a vector, $\mathbf{v}$, that contains values from 0 to v_{max}. This was created in an equally spaced manner with a predetermined step size of dv. The second input is the vector, $\mathbf{y}$, of the integrand evaluated at each of the points in $\mathbf{v}$ i.e.

$$\mathbf{y} = e^{ik\mathbf{v}} * \psi(\mathbf{v}).$$

Invoking trapz with $\mathbf{v}$ and $\mathbf{y}$, returns the approximated value of the integral. The accuracy can be increased by choosing a larger value for v_{max} or by decreasing dv.

Carr and Madan (1999) show that if a left-hand Riemman sum is chosen to approximate the truncated integral, the Fast Fourier transform (FFT) can be applied to efficiently price many strikes at once. However, the application of the FFT leads to an inflexibility in terms of the set of strikes that can be priced and hence, it was not considered here. In section 4.2.4, a more flexible method, that relies on MATLAB's vectorisation capabilities, is proposed.

4.2.3 Implementing the HHW Characteristic Functions

The HHW-1 characteristic function is the simplest out of the three and it is relatively easy to see that its implementation is vectorisable[2]. Due to the different possibilities for evaluating the expectation of $\sqrt{v_t}$ and the subsequent integrand in (3.8), the HHW-2D characteristic function is slightly more complex to implement. However, notice that in (3.9), u is factored out of the integral. This means that regardless of

[2] In this context, vectorisable means that the implementation of the characteristic function can accept a vector input, $\mathbf{u}$, and return a vector of the characteristic function evaluated at each element of $\mathbf{u}$. In MATLAB this is usually significantly faster than a for loop.

how the expectation is evaluated, the integral only needs to be computed once for a range of different input values, i.e., the HHW-2D characteristic function is also vectorisable. Due to this vectorisable nature, computing **y** for HHW-1 and HHW-2D cases is straightforward and efficient.

However, for the HHW-2S characteristic function, a loop is needed to compute **y**. Due to the loop and the system of ODEs that needs to be solved every time the HHW-2S characteristic function is evaluated, computing **y** takes longer than for the other two cases.

Note that for the characteristic functions that are vectorisable, MATLAB's integral function can be used to approximate the truncated integral. The advantage of this function is that it chooses the appropriate method and step size to achieve a given tolerance. However, trapezoidal integration was used so that there was a consistent approach across all the characteristic functions. Nevertheless, when it can be used, the integral function is a good way to check the results of the trapezoidal integration.

4.2.4 Speeding Up the Implementation

To speed up the execution of the Car-Madan method, the implementation was altered to simultaneously price a set of strikes. Instead of evaluating the whole integrand in (4.1) at each point in **v**, only $\psi(\mathbf{v})$ was evaluated and the results were stored. The different strikes, denoted by k_j, were then looped through and

$$\mathbf{y}_j = e^{ik_j\mathbf{v}}\psi(\mathbf{v})$$

was computed using vectorisation. For each of the strikes, `trapz` was then used to approximate the integral in (4.1) and hence determine the price. This implementation is faster because the evaluation of $\psi(\mathbf{v})$, which is what causes the inefficiency for the HHW-2S model, is only performed once for a range of different strikes.

For the HHW-2S model, parallelisation can be used to gain further speed enhancements. In particular, one simple option is that the loop that is need to compute **y**, can be replaced with a parallelised version i.e. `parfor` in MATLAB. However, this was not considered here.

Chapter 5

Monte Carlo Methods

In terms of accuracy and execution time, Monte Carlo methods are not the most efficient numerical methods for pricing options[1]. They are, however, highly flexible, in that they can used to price a wide range of claims. Aside from the HHW-1 case, where the exact characteristic function is available, Monte Carlo methods are also the simplest way of obtaining an accurate reflection of the HHW's true prices. This can then be used as a benchmark in assessing the accuracy of the deterministic and stochastic approximations.

A number of authors, including Andersen (2007) and Bégin *et al.* (2015), have documented that general discretisation schemes, such as the Euler and Milstein schemes, are not the ideal choice for simulating the Heston model. These schemes result in negative values for the variance process and are inefficient, in the sense that a large number of simulations are required to obtain an acceptable degree of accuracy. Furthermore, for some parameter sets, this inaccuracy remains even as the number of simulations becomes increasingly large.

Many schemes have been developed to address these issues. Some of these, such as the full truncation Milstein scheme, are alterations of the general schemes. Others, such as the Quadratic Exponential (QE) scheme of Andersen (2007), were specifically developed for the Heston model. The QE scheme is widely considered as being one the most efficient schemes for simulating the Heston model (Bégin *et al.*, 2015).

When benchmarking the deterministic and stochastic approximations against Monte Carlo prices, Maze (2014) used the full truncation Milstein scheme and Grzelak and Oosterlee (2011) used the QE scheme. However, the literature does not provide the details around how these schemes can be extended to the HHW model. It is also not clear how the performance of these schemes are affected when they are extended to the HHW model. In section 5.1 and 5.3, the full truncation Milstein and

[1] Authors such as Glasserman (2013) and Wilmott (2006) have noted that Monte Carlo methods are often outperformed by finite difference methods and Fourier methods, when they are available.

QE schemes are outlined. In section 5.2, a slight modification of the full truncation Milstein scheme is proposed and in section 5.4, the schemes are compared.

5.1 The Full Truncation Milstein Scheme

The details of the Milstein scheme are summarised in Theorem 5.1.

Theorem 5.1. *Let X_t be a process that is governed by*

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad X_0 > 0,$$

The solution to this SDE can be approximated by the following discrete process:

$$\hat{X}_{t+dt} = \hat{X}_t + \mu(\hat{X}_t)dt + \sigma(\hat{X}_t)\sqrt{dt}Z + \frac{1}{2}\sigma'(\hat{X}_t)dt(Z^2 - 1), \quad \hat{X}_0 = X_0,$$

where Z is a standard normal random variable.

Proof. The proof of this result can be found in Kloeden and Platen (2013). □

Applying Theorem 5.1 to each SDE of the HHW log dynamics produces

$$x_{t+dt} = x_t + \left(r_t - \frac{1}{2}v_t\right)dt + \sqrt{v_t}\sqrt{dt}Z_s, \tag{5.1}$$

$$v_{t+dt} = \max\left(v_t + k(\bar{v} - v_t)dt + \sigma\sqrt{v_t}\sqrt{dt}Z_v + \frac{1}{4}\sigma^2 dt(Z_v^2 - 1), 0\right), \tag{5.2}$$

$$r_{t+dt} = r_t + \lambda\big(\theta - r_t\big)dt + \eta\sqrt{dt}Z_r, \tag{5.3}$$

where Z_s, Z_v and Z_r are standard normal variables with the same correlation structure as the Brownian motions driving the state variables.

The presence of the max function in (5.2) is what differentiates this as the full truncation Milstein scheme. Under the Euler and Milstein schemes, the discretised variance process has a positive probability of becoming negative, regardless of whether or not the Feller condition is satisfied. This is because the Feller condition only holds for continuous time processes. Negative variances are a problem because the square root of v_t needs to be computed in (5.1) and (5.2). The inclusion of the max function prevents this issue. There are other options for fixing this issue, for example, replacing the max function with the absolute value function results in the full reflection scheme. Lord *et al.* (2010) tested various of these fixes and found that the full truncation scheme has the smallest bias.

