# Exact Simulation of Wishart Multidimensional Stochastic Volatility Model 

Chulmin Kang* Wanmo Kang ${ }^{\dagger}$

November 19, 2018


#### Abstract

In this article, we propose an exact simulation method of the Wishart multidimensional stochastic volatility (WMSV) model, which was recently introduced by Da Fonseca et al. [14. Our method is based on analysis of the conditional characteristic function of the logprice given volatility level. In particular, we found an explicit expression for the conditional characteristic function for the Heston model. We perform numerical experiments to demonstrate the performance and accuracy of our method. As a result of numerical experiments, it is shown that our new method is much faster and reliable than Euler discretization method.


Keywords: Wishart processes, stochastic volatility, Monte-Carlo method, exact simulation

## 1 Introduction

Ever since the Heston's stochastic volatility model [24] was introduced in 1993, it has been the most important and widely used model among stochastic volatility models. Its popularity relies on the clear financial interpretation of parameters and computational tractability of the model. Heston [24] found the characteristic function of logarithmic asset price in a closed-form and showed that European call options can be priced by inverting the characteristic function.

Despite its popularity, extensive empirical research has documented limitations of the Heston model [4, 5, 10, 11, 13]. The most critical deficiency of the model is that it can not generate realistic term structure of volatility smiles; Heston model provides too flat implied volatility surface to capture reality. But empirical studies revealed that the implied volatility curve for a short maturity has rather steep slope and it is convex, and that for a long maturity tends to be linear [11, 13. Consequently, much effort has been focused on generalizing the Heston model so as to accommodate such stylized facts. There are two approaches to generalize the Heston model. The first one is to add jumps in the dynamics of stock return, the volatility, or both of them [4, [5, 16]. And the other stream has been investigating the multifactor nature of the implied volatility [5, 11, 14, 16].

Among multifactor stochastic volatility models, the Wishart multidimensional stochastic volatility (WMSV) model is the most flexible one, and it matches the term structure of implied volatilities well [14, 13]. Term structure of the realized volatilities in this model is described by a positive semidefinite matrix valued stochastic process, namely the Wishart process which was developed by Bru [8] and introduced in the financial literature by Gourieroux and Sufana [22]. The dependence between the asset price and the volatility factor is also characterized by a square matrix. Such a matrix specification of the model make it possible to capture the

[^0]stylized facts observed in the option markets. The model can fit both the long-term volatility level and the short-term skew at the same time. In addition, the model exhibits the stochastic leverage effect, and it makes the model adequate to deal with a stochastic skew effect. Even with these flexible parameter specifications, it is still analytically tractable; the problem of pricing European vanilla options can be handled through the transform methods of Duffie et al. [16] and the FFT methods of Carr and Madan (9).

Even though analytic aspects of the WMSV model are well explored [6, 14, 13, 18, there are only few studies on the simulation of the model. Gauthier and Possamaï [19] proposed some discretization schemes beyond the crude Euler scheme, but their schemes are less satisfactory in terms of accuracy. Their schemes entails severe bias errors, and its accuracy is sensitive to the model parameters.

In this paper, we propose an exact simulation method of the WMSV model, which does not suffer from discretization bias. Our new simulation method for the WMSV model is motivated by the exact simulation method of Broadie and Kaya [7] for the Heston model. Their key observation was that the Heston model can be sampled exactly, provided that the endpoints and the integral of variance process are sampled exactly. For this purpose, they derived the conditional characteristic function for the integral of the variance process up to the endpoint conditional on its endpoints, and they used it to sample the integral by Fourier inversion techniques. We take a similar, but rather direct approach to generate a sample from the WMSV model. We first sample the terminal value for the volatility factor process. Then we use the Fourier inversion techniques to generate the log-price of the stock conditional on the given terminal value of the volatility factor.