With a given time discretisation and initial values for the state variables, (5.1) to (5.3) can be successively applied to simulate paths of the HHW model. The Cholesky decomposition is used to generate the correlated standard normal variates. In order to price claims, the evolution of the integrated short rate process

must also be simulated, i.e., Y_{t+dt} from (2.4). With r_t and r_{t+dt}, trapezoidal integration can be used to approximate Y_{t+dt}. The discount factor process, N_{t+dt}, can then be obtained by exponentiating the negative of Y_{t+dt}. Having simulated all of the required paths, the price of a claim can be computed in the usual Monte Carlo way.

5.2 Exact Simulation of Rates

Under the Vasicek model, the short rate and its integrated process can be simulated exactly. This is summarised in Theorem 5.2.

Theorem 5.2. *For the Vasicek model, the short rate and the integrated short rate can be simulated as follows:*

$$\begin{aligned} r_{t+dt} &= \mu_r(t) + \sigma_r(t) Z_r, \\ Y_{t+dt} &= \mu_Y(t) + \sigma_Y(t)\Big[\rho_{rY}(t) Z_r + \sqrt{1-\rho_{rY}^2(t)}\, Z_Y\Big], \end{aligned} \tag{5.4}$$

where

$$\begin{aligned} \mu_r(t) &= r_t + \lambda\big(\theta - r_t\big)A(t), \\ \sigma_r^2(t) &= \eta^2\left(A(t) - \frac{\lambda}{2}A^2(t)\right), \\ \mu_Y(t) &= Y_t + \theta dt + \big(r_t - \theta\big)A(t), \\ \sigma_Y^2(t) &= \frac{\eta^2}{\lambda^2}\left(dt - A(t) - \frac{\lambda}{2}A^2(t)\right), \\ \rho_{rY}(t) &= \frac{\sigma^2 A^2(t)}{2\sigma_r(t)\sigma_Y(t)}, \\ A(t) &= \frac{1}{\lambda}\left(1 - e^{-\lambda dt}\right) \end{aligned}$$

and, Z_r and Z_Y are independent standard normal random variables.

Proof. The proof of this result can be found in Brigo and Mercurio (2007). □

Replacing (5.3) with (5.4) results in a second option for simulating the HHW model. This is expected to be slightly more accurate than the full truncation Milstein scheme. In terms of the implementation, all that changes from the previous section, is that an additional independent standard normal variate needs to sampled to simulate the integrated short rate process.

5.3 The QE Scheme

Recall from Theorem 3.4, that the distribution of the variance process is known. Making it possible to simulate exactly. However, the algorithm for exact simulation, which was developed by Broadie and Kaya (2006), is inefficient and difficult to implement (Andersen, 2007). The goal of the subsequent Heston specific schemes, such as the QE scheme, was to find an approximation for the distribution of the variance process that results in fast and accurate simulation. In Theorem 5.3, the QE scheme for simulating the variance process is outlined.

Theorem 5.3. *The QE scheme for simulating the variance process is as follows:*

$$v_{t+dt} \approx \begin{cases} a\big(b + \Phi^{-1}(U_v)\big)^2 & \psi \leq \psi_c \\ \Psi^{-1}\big(U_v; p, \beta\big) & \psi > \psi_c, \end{cases}$$

where U_v is a uniform random variable, $\Phi(x)$ is the standard normal distribution function,

$$\Psi^{-1}(u) = \begin{cases} 0 & 0 \leq u \leq p \\ \frac{1}{\beta} \log \frac{1-p}{1-u} & p \leq u \leq 1, \end{cases}$$

$$\psi = \frac{S_t^2}{m_t^2}, \quad b = \left(\frac{2}{\psi} - 1 + \sqrt{\frac{2}{\psi}\left(\frac{2}{\psi} - 1\right)}\right), \quad a = \frac{m}{1+b^2}, \quad p = \frac{\psi - 1}{\psi + 1}, \quad \beta = \frac{1-p}{m}.$$

The expressions for m_t and S_t^2 are given in Theorem 2.1 and $\psi_c = 1.5$.

Proof. Building on the work of other authors, Andersen (2007) proposed that conditional on v_t, v_{t+dt} be approximated by a scaled version of squared normal distribution, i.e.,

$$v_{t+dt} \approx a\big(b + Z_v\big)^2, \tag{5.5}$$

where $a, b > 0$ and Z_v is a standard normal random variable.

The distribution of v_t is proportional to a noncentral chi-squared distribution. The intuition behind this approximation becomes clear when one considers that a standard normal random variable squared follows a central chi-squared distribution. Also, notice that simulating with (5.5), leads to positive values for v_{t+dt}.

However, Andersen (2007) found that the approximation in (5.5) only works well when v_t is large. For small values of v_t, he proposed that

$$\mathbb{P}\big(v_{t+dt} \in [x, x+dx] \mid v_t\big) \approx \big(p\delta(0) + \beta(1-p)e^{-\beta x}\big)dx, \tag{5.6}$$

where $0 \leq p \leq 1$, $\beta > 0$ and δ is the Dirac delta-function.

This approximation was formulated through the insights of the Truncated Guassian (TG) scheme, which is another Heston specific scheme that Andersen (2007)

developed. Integrating the probability in (5.6) yields a distribution function that has an analytical inverse, i.e., Ψ^{-1} in the theorem. This means that the inverse transform and uniform random variates can be used to simulate from the resulting distribution. Since the standard normal random variable in (5.5) can also be simulated through the inverse transform, (5.5) and Ψ^{-1} can be put together to form Theorem 5.3.

Through the expression for ψ, Andersen (2007) quantified exactly what it means for v_t to be large or small. With all else equal, higher values of v_t will result in lower values of ψ. The rule for switching between the approximations is based on the value of ψ in relation to ψ_c. The expressions for a, b, p, β are found by matching the relevant approximate distribution's expectation and variance to the actual expectation and variance of v_t. Further details can be found in the original paper. □

Theorem 5.3 leads an efficient method for simulating the variance process. Unfortunately, when it is used, the Euler or Milstein schemes cannot be applied to simulate the stock price and short rate. This is because the variance is simulated with uniform random variates, which means that the correct correlation structure between the state variables cannot be imparted through the Cholesky decomposition.

To solve this problem, Andersen (2007) expressed the SDE of the log stock price in terms of independent Brownian motions. This was then solved with the help of some approximations, leading to a method for imparting the correct correlation structure into the simulated stock price. In Theorem 5.4, this is outlined and extended to the Vasicek short rate process.

Theorem 5.4. *Having simulated the variance process as per Theorem 5.3, the stock price and the short rate can be simulated as follows:*

$$\begin{aligned} x_{t+dt} &\approx x_t + r_t dt + K_0 + K_1 v_t + K_2 v_{t+dt} + \sqrt{K_3 v_t}\hat{Z}_s, \\ r_{t+dt} &\approx r_t + \lambda\big(\theta - r_t\big)dt + L_1\hat{Z}_s + L_2\hat{Z}_r, \end{aligned}$$

where

$$K_0 = -\frac{\kappa\rho_{s,v}\bar{v}}{\sigma}dt, \quad K_1 = \Big(\frac{\kappa\rho_{s,v}}{\sigma} - \frac{1}{2}\Big)dt - \frac{\rho_{s,v}}{\sigma}, \quad K_2 = \frac{\rho_{s,v}}{\sigma},$$

$$K_3 = (1-\rho_{s,v}^2)dt, \quad L_1 = \eta\sqrt{dt}\frac{\rho_{s,r}}{\sqrt{1-\rho_{s,v}^2}}, \quad L_2 = \eta\sqrt{dt}\sqrt{1-\frac{\rho_{s,r}^2}{1-\rho_{s,v}^2}}$$

and, $\hat{Z}_s$, $\hat{Z}_v$ and $\hat{Z}_r$ are independent standard normal random variables.

Proof. Recall from Theorem 2.3 that the log dynamics of the HHW model were expressed, through the Cholesky decomposition, in terms of independent Brownian motions. The same approach is taken here but the order of the state vector is changed to $\mathbf{X}_t = \begin{bmatrix} v_t & x_t & r_t \end{bmatrix}^\top$. This changes the order of $\mathbf{W}_t$ and the correlation matrix $\mathbf{p}$. Furthermore, in the resulting correlation matrix, $\rho_{v,r}$ is set to zero. Defining $\hat{\mathbf{B}}_t$ as a standard 3-dimensional Brownian motion and computing $\mathbf{W}_t = \mathbf{L}\hat{\mathbf{B}}_t$ yields

$$\begin{aligned} W_t^v &= \hat{B}_t^v, \\ W_t^s &= \rho_{s,v}\hat{B}_t^v + \sqrt{1-\rho_{s,v}^2}\hat{B}_t^s, \\ W_t^r &= \frac{\rho_{s,r}}{\sqrt{1-\rho_{s,v}^2}}\hat{B}_t^s + \sqrt{1-\frac{\rho_{s,r}^2}{1-\rho_{s,v}^2}}\hat{B}_t^r. \end{aligned}$$

Replacing the correlated Brownian motions in the HHW log dynamics with the independent ones, integrating over $[t, t+dt]$ and simplifying yields

$$\begin{aligned} v_{t+dt} &= v_t + \kappa\bar{v}dt - \kappa\int_t^{t+dt} v_u du + \sigma\int_t^{t+dt}\sqrt{v_u}d\hat{B}_u^v, \\ x_{t+dt} &= x_t + \int_t^{t+dt} r_u du - \frac{1}{2}\int_t^{t+dt} v_u du + \rho_{s,v}\int_t^{t+dt}\sqrt{v_u}d\hat{B}_u^v + \sqrt{1-\rho_{s,v}^2}\int_t^{t+dt}\sqrt{v_u}d\hat{B}_u^s, \\ r_{t+dt} &= r_t + \lambda\theta dt - \int_t^{t+dt} r_u du + \eta\frac{\rho_{s,r}}{\sqrt{1-\rho_{s,v}^2}}\left(\hat{B}_{t+dt}^s - \hat{B}_t^s\right) + \sqrt{1-\frac{\rho_{s,r}^2}{1-\rho_{s,v}^2}}\left(\hat{B}_{t+dt}^r - \hat{B}_t^r\right). \end{aligned}$$

The expression for v_{t+dt} can be rearranged as follows:

$$\int_t^{t+dt}\sqrt{v_u}d\hat{B}_u^v = \frac{1}{\sigma}\left(v_{t+dt} - v_t - \kappa\bar{v}dt + \kappa\int_t^{t+dt} v_u du\right). \tag{5.7}$$

Substituting (5.7) into the expression for x_{t+dt} yields

$$\begin{aligned} x_{t+dt} = x_t + \int_t^{t+dt} r_u du + \left(\frac{\kappa\rho_{s,v}}{\sigma} - \frac{1}{2}\right)\int_t^{t+dt} v_u du + \\ \frac{\rho_{s,v}}{\sigma}\left(v_{t+dt} - v_t - \kappa\bar{v}dt\right) + \sqrt{1-\rho_{s,v}^2}\int_t^{t+dt}\sqrt{v_u}d\hat{B}_u^s. \end{aligned}$$

The following approximations can then used to simplify the remaining integrals in the expressions for x_{t+dt} and r_{t+dt}. Recognising K_0 to L_2 in Theorem 5.4 and performing the appropriate simplifications completes the proof.

$$\int_t^{t+dt} r_u du \approx r_t dt, \tag{5.8}$$

$$\int_t^{t+dt} v_u du \approx v_t dt, \tag{5.9}$$

$$\int_t^{t+dt}\sqrt{v_u}d\hat{B}_u^s \approx \sqrt{dt}\sqrt{v_t}\hat{Z}_s. \tag{5.10}$$

The approximations in (5.9) and (5.10) were proposed by Andersen (2007). They are relatively straightforward and follow from assuming that the integrands are constant over $[t, t + dt]$. The same approach was used in determining an approximation for the integral of the short rate process in (5.8). There are other more complex methods for approximating these integrals. Andersen (2007) explores some of these, but for this dissertation the simplest approximations were used. □

Note that under this specification, the integrated short rate and discount factor processes are simulated in the same way as they were for the full truncation Milstein scheme.

5.4 Comparing the Discretisation Schemes

For this experiment, $\rho_{s,r}$ and $\rho_{v,r}$ were set to zero. Call options, with varying strikes, were then priced using the discretisation schemes that were presented in this chapter. These prices were compared to the prices obtained from the HHW-1 characteristic function and the Carr-Madan method.

Tab. 5.1: Parameter sets used in assessing the discretisation schemes.

Set	S_0	T	v_0	k	σ	v_b	r_0	λ	θ	η	ρ_{sv}
A	100	{1, 10}	0.06	2.5	0.5	0.06	0.07	0.05	0.07	{0.01, 0.1}	−0.3
B	100	{1, 10}	0.05	0.3	0.6	0.05	0.07	0.05	0.07	{0.01, 0.1}	−0.3

The model specific parameter sets that were used are summarised in Table 5.1. For the Heston model, differences between the discretisation schemes became apparent when the Feller condition was not satisfied and the maturity was large (Bégin *et al.*, 2015). Recall from Chapter 2, that the Feller condition is often violated in practice, making it important for the discretisation schemes to accurately price under these conditions. The parameter sets in Table 5.1 were chosen with this in mind. The Feller condition is satisfied in parameter set A, but not in parameter set B. For each of these parameter sets, two maturities were considered. In order to assess the effect of the volatility of the short rate process, two levels for η were also considered. The various combinations of maturities and volatilities meant that a total of eight different scenarios were investigated. For the discretisation schemes, the parameters that were used are $N = 500$ and $n = 100000$. For the Carr-Madan method $\alpha = 0.75$, $v_{\max} = 150$ and $dv = 0.05$.

5.4.1 Results

For ease of comparison, the eight scenarios were grouped by maturity. Figure 5.1 contains all the scenarios where T $= 1$ and Figure 5.2 contains all the scenarios where T $= 10$. The box within the plots are zoomed-in views of the at-the-money (ATM) call option prices.

Figure 5.1 shows that when the maturity of the options are short-dated, the three discretisation schemes produced prices that were very close to the Carr-Madan prices. This is a good indicator that the HHW-1 characteristic function, Carr-Madan method and discretisation schemes were correctly implemented. The figure also shows that when the Feller condition was not satisfied, the QE scheme was slightly more accurate than the Milstein-based schemes. Furthermore, the increase in η did not seem to result in any noticeable changes to the performance of the schemes.

Figure 5.2 shows that when the maturity increased, the overall performance of the discretisation schemes decreased. However, when the Feller condition was satisfied, the three schemes still produced prices that were reasonably close to the Carr-Madan prices. On the other hand, when the Feller condition was not satisfied, the QE scheme was significantly more accurate than the Milstein-based schemes. So much so, that the price differences are clearly visible in the standard, as apposed to the zoomed, portion of the plot. Also notice that in this case, the errors of the Milstein-based schemes increased as the options moved more deeply out-the-money. Lastly, with the longer maturity, the increase in η lead to further differences between the discretisation schemes and the Carr-Madan method. This was particularly noticeable when the Feller condition was not satisfied, with the increase in η further magnifying the errors of the Milstein-based schemes.