In order to apply the Fourier inversion techniques, it is necessary to find the relevant characteristic function. In our case, it is the conditional characteristic function of the log-price given the terminal value of volatility factor. We prove that it can be obtained by solving a certain system of ordinary differential equations. In particular, we provide an explicit formula of the conditional characteristic function for the Heston model. In general, the system of ordinary differential equations does not admit a closed-form solution due to non-commutativity of matrix multiplications, but it can be efficiently solved by numerical methods.

The rest of this paper is organized as follows. Section 2 introduces the WMSV model specifications and derive the conditional characteristic function of the log-price. Section 3 provides a brief review on the Broadie-Kaya's exact simulation method of the Heston model. Section 4 presents our exact simulation method for the WMSV model in detail. In Section 區, we give some numerical results and compare our method with a standard discretization method and Broadie-Kaya's method. Section 6 conclude the paper. Some detailed derivations are deferred to appendices.

## 2 The Wishart Multidimensional Stochastic Volatility Model

Within the WMSV model, the dynamics of asset price, under the risk neutral measure, is described by

$$
\begin{equation*}
\frac{d S_{t}}{S_{t}}=r d t+\operatorname{tr}\left[\sqrt{X_{t}}\left(d W_{t} R^{\top}+d Z_{t} \sqrt{I_{d}-R R^{\top}}\right)\right], \quad S_{0}=s>0, \tag{1}
\end{equation*}
$$

where $\operatorname{tr}$ is the trace operator, $r$ is a constant which represents the risk neutral drift, $W$ and $Z$ are independent $d \times d$ standard matrix Brownian motions, i.e., all the entries of $W$ and $Z$ are independent standard 1-dimensional Brownian motions. In this specification, the volatility is determined by the $d \times d$ symmetric positive semidefinite matrix-valued process $X_{t}$. In the followings, we denote by $S_{d}^{++}\left(S_{d}^{+}\right)$the set of symmetric positive (semi)definite matrices. The $d \times$
$d$ matrix $R$ specifies the correlation between asset price and volatility factors, and it determines the skewness of the distribution of the return.

The volatility factor process is assumed to be a Wishart process which solves the equation

$$
\begin{equation*}
d X_{t}=\left(\delta \Sigma^{\top} \Sigma+H X_{t}+X_{t} H^{\top}\right) d t+\sqrt{X_{t}} d W_{t} \Sigma+\Sigma^{\top} d W_{t}^{\top} \sqrt{X_{t}}, \quad X_{0}=x \in S_{d}^{+} \tag{2}
\end{equation*}
$$

where $\Sigma, H$ are $d \times d$ matrices, and $\delta \geq d-1$. The parameter restriction $\delta \geq d-1$ ensures the existence of the unique weak solution of (2) [12]. The Wishart process is a matrix analog of the square-root mean-reverting process. In order to grant the typical mean-reverting feature of the volatility, the matrix $H$ is assumed to be negative definite.

Throughout the paper, we express the asset price in terms of the $\log$-price $Y_{t}=\log \left(S_{t}\right)$, so that (1) becomes

$$
\begin{equation*}
d Y_{t}=\left(r-\frac{1}{2} \operatorname{tr}\left[X_{t}\right]\right) d t+\operatorname{tr}\left[\sqrt{X_{t}}\left(d W_{t} R^{\top}+d Z_{t} \sqrt{I_{d}-R R^{\top}}\right)\right], \quad Y_{0}=y \tag{3}
\end{equation*}
$$

We first review the affine transform formula for the WMSV model [14], and we derive the conditional Laplace transform of log-price $Y_{T}$ given the terminal volatility $X_{T}$ using the affine transform formula and the change of measure techniques.

### 2.1 Laplace Transform of $Y_{T}$

Due to the affine nature of the WMSV model, the Laplace transform of the log-price process is exponentially affine in the initial values $(x, y)$. In particular, the Laplace transform is of the following form.