In Figures 5.1 and 5.2, there were no real differences between the Milstein-based schemes where the short rate was approximated and exactly simulated. This indicates that the Milstein scheme is an accurate approximation for the Vasicek model.

In summary, when the Feller condition was satisfied, the three schemes performed equally well. When the Feller condition was not satisfied, the QE scheme was more accurate than the Milstein-based schemes. With the differences between them becoming increasingly more apparent as the maturity and volatility of the short rate process increased. The QE scheme is therefore recommended as the scheme of choice when simulating the HHW model. As a last remark in this chapter, note that since Maze (2014) only considered parameter sets where the Feller condition was satisfied, he would have obtained accurate results when using the full truncation Milstein scheme.

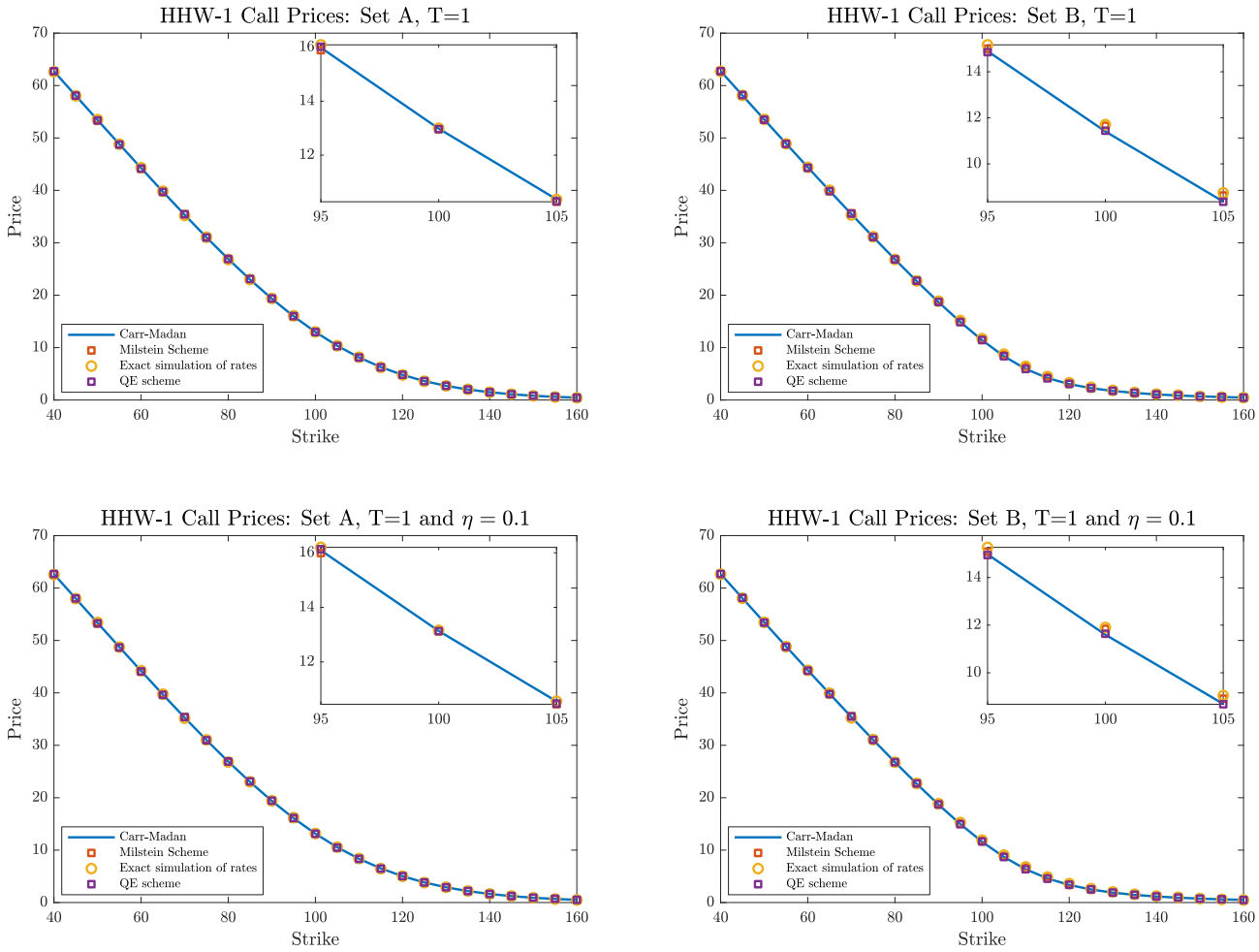


Fig. 5.1: HHW-1 call options prices with the parameter sets from Table 5.1 and T = 1.

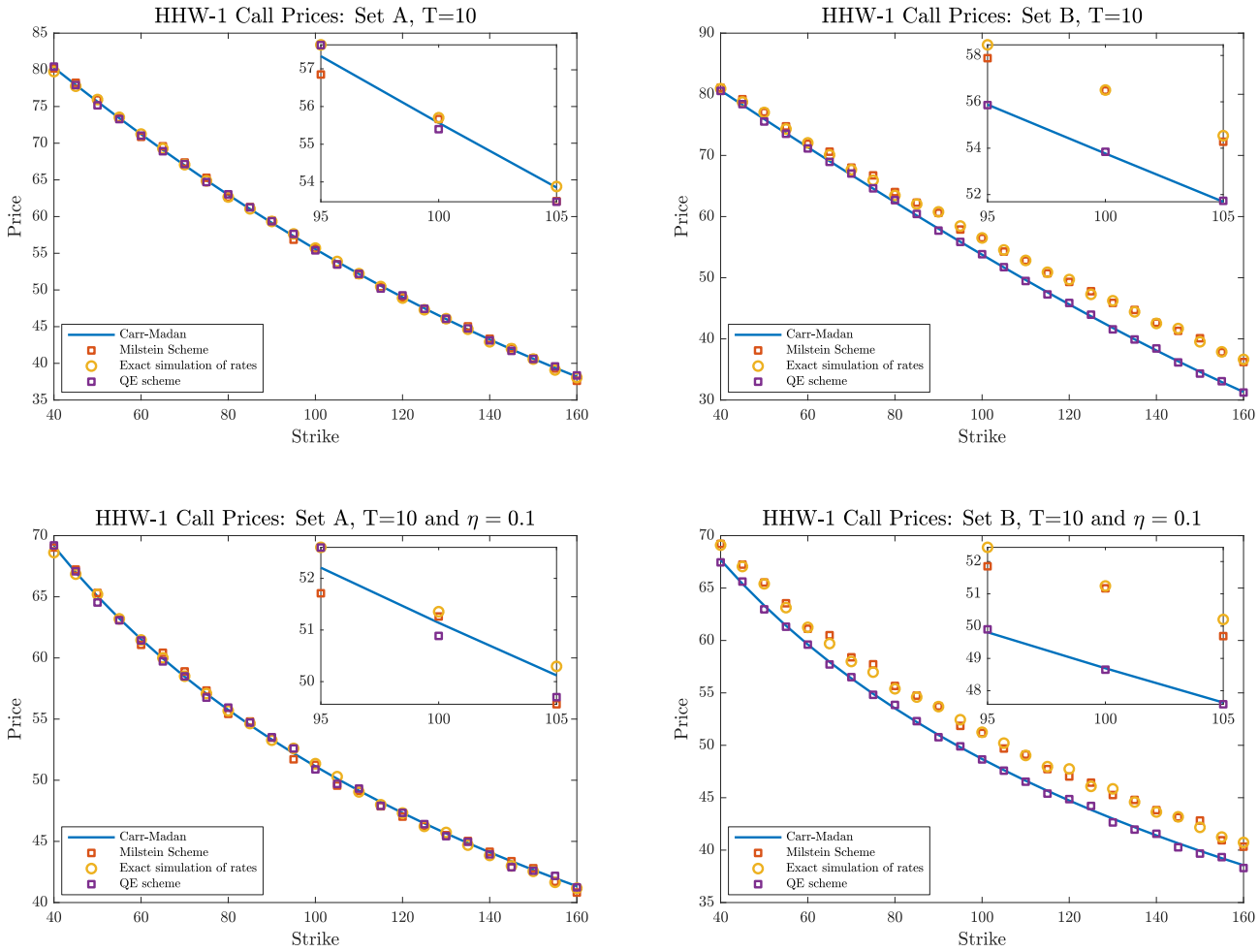


Fig. 5.2: HHW-1 call options prices with the parameter sets from Table 5.1 and T = 10.