Proposition 2.1 (Da Fonseca et al. [14]). The Laplace transform of the log-price $Y_{T}$ is given by

$$
\begin{equation*}
\mathbb{E}_{x, y}\left[e^{-u Y_{T}}\right]=e^{-\phi(0, u)-\operatorname{tr}[\psi(0, u) x]-u y}, \quad \text { if LHS is finite, } \tag{4}
\end{equation*}
$$

where $(\phi, \psi)$ is the solution of

$$
\left\{\begin{align*}
\partial_{t} \psi(t, u)= & 2 \psi(t, u) \Sigma^{\top} \Sigma \psi(t, u)  \tag{5}\\
& -\left(H^{\top}-u R \Sigma\right) \psi(t, u)-\psi(t, u)\left(H-u \Sigma^{\top} R^{\top}\right)+\frac{u(u+1)}{2} I_{d}, \\
\partial_{t} \phi(t, u)= & -\delta \operatorname{tr}\left[\psi(t, u) \Sigma^{\top} \Sigma\right]-u r,
\end{align*}\right.
$$

for $0 \leq t \leq T$, with the terminal value $\psi(T, u)=0$ and $\phi(T, u)=0$.
Remark 2.2. In this paper, we take the equations (5) as backward equations for notational convenience. With this backward equations, the conditional Laplace transform given $\mathcal{F}_{t}$ can be written as

$$
\mathbb{E}_{x, y}\left[e^{-u Y_{T}} \mid \mathcal{F}_{t}\right]=\mathbb{E}_{X_{t}, Y_{t}}\left[e^{-u Y_{T-t}}\right]=e^{-\tilde{\phi}(0, u)-\operatorname{tr}\left[\tilde{\psi}(0, u) X_{t}\right]-u Y_{t}}
$$

where $(\tilde{\phi}, \tilde{\psi})$ is the solution of (5) with $\tilde{\psi}(T-t, u)=0$ and $\tilde{\phi}(T-t, u)=0$. Notice that $\psi(\cdot+t, u)=\tilde{\psi}(\cdot, u)$ and $\phi(\cdot+t, u)=\tilde{\phi}(\cdot, u)$ by the uniqueness of the solution. So we have $\tilde{\psi}(0, u)=\psi(t, u)$ and $\tilde{\phi}(0)=\phi(t)$. Hence we have

$$
\begin{equation*}
\mathbb{E}_{x, y}\left[e^{-u Y_{T}} \mid \mathcal{F}_{t}\right]=e^{-\phi(t, u)-\operatorname{tr}\left[\psi(t, u) X_{t}\right]-u Y_{t}} . \tag{6}
\end{equation*}
$$

### 2.2 Conditional Laplace Transform of $Y_{T}$ given $X_{T}$

The aim of this subsection is to find the conditional Laplace transform of $Y_{T}$ given $X_{T}=x_{T} \in$ $S_{d}^{++}$. We apply the affine transform formula (6) and the change of measure technique (e.g. see Theorem XI.3.2 of (30) to find the conditional Laplace transform. From this section, we assume that $\Sigma$ is nonsingular, and $\delta>d-1$ to ensure the nonsingularity of Wishart process $X_{t}$.

To state and prove the main result of this section, it is necessary to recall the definitions of the noncentral Wishart distribution and some multivariate special functions. We collect them in Appendix A.1.
Theorem 2.3. The conditional Laplace transform of log-price $Y_{T}$ given $X_{T}=x_{T} \in S_{d}^{++}$ satisfies