Chapter 6

Assessing the Deterministic and Stochastic Approximations

The goal of this section is to assess the deterministic and stochastic approximations from chapter 3. Section 6.1 describes how the approximations will be compared. In section 6.2 the results are presented and analysed. Finally, in section 6.3 some interesting results that were observed during these experiments are discussed.

6.1 Methodology

In the previous section, it was shown that when pricing under the HHW-1 model, the QE scheme was the best performing discretisation scheme. Hence, It is used as a benchmark in assessing the accuracy of the deterministic and stochastic approximations. Call options were priced using the HHW-2D and HHW-2S characteristic functions and the Carr-Madan method. These prices were then compared to the prices obtained using the QE scheme, with $\rho_{v,r}$ set to zero.

Tab. 6.1: Parameter sets used in assessing the accuracy of the HHW-2D and HHW-2S characteristic functions.

Set	S_0	T	v_0	k	σ	v_b	r_0	λ	θ	η	ρ_{sv}	ρ_{sr}
A	100	$\{1, 10\}$	0.06	2.5	0.5	0.06	0.07	0.05	0.07	0.01	-0.3	$\{0.2, 0.6\}$
B	100	$\{1, 10\}$	0.05	0.3	0.6	0.05	0.07	0.05	0.07	0.01	-0.3	$\{0.2, 0.6\}$

The model specific parameter sets that were used for this experiment are given in Table 6.1. With the exception of $\rho_{s,r}$, these parameters are the same as the ones considered in the previous chapter. The various combinations of maturities and correlations means that a total of 8 scenarios were considered. For each scenario, 5 strikes representing different levels of moneyness were priced. Since the Feller condition is satisfied in parameter set A, the exact value of $\mathbb{E}\left[\sqrt{v_t}\right]$ or either of its

approximations (i.e., $\Lambda(t)$ and $\hat{\Lambda}(t)$ from Chapter 3) can be used when evaluating the characteristic functions. For parameter set B, the Feller condition is not satisfied and $\frac{8k\bar{v}}{\sigma^2} < 1$. Only the exact value of the $\mathbb{E}[\sqrt{v_t}]$ can therefore be used when evaluating the characteristic functions. However, in order to get a fair comparison between cases where the Feller condition is and is not satisfied, the exact value of the $\mathbb{E}[\sqrt{v_t}]$ was used across all the scenarios that were considered. For the Carr-Madan method, the parameters that were used are: $\alpha = 0.75$, $v_{\text{max}} = 150$ and $dv = 0.05$. For the QE scheme: $N = 1000$ and $n = 100000$.

Grzelak and Oosterlee (2011) state that they expect the stochastic approximation to be more accurate than the deterministic approximation as the volatility of the short rate process increases. In the current literature, there is a lack of evidence to support this claim. In light of this, a second experiment was performed. The same scenarios as in Table 6.1 were considered but η was set to 0.1. The parameters for the Carr-Madan method and QE scheme were also unchanged from those that where used in the first experiment.

Finally, the efficiency of the deterministic and stochastic approximations was assessed. The time that it took to price the 5 strikes under each of the scenarios was measured[1]. The results were then grouped by maturity and averaged. The grouping by maturity is due to the expectation that the run-time for the stochastic approximation should increase as the maturity increases. This is because, for a given tolerance, the time that it takes to solve the HHW-2S system of ODEs should be dependant on the maturity. The average was computed because run-time is generally quite variable and so the average is a better indicator of the actual efficiency.

6.2 Results

The results of the first experiment are presented in Table 6.3. The error percentage column is the relative error between the QE scheme and Carr-Madan prices, expressed as a percentage. The numbers in brackets indicate the standard deviation of the QE scheme prices. Across all the scenarios that were considered, the average errors for the deterministic and stochastic approximations were 0.64% and 0.63% respectively. Furthermore, for the 40 options that were priced, the deterministic approximation was more accurate 21 times. The stochastic approximation was more accurate 14 times and the two approximations yielded the same errors 5 times. The maximum errors for the deterministic and stochastic approximations were 2.38% and 2.89% respectively. For the first experiment, the results indicate that both ap-

[1] Note that all run-times were obtained from a 2.4 Ghz Intel Core i5 CPU (I5-4258U) with 8 GB ram.

proximations performed similarly well and it is difficult to make a statement regarding which of the two is more accurate.

The results of the second experiment are presented in Table 6.4. For parameter set A, with $\rho_{s,r} = 0.6$ and T $= 10$, the stochastic approximation did not produce a result. This scenario was excluded from the analysis to follow. The next section will explore possibilities as to why the stochastic approximation failed in this case.

Across the remaining scenarios, the average errors for the deterministic and stochastic approximations were 1.6% and 0.70% respectively. Compared to the first experiment, the average error for the deterministic approximation increased by a factor of about 2.4 and was much higher than the relatively unchanged error of the stochastic approximation. For the 35 options that were priced, the stochastic approximation was more accurate 25 times. The deterministic approximation was more accurate 8 times and the two approximations yielded the same errors twice. The maximum errors for the deterministic and stochastic approximations were 7.45% and 3.12% respectively. Therefore, when the volatility of the short rate process is high, the stochastic approximation seems to be a lot more accurate than the deterministic approximation, confirming the claim made by Grzelak and Oosterlee (2011).

By computing the relevant averages across Table 6.3 and 6.4, the following additional observations can be made. For both approximations, the average errors increased as $\rho_{s,r}$ increased. The average errors also increased as the options moved more deeply out-the-money. The average errors were highest for the short maturities and lowest for the long maturities. Lastly, for the first experiment, the Feller condition did not have any noticeable impact on the average errors. On the other hand, for the second experiment, the average errors were higher when the Feller condition was not satisfied. Due to the small sample size that these additional observations are based on, it is difficult to make any firm conclusions. However, these additional points should be kept in mind, especially when selecting what market data will be used for calibration.

Tab. 6.2: Average execution time grouped by maturity

Maturity	QE Scheme	HHW-2D	HHW-2S
T = 1	234.0929	0.0129	37.5207
T = 10	187.27271	0.0119	61.6487

Table 6.2 summarises the average run-times, measured in seconds, for the two approximations and the QE scheme. From the table, it is clear that the deterministic approximation is significantly faster than the stochastic approximation and both of

the approximations are significantly faster than the QE scheme. Furthermore, as expected, the run-time for the stochastic approximation increased by a factor of about 1.6, when the maturity increased. On the other hand, the run-times for the deterministic approximation remained stable as the maturity increased.

In using the deterministic approximation to calibrate the HHW model to 27 options, Maze (2014) observed that his pricing function, which priced each of the 27 options, was called on average 2600 times. For calibration, the method for simultaneously pricing multiple strikes, that was proposed in section 4.2.4, is difficult to use. This is because many different options with different maturities will be observed in the market. To price a single option, the deterministic and stochastic approximation took an average of 0.0063 and 32.2268 seconds respectively. This average was computed from 10 runs, the exact value of the expectation of $\sqrt{v_t}$ and a relatively short maturity of T $= 1$. Based on the results of Maze (2014), calibrating with the deterministic and stochastic approximations, would take approximately 442.26 and 2262321.36 seconds, respectively. In terms of daily calibration, it is clear that the deterministic approximation is a feasible method, whereas this dissertation's implementation of the stochastic approximation is not feasible.

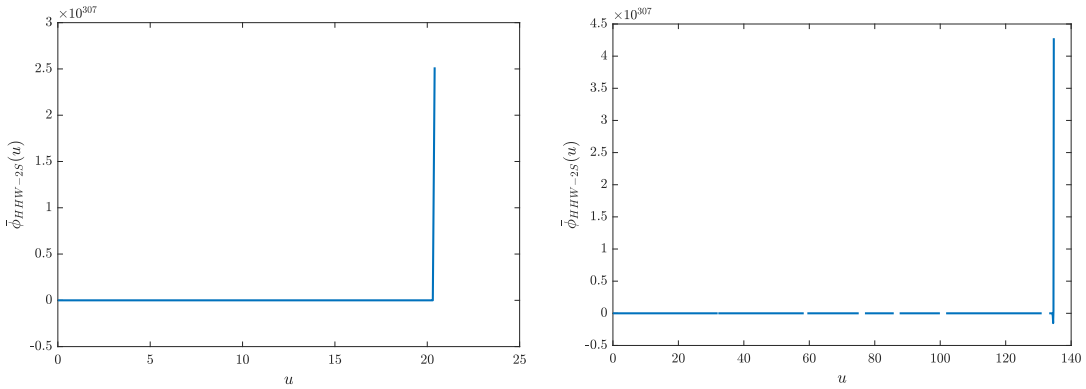
Recall from section 3.3.3 that an alternative implementation of the hypergeometric function by Brookes (2016) was used. Using the built-in MATLAB hypergeometric function implementation, the minimum time that the HHW-2S characteristic function took for a single evaluation was 2.1527 seconds. This minimum was based on 10 runs and a relatively short maturity of T $= 1$. To price the 5 strikes under one of the scenarios that were considered in the experiments above, the characteristic function needs to be evaluated 3001 times. It is therefore estimated that if the built-in MATLAB hypergeometric function was used, it would have taken at least 6460.2527 seconds to price one of the scenarios under the HHW-2S model. This is significantly longer than the average execution time of 35.507 seconds from table 6.2 and is even longer than the QE scheme's execution time. This highlights the inefficiency of the built-in MATLAB symbolic function[2]. It also illustrates how one change to the implementation can lead to significant performance gains.

Note that for the QE scheme and the Carr-Madan method, the execution time depends on the parameters that were used. In particular, for the QE scheme, N and n are the variables that impact the execution time. While for the Carr-Madan method, the variables of importance are v_{max} and dv. To ensure the best results for the accuracy comparisons, large values for N, n and v_{max} were used and a small value for dv was used. This resulted in long run-times. In reality, much more rea-

[2] Note that this could be a MATLAB specific problem and other programming languages could have more efficient options for evaluating the hypergeometric function.

sonable parameters could be used, which should decrease the run-times. However, even with these decreased run-times, the stochastic approximation is expected to be too inefficient for calibration.

6.3 Instability of the Stochastic Approximation



(a) $S_0 = 100$, $T = 10$, $v_0 = 0.06$, $\kappa = 2.5$, $\bar{v} = 0.06$, $\sigma = 0.5$, $r_0 = 0.07$, $\lambda = 0.05$, $\theta = 0.07$, $\eta = 0.1$, $\rho_{s,v} = -0.3$ and $\rho_{s,r} = 0.6$.

(b) $S_0 = 100$, $T = 10$, $v_0 = 0.06$, $\kappa = 5$, $\bar{v} = 0.06$, $\sigma = 0.3$, $r_0 = 0.07$, $\lambda = 0.05$, $\theta = 0.07$, $\eta = 0.01$, $\rho_{s,v} = -0.3$ and $\rho_{s,r} = 0.85$.

Fig. 6.1: The unstable behaviour of the HHW-2S characteristic function.

In figure 6.1a, the real part of the HHW-2S characteristic function is plotted for values of u ranging from 0 to 150. The parameter set that was used for this figure, was the one that failed to produce prices in the previous section. The figures show that for certain values of u, the implementation of the HHW-2S characteristic function did not evaluate. The characteristic function also exploded at certain points. Figure 6.1b shows one of the many other parameter sets where similar behaviour was observed.

This behaviour seems to occur when κ, $\rho_{s,r}$ and T are large. Analysing the component functions of the HHW-2S characteristic function, revealed that for these scenarios, the numerical solution to the system of ODEs is a very large number. The issue seems to arise because in evaluating the characteristic function, this large number must then be exponentiated.

Various factors that could have influenced these results were investigated. Instead of `ode23`, all the other MATLAB ODE solvers were tested. Different implementations of the hypergeometric function were also tested. However, these investigations did not affect the behaviour that was observed. Furthermore, due to the fact that only the characteristic function is plotted in these figures, any concerns

relating to the Carr-Madan dampening factor can be ruled out.

It therefore seems as if there is some sort of theoretical or numerical instability with the HHW-2S characteristic function. A situation like this was seen with the original derivation of the Heston characteristic function and was later improved upon by authors such as Albrecher *et al.* (2007).

It is difficult to verify this finding with other authors that have studied the HHW model. In Grzelak and Oosterlee (2011), results are published for parameter sets where this sort of instability was not observed. Maze (2014) and Wang (2011) only implemented the deterministic approximation, citing that the stochastic approximation is too inefficient. Their results for the deterministic approximation were matched almost exactly and the full comparison can be found in Appendix A.

Tab. 6.3: Accuracy of the HHW-2D and HHW-2S models for parameter sets A and B from table 6.1.