$$
\begin{align*}
& \mathbb{E}_{x, y}\left[e^{-u Y_{T}} \mid X_{T}=x_{T}\right]=\left(\frac{\operatorname{det}[V(0,0)]}{\operatorname{det}[V(0, u)]}\right)^{\delta / 2} \exp \{-\phi(0, u)-u y\} \\
& \quad \times \exp \left\{-\frac{1}{2} \operatorname{tr}\left[\left(2 \psi(0, u)+\Psi(0, u) V(0, u)^{-1} \Psi(0, u)^{\top}-\Psi(0,0) V(0,0)^{-1} \Psi(0,0)^{\top}\right) x\right]\right\} \\
& \quad \times \exp \left\{-\frac{1}{2} \operatorname{tr}\left[\left(V(0, u)^{-1}-V(0,0)^{-1}\right) x_{T}\right]\right\}  \tag{7}\\
& \quad \times \frac{{ }_{0} F_{1}\left(\frac{1}{2} \delta ; \frac{1}{4} V(0, u)^{-1} \Psi(0, u)^{\top} x \Psi(0, u) V(0, u)^{-1} x_{T}\right)}{{ }_{0} F_{1}\left(\frac{1}{2} \delta ; \frac{1}{4} V(0,0)^{-1} \Psi(0,0)^{\top} x \Psi(0,0) V(0,0)^{-1} x_{T}\right)},
\end{align*}
$$

where ${ }_{0} F_{1}$ is the hypergeometric function of matrix argument defined in Appendix A.1, the matrix-valued functions $\psi, \Psi, V$, and the real-valued function $\phi$ are the solution of the system of ordinary differential equations:

$$
\left\{\begin{align*}
\partial_{t} \psi(t, u)= & 2 \psi(t, u) \Sigma^{\top} \Sigma \psi(t, u)  \tag{8}\\
& -\left(H^{\top}-u R \Sigma\right) \psi(t, u)-\psi(t, u)\left(H-u \Sigma^{\top} R^{\top}\right)+\frac{u(u+1)}{2} I_{d}, \\
\partial_{t} \phi(t, u)= & -\delta \operatorname{tr}\left[\psi(s, u) \Sigma^{\top} \Sigma\right]-u r, \\
\partial_{t} \Psi(t, u)= & -\left(H^{\top}-u R \Sigma-2 \psi(t, u) \Sigma^{\top} \Sigma\right) \Psi(t, u), \\
\partial_{t} V(t, u)= & -\Psi(t, u)^{\top} \Sigma^{\top} \Sigma \Psi(t, u),
\end{align*}\right.
$$

for $0 \leq t \leq T$, with terminal values $\psi(T, u)=V(T, u)=0, \Psi(T, u)=I_{d}$, and $\phi(T, u)=0$.
Proof. Using the affine transform formula (6), we define a positive martingale

$$
\begin{aligned}
Z_{t} & =\mathbb{E}_{x, y}\left[\exp \left\{-u Y_{T}+\phi(0, u)+\operatorname{tr}[\psi(0, u) x]+u y\right\} \mid \mathcal{F}_{t}\right] \\
& =\exp \left\{-\phi(t, u)+\phi(0, u)-\operatorname{tr}\left[\psi(t, u) X_{t}\right]+\operatorname{tr}[\psi(0, u) x]-u\left(Y_{t}-y\right)\right\},
\end{aligned}
$$

for $0 \leq t \leq T$, where $(\phi, \psi)$ is the solution of the equations (5). Since $Z_{0}=1$ and $Z_{t}>0$ for all $0 \leq t \leq T$, it can be used as a Radon-Nikodym derivative process. We define an equivalent measure $\tilde{\mathbb{P}}$ by

$$
\frac{d \tilde{\mathbb{P}}}{d \mathbb{P}}=Z_{T}, \quad \text { on } \quad \mathcal{F}_{T}
$$

In order to apply Girsanov theorem, we need to find a martingale $\left(M_{t}\right)_{0 \leq t \leq T}$ such that $Z_{t}=$ $\mathcal{E}(M)_{t}$. We apply the integration by parts formula to have

$$
\begin{aligned}
& \operatorname{tr}\left[\psi(t, u) X_{t}\right]-\operatorname{tr}[\psi(0, u) x]=\int_{0}^{t} \operatorname{tr}\left[\partial_{s} \psi(s, u) X_{s}\right] d s+\int_{0}^{t} \operatorname{tr}\left[\psi(s, u) d X_{s}\right] \\
& = \\
& \quad \int_{0}^{t} \operatorname{tr}\left[\left(\partial_{s} \psi(s, u)+H^{\top} \psi(s, u)+\psi(s, u) H\right) X_{s}\right] d s \\
& \quad+\int_{0}^{t} \delta \operatorname{tr}\left[\psi(s, u) \Sigma^{\top} \Sigma\right] d s+\int_{0}^{t} \operatorname{tr}\left[2 \Sigma \psi(s, u) \sqrt{X_{s}} d W_{s}\right]
\end{aligned}
$$

and

$$
u\left(Y_{t}-y\right)=\int_{0}^{t}\left(u r-\frac{1}{2} \operatorname{tr}\left[u X_{s}\right]\right) d s+\int_{0}^{t} \operatorname{tr}\left[u R^{\top} \sqrt{X_{s}} d W_{s}\right]+\int_{0}^{t} \operatorname{tr}\left[u \sqrt{I_{d}-R R^{\top}} \sqrt{X_{s}} d Z_{s}\right]
$$

Since $(\phi, \psi)$ satisfies the equations (5),

$$
\begin{aligned}
&-\phi(t, u)+\phi(0, u)-\operatorname{tr}\left[\psi(t, u) X_{t}\right]+\operatorname{tr}[\psi(0, u) x]-u\left(Y_{t}-y\right) \\
&=-\int_{0}^{t} \operatorname{tr}\left[\left(2 \Sigma \psi(s, u)+u R^{\top}\right) \sqrt{X_{s}} d W_{s}\right]-\int_{0}^{t} \operatorname{tr}\left[u \sqrt{I_{d}-R R^{\top}} \sqrt{X_{s}} d Z_{s}\right] \\
&-\frac{1}{2} \int_{0}^{t} \operatorname{tr}\left[\left(4 \psi(s, u) \Sigma^{\top} \Sigma \psi(s, u)+2 u R \Sigma \psi(s, u)+2 u \psi(s, u) \Sigma^{\top} R^{\top}+u^{2} I_{d}\right) X_{s}\right] d s .
\end{aligned}
$$

Set

$$
M=-\int_{0} \operatorname{tr}\left[\left(2 \Sigma \psi(s, u)+u R^{\top}\right) \sqrt{X_{s}} d W_{s}\right]-\int_{0} \operatorname{tr}\left[u \sqrt{I_{d}-R R^{\top}} \sqrt{X_{s}} d Z_{s}\right]
$$

Then

$$
\langle M\rangle_{t}=\int_{0}^{t} \operatorname{tr}\left[\left(4 \psi(s, u) \Sigma^{\top} \Sigma \psi(s, u)+2 u R \Sigma \psi(s, u)+2 u \psi(s, u) \Sigma^{\top} R^{\top}+u^{2} I_{d}\right) X_{s}\right] d s
$$

Therefore, we have $Z_{t}=\mathcal{E}(M)_{t}, 0 \leq t \leq T$. By Girsanov theorem, the dynamics of $X$ is as follows

$$
d X_{t}=\left(\delta \Sigma^{\top}+H(t, u) X_{t}+X_{t} H(t, u)^{\top}\right) d t+\sqrt{X_{t}} d \tilde{W}_{t} \Sigma+\Sigma^{\top} d \tilde{W}_{t}^{\top} \sqrt{X_{t}}
$$

where $H(t, u)=H-u \Sigma^{\top} R^{\top}-2 \Sigma^{\top} \Sigma \psi(t, u)$, and $\tilde{W}$ is a $d \times d$ matrix Brownian motion under $\tilde{\mathbb{P}}$. Therefore $X$ is a Wishart process with time-varying linear drift in the sense of Appendix A. 2.