			T=1					T=10				
Set	p_{sr}	Strike	QE Scheme	HHW-2D	Error (%)	HHW-2S	Error (%)	QE Scheme	HHW-2D	Error (%)	HHW-2S	Error (%)
		40	62.68 (0.08)	62.71	0.06	62.71	0.06	80.07 (0.27)	80.34	0.33	80.36	0.36
		80	26.89 (0.07)	26.92	0.13	26.92	0.14	62.95 (0.26)	63.23	0.44	63.26	0.49
	20%	100	12.97 (0.05)	13.02	0.36	13.02	0.36	55.68 (0.26)	55.97	0.52	55.99	0.57
S		120	4.79 (0.04)	4.83	0.90	4.83	0.82	49.24 (0.25)	49.54	0.62	49.56	0.66
E		160	0.45 (0.01)	0.46	1.77	0.46	1.85	38.61 (0.23)	38.94	0.86	38.94	0.87
T		40	62.68 (0.08)	62.71	0.06	62.72	0.06	80.19 (0.27)	80.45	0.32	80.53	0.42
A		80	26.91 (0.07)	26.95	0.13	26.95	0.14	63.49 (0.26)	63.76	0.43	63.85	0.56
	60%	100	13.04 (0.05)	13.09	0.38	13.08	0.36	56.46 (0.26)	56.76	0.52	56.83	0.64
		120	4.86 (0.04)	4.91	1.03	4.90	0.81	50.27 (0.25)	50.58	0.62	50.63	0.72
		160	0.47 (0.01)	0.48	1.73	0.48	1.94	40.02 (0.24)	40.38	0.91	40.38	0.92
		40	62.71 (0.07)	62.76	0.07	62.76	0.07	80.73 (0.24)	80.63	0.12	80.63	0.11
		80	26.82 (0.06)	26.85	0.11	26.85	0.12	62.51 (0.24)	62.43	0.13	62.46	0.08
	20%	100	11.39 (0.05)	11.44	0.41	11.43	0.38	53.94 (0.23)	53.89	0.09	53.93	0.02
S		120	3.01 (0.03)	3.08	2.27	3.07	2.02	45.85 (0.23)	45.86	0.03	45.89	0.08
E		160	0.47 (0.02)	0.48	2.38	0.48	2.88	31.67 (0.22)	31.89	0.70	31.74	0.23
T		40	62.71 (0.07)	62.76	0.07	62.76	0.07	80.77 (0.24)	80.65	0.15	80.69	0.10
B		80	26.83 (0.06)	26.86	0.09	26.87	0.12	62.68 (0.24)	62.56	0.19	62.69	0.01
	60%	100	11.44 (0.05)	11.50	0.49	11.49	0.39	54.22 (0.24)	54.16	0.11	54.30	0.15
		120	3.07 (0.03)	3.15	2.79	3.13	2.02	46.27 (0.23)	46.35	0.18	46.42	0.33
		160	0.48 (0.02)	0.48	1.42	0.49	2.89	32.44 (0.22)	33.01	1.76	32.54	0.32

Tab. 6.4: Accuracy of the HHW-2D and HHW-2S models for parameter sets A and B from table 6.1 with $\eta = 0.1$.

			T=1					T=10				
Set	p_{sr}	Strike	QE Scheme	HHW-2D	Error (%)	HHW-2S	Error (%)	QE Scheme	HHW-2D	Error (%)	HHW-2S	Error (%)
		40	62.61 (0.08)	62.66	0.07	62.66	0.07	70.97 (0.27)	71.37	0.57	71.11	0.20
		80	26.99 (0.07)	27.02	0.14	27.03	0.18	58.57 (0.27)	59.02	0.76	58.79	0.37
	20%	100	13.40 (0.06)	13.47	0.50	13.45	0.41	54.22 (0.26)	54.70	0.88	54.49	0.49
S		120	5.33 (0.04)	5.41	1.42	5.38	0.85	50.60 (0.26)	51.11	1.01	50.91	0.61
E		160	0.61 (0.01)	0.62	1.55	0.62	1.78	44.85 (0.25)	45.39	1.22	45.21	0.82
T		40	62.62 (0.08)	62.66	0.07	62.66	0.07	74.80 (0.29)	75.25	0.60	NaN	NaN
A		80	27.27 (0.07)	27.29	0.10	27.32	0.19	64.07 (0.29)	64.62	0.86	NaN	NaN
	60%	100	14.01 (0.06)	14.10	0.66	14.06	0.39	60.25 (0.29)	60.84	0.97	NaN	NaN
		120	6.00 (0.04)	6.12	2.08	6.05	0.77	57.04 (0.28)	57.65	1.08	NaN	NaN
		160	0.84 (0.02)	0.86	2.32	0.86	2.12	51.83 (0.28)	52.50	1.29	NaN	NaN
		40	62.65 (0.07)	62.70	0.07	62.70	0.08	68.59 (0.25)	68.96	0.54	68.73	0.21
		80	26.82 (0.06)	26.84	0.06	26.87	0.16	54.86 (0.25)	55.41	1.01	55.04	0.34
	20%	100	11.79 (0.05)	11.91	1.00	11.85	0.51	50.13 (0.25)	50.74	1.23	50.34	0.43
S		120	3.57 (0.04)	3.71	4.18	3.63	1.84	46.23 (0.25)	46.90	1.44	46.48	0.53
E		160	0.55 (0.02)	0.54	1.73	0.57	2.57	40.17 (0.24)	40.86	1.73	40.43	0.65
T		40	62.66 (0.07)	62.70	0.06	62.71	0.07	70.48 (0.26)	71.32	1.19	70.28	0.28
B		80	27.00 (0.06)	26.97	0.10	27.05	0.18	57.57 (0.26)	58.85	2.23	57.27	0.53
	60%	100	12.27 (0.05)	12.51	1.93	12.33	0.51	53.08 (0.26)	54.50	2.68	52.80	0.54
		120	4.08 (0.04)	4.39	7.45	4.15	1.72	49.36 (0.26)	50.89	3.09	49.12	0.49
		160	0.68 (0.02)	0.63	7.13	0.71	3.12	43.49 (0.26)	45.14	3.80	43.36	0.31

Chapter 7

Conclusion

Three specific themes were explored as part of this dissertation. These included the theory and implementation of pricing under the deterministic and stochastic approximations, as well as the formation of an appropriate Monte Carlo discretisation scheme for benchmarking purposes. Lastly, the deterministic and stochastic approximations were assessed. A summary of the research and the key findings under each of these themes, will be discussed in the paragraphs to follow.

The presence of $\sqrt{v_t}$ in its covariance matrix means that the HHW-2 model is not affine. The deterministic and stochastic approximations are two techniques for dealing with the terms containing $\sqrt{v_t}$. As part of the deterministic approximation, the terms involving $\sqrt{v_t}$ are replaced by their expectation, while in the stochastic approximation, an approximation of $\sqrt{v_t}$ is added to the system of state variables. The approximated models that result from these techniques, i.e. the HHW-2D and HHW-2S models, are affine. The method by Duffie *et al.* (2000) for determining the discounted characteristic function of an affine model was then applied to the affine HHW-1 model and these two approximated models.

With a closed-form expression that was easily vectorisable, the HHW-1 characteristic function was the simplest out the three. The final expression for the HHW-2D characteristic function, was dependent on how $\mathbb{E}\left[\sqrt{v_t}\right]$ evaluated. If the second approximation was chosen, the characteristic function could be simplified to a closed-from expression. In the other cases, numerical integration was needed to evaluate the characteristic function. However, in all of these cases, the HHW-2D characteristic function was vectorisable. Evaluating the HHW-2S characteristic function requires a system of ODEs to be solved. In MATLAB, this system was solved with `ode23` and the hypergeometric function implementation by Brookes (2016). Due to its complexity, the HHW-2S characteristic function was not vectorisable.

In using these characteristic functions to price vanilla options, the Carr-Madan Fourier method was selected. The extension to a stochastic interest rate altered the

formulation in a sensible way, with the discount factor being dropped and the characteristic function being replaced by its discounted version. As part of this discussion, a method that was based on MATLAB's vectorisation ability, was proposed for the simultaneous pricing of many strikes.

Due to the complications in simulating the Heston model's variance process, the general Euler or Milstein discretisation schemes were not suitable choices. Drawing on the research from the Heston model, the full truncation Milstein scheme, which was the best performing adaptation of the general schemes, and the QE scheme, which was the best performing Heston specific scheme, were extended to the HHW model. These schemes, in addition to an alteration to the full truncation Milstein scheme where the interest rate is simulated exactly, were then compared to the exact HHW-1 prices. When the Feller condition was satisfied, the three schemes performed equally well. However, when the Feller condition was not satisfied and maturities were large, the QE scheme outperformed the Milstein-based schemes. For the extension to the HHW model, the effect of volatility of the short rate process was an additional consideration. It was observed that large values of η negatively impacted upon the Milstein-based schemes, while the QE scheme remained relatively unaffected.