According to Proposition A. 6 in Appendix A.2, under $\tilde{\mathbb{P}}, X_{T}$ has the noncentral Wishart distribution $\mathcal{W}_{d}\left(\delta, V(0, u), V(0, u)^{-1} \Psi(0, u)^{\top} x \Psi(0, u)\right)$, and it has the transition density from time 0 to $T$ :

$$
\begin{align*}
p_{0, T}\left(x, x_{T} ; u\right)= & \frac{\left(\operatorname{det}\left[x_{T}\right]\right)^{(\delta-d-1) / 2}}{2^{d \delta / 2} \Gamma_{d}\left(\frac{1}{2} \delta\right)(\operatorname{det}[V(0, u)])^{\delta / 2}} \exp \left\{-\frac{1}{2} \operatorname{tr}\left[V(0, u)^{-1}\left(x_{T}+\Psi(0, u)^{\top} x \Psi(0, u)\right)\right]\right\} \\
& \times{ }_{0} F_{1}\left(\frac{1}{2} \delta ; \frac{1}{4} V(0, u)^{-1} \Psi(0, u)^{\top} x \Psi(0, u) V(0, u)^{-1} x_{T}\right) \tag{9}
\end{align*}
$$

where $\Psi(t, u)$ and $V(t, u)$ solve the corresponding equations in (8). Notice that the transition density of $X$ under $\mathbb{P}$ can be obtained by taking $u=0$ because $H(t, 0)=H$ for all $0 \leq t \leq T$.

Now we are ready to compute the conditional Laplace transform of $Y_{T}$ given $X_{T}=x_{T}$. Observe that

$$
\begin{aligned}
\int_{S_{d}^{++}} & \mathbb{E}_{x, y}\left[e^{-u Y_{T}} \mid X_{T}=x_{T}\right] f\left(x_{T}\right) p_{0, T}\left(x, x_{T} ; 0\right) d x_{T}=\mathbb{E}_{x, y}\left[e^{-u Y_{T}} f\left(X_{T}\right)\right] \\
& =e^{-\phi(0, u)-\operatorname{tr}[\psi(0, u) x]-u y} \mathbb{E}_{x, y}\left[Z_{T} f\left(X_{T}\right)\right]=e^{-\phi(0, u)-\operatorname{tr}[\psi(0, u) x]-u y} \tilde{\mathbb{E}}_{x, y}\left[f\left(X_{T}\right)\right] \\
& =e^{-\phi(0, u)-\operatorname{tr}[\psi(0, u) x]-u y} \int_{S_{d}^{++}} f\left(x_{T}\right) p_{0, T}\left(x, x_{T} ; u\right) d x_{T}
\end{aligned}
$$

for all nonnegative measurable function $f$ on $S_{d}^{++}$. Hence the conditional Laplace transform satisfies

$$
\mathbb{E}_{x, y}\left[e^{-u Y_{T}} \mid X_{T}=x_{T}\right]=e^{-\phi(0, u)-\operatorname{tr}[\psi(0, u) x]-u y} \frac{p_{0, T}\left(x, x_{T} ; u\right)}{p_{0, T}\left(x, x_{T} ; 0\right)}
$$

By substituting (9) into the above equation, we complete the proof.

### 2.3 Conditional Laplace Transform for Heston Model

If the volatility has a single factor, the WMSV model reduces to the classical Heston's stochastic volatility model [24]. Therefore the analysis in the previous subsection can be readily applied to the Heston model. Moreover, (8) admits a closed-form solution, and it opens possibility of further analysis of the conditional Laplace transform for the Heston model.