In order to assess the approximations, two experiments were conducted with a range of challenging parameter sets. In the first experiment, both approximations were similarly accurate. In the second experiment, η was increased. This resulted in a deterioration in the performance of the deterministic approximation, while the stochastic approximation remained relatively unchanged. However, this experiment brought to light the potential instability of the stochastic approximation. In terms of efficiency, the deterministic approximation was significantly faster than the stochastic approximation and as expected, both the approximations were significantly faster than the QE scheme.

Thus, the aim of assessing the deterministic and stochastic approximations was achieved. In terms of calibration, this dissertation did not outline the specific procedure, however, as discussed previously, the efficient calibration of the HHW model is dependent on the ability to both accurately and efficiently price vanilla options. Therefore, the findings of this dissertation can be used to make inferences about the suitability of these approximations for the purposes of calibration.

In practice, when the HHW model is calibrated, the parameters of the short rate process are determined first (Grzelak and Oosterlee, 2011). This means that a sense of the size of η will be available before the stochastic volatility component is calibrated. When η is small, the deterministic approximation is a recommended choice that should result in satisfactory calibration. Furthermore, in this situation,

the additional observations from section 6.2 could be used to ensure the selection of optimal data for accurate calibration. However, when η is large, the deterministic approximation did not perform as well. As such, the deterministic approximation is not an ideal choice when the volatility of the short rate process is large.

Given the uncertain nature of the parameters that are inputted during the calibration process, the potential instability of the stochastic approximation is of particular concern. This, in addition to its inefficiency, makes it difficult to recommend the stochastic approximation as a viable option for calibration. Despite its limitations, it must be noted that in cases where η was large, the stochastic approximation exhibited promising signs of accuracy. It was also shown to be more efficient than what other authors had previously postulated. Therefore, further research into the efficiency and stability of the stochastic approximation, could result in it becoming a suitable option for situations in which the volatility of the short rate process is large.

Bibliography

Albrecher, H., Mayer, P., Schoutens, W. and Tistaert, J. (2007). The little heston trap, *Wilmott* (1): 83–92.

Andersen, L. B. (2007). Efficient simulation of the heston stochastic volatility model.

Bégin, J.-F., Bédard, M. and Gaillardetz, P. (2015). Simulating from the heston model: A gamma approximation scheme, *Monte Carlo Methods and Applications* **21**(3): 205–231.

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

Broadie, M. and Kaya, Ö. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes, *Operations research* **54**(2): 217–231.

Brookes, M. (2016). v_hypergeom1f1. http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/doc/voicebox/hypergeom1f1.html#_subfunctions.

Carr, P. and Madan, D. (1999). Option valuation using the fast fourier transform, *Journal of computational finance* **2**(4): 61–73.

Cox, J. C., Ingersoll Jr, J. E. and Ross, S. A. (2005). A theory of the term structure of interest rates, *Theory of Valuation*, World Scientific, pp. 129–164.

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

Dufresne, D. (2001). *The integrated square-root process*.

Fang, F. and Oosterlee, C. W. (2008). A novel pricing method for european options based on fourier-cosine series expansions, *SIAM Journal on Scientific Computing* **31**(2): 826–848.

Feller, W. (1951). Two singular diffusion problems, *Annals of mathematics* pp. 173–182.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient, *Proc. R. Soc. Lond. A* **121**(788): 654–673.

Glasserman, P. (2013). *Monte Carlo methods in financial engineering*, Vol. 53, Springer Science & Business Media.

Grzelak, L. A. and Oosterlee, C. W. (2011). On the heston model with stochastic interest rates, *SIAM Journal on Financial Mathematics* **2**(1): 255–286.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The review of financial studies* **6**(2): 327–343.

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

Jäckel, P. (2004). Stochastic volatility models: past, present and future, *The Best of Wilmott* **1**: 355–377.

Kloeden, P. E. and Platen, E. (2013). *Numerical solution of stochastic differential equations*, Vol. 23, Springer Science & Business Media.

Lord, R., Koekkoek, R. and Dijk, D. V. (2010). A comparison of biased simulation schemes for stochastic volatility models, *Quantitative Finance* **10**(2): 177–194.

Maze, S. (2014). Efficient implementation of the heston-hull & white model.

Patnaik, P. (1949). The non-central χ 2-and f-distribution and their applications, *Biometrika* **36**(1/2): 202–232.

Rouah, F. D. (2013). *The Heston Model and Its Extensions in Matlab and C*, John Wiley & Sons.

Simons, E., Schoutens, T. J. and TISTAERT, J. (2004). W. a perfect calibration! now what, *Wilmott Magazine, March* .

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

Wang, G. (2011). An equity and foreign exchange heston-hull-white model for variable annuities.

Wilmott, P. (2006). *Paul Wilmott on Quantitative Finance-3 Volume Set*, John Wiley & Sons.

Appendix A

Comparing the HHW-2D Implementations

As was previously discussed in Chapter 6, the implementation of the deterministic approximation was compared to that of Maze (2014) and Wang (2011). The model specific parameters that were used are $S_0 = 100$, $v_0 = 0.0175$, $\kappa = 1.5768$, $\sigma = 0.0571$, $\bar{v} = 0.0398$, $r_0 = 0.07$, $\lambda = 0.05$, $\theta = 0.07$, $\eta = 0.005$, $\rho_{s,v} = -0.5711$, $\rho_{s,r} = 0.2$ and $\rho_{r,v} = 0$. The results of this are presented in Table A.1.

Tab. A.1: Comparing the HHW-2D implementations.

	T=1			T=10		
Strike	HHW-2D	Maze (2014)	Wang (2011)	HHW-2D	Maze (2014)	Wang (2011)
50	53.3802	53.3802	53.3802	75.2871	75.2871	75.2837
55	48.7188	48.7188	48.7188	72.8989	72.8989	72.8956
60	44.0594	44.0594	44.0595	70.5437	70.5437	70.5405
65	39.4076	39.4076	39.4077	68.2258	68.2258	68.2222
70	34.7773	34.7773	34.7775	65.9492	65.9492	65.9459
75	30.1978	30.1978	30.1981	63.7175	63.7175	63.7143
80	25.7199	25.7199	25.7204	61.5335	61.5335	61.5303
85	21.4184	21.4184	21.4190	59.3999	59.3999	59.3969
90	17.3856	17.3856	17.3863	57.3186	57.3186	57.3157
95	13.7185	13.7185	13.7190	55.2912	55.2912	55.2881
100	10.4998	10.4998	10.5001	53.3190	53.3190	53.3159
105	7.7828	7.7828	7.7827	51.4027	51.4027	51.4001
110	5.5814	5.5814	5.5809	49.5429	49.5429	49.5397
115	3.8711	3.8711	3.8703	47.7396	47.7396	47.7361
120	2.5968	2.5968	2.5958	45.9928	45.9928	45.9896
125	1.6856	1.6856	1.6846	44.3021	44.3021	44.2992
130	1.0597	1.0597	1.0587	42.6670	42.6670	42.6632
135	0.6458	0.6458	0.6450	41.0868	41.0868	41.0849
140	0.3820	0.3820	0.3813	39.5605	39.5605	39.5580
145	0.2196	0.2196	0.2191	38.0873	38.0873	38.0851
150	0.1229	0.1229	0.1225	36.6660	36.6660	36.6641