In the Heston model, the dynamics of $\log$-price process $Y$ and variance process $X$, under the risk neutral measure, are described by the following system of stochastic differential equations:

$$
\left\{\begin{array}{l}
d X_{t}=\kappa\left(\theta-X_{t}\right) d t+\sigma \sqrt{X_{t}} d W_{t},  \tag{10}\\
d Y_{t}=\left(r-\frac{1}{2} X_{t}\right) d t+\sqrt{X_{t}}\left[\rho d W_{t}+\sqrt{1-\rho^{2}} d Z_{t}\right]
\end{array}\right.
$$

with initial value $X_{0}=x \geq 0$ and $Y_{0}=y \in \mathbb{R}$. The second equation gives the dynamics of the log-price process: the stock price $S_{t}$ at time $t$ is given by $S_{t}=e^{Y_{t}}, r$ is the risk-neutral drift and $\sqrt{X_{t}}$ is the volatility. The first equation describes the dynamics of the stochastic variance process. The parameter $\kappa>0$ determines the speed of the mean reversion, $\theta>0$ represents the long-run mean of the variance process, and $\sigma>0$ is the volatility of the variance process. $W$ and $Z$ are independent standard 1-dimensional Brownian motions.

The parameters of the WMSV model and those of the Heston model are related in the following way:

$$
\kappa \theta=\delta \Sigma^{\top} \Sigma=\delta \Sigma^{2}, \quad \sigma=2 \Sigma, \quad \kappa=-2 H, \quad \rho=R .
$$

And (8) are reduced to

$$
\left\{\begin{align*}
\partial_{t} \psi(t, u) & =\frac{1}{2} \sigma^{2} \psi(t, u)+(\kappa+u \sigma \rho) \psi(t, u)+\frac{1}{2} u(u+1),  \tag{11}\\
\partial_{t} \phi(t, u) & =-\kappa \theta \psi(t, u)-u r, \\
\partial_{t} \Psi(t, u) & =\frac{1}{2}\left(\kappa+u \sigma \rho+\sigma^{2} \psi(t, u)\right) \Psi(t, u), \\
\partial_{t} V(t, u) & =-\frac{1}{4} \sigma^{2} \Psi(t, u)^{2},
\end{align*}\right.
$$

with terminal values $\psi(T, u)=\phi(T, u)=V(T, u)=0$ and $\Psi(T, u)=1$.
Through a straightforward calculations, we can derive the conditional Laplace transform for the Heston model. The detailed calculation is given in Appendix A.3.

Corollary 2.4. For the Heston model, the conditional Laplace transform of $Y_{T}$ given $X_{T}=$ $x_{T}>0$ satisfies

$$
\begin{align*}
& \mathbb{E}_{x, y}[ \left.e^{-u Y_{T}} \mid X_{T}=x_{T}\right] \\
&= \frac{\eta(u)\left(1-e^{-\kappa T}\right)}{\kappa\left(1-e^{-\eta(u) T}\right)} \exp \left\{-u\left(y+\left(r-\frac{\kappa \theta \rho}{\sigma}\right) T\right)-\frac{1}{2}(\eta(u)-\kappa) T\right\} \\
& \times \exp \left\{-\frac{1}{\sigma^{2}}\left(\frac{\eta(u)\left(1+e^{-\eta(u) T}\right)-(\kappa+u \sigma \rho)\left(1-e^{-\eta(u) T}\right)}{1-e^{-\eta(u) T}}-\frac{2 \kappa e^{-\kappa T}}{1-e^{-\kappa T}}\right) x\right\}  \tag{12}\\
&\left.\quad \times \exp \left\{-\frac{1}{\sigma^{2}} \frac{\eta(u)\left(1+e^{-\eta(u) T}\right)+(\kappa u \sigma \rho)\left(1-e^{-\eta(u) T}\right)}{1-e^{-\eta(u) T}}-\frac{2 \kappa}{1-e^{-\kappa T}}\right) x_{T}\right\} \\
& \quad \times \frac{I_{0.5 \delta-1}\left[\sqrt{x x_{T}} \frac{4 \eta(u) e^{-0.5 \eta(u) T}}{\sigma^{2}\left(1-e^{-\eta(u) T)}\right.}\right]}{I_{0.5 \delta-1}\left[\sqrt{x x_{T}} \frac{4 e^{-0.5 \kappa T}}{\sigma^{2}\left(1-e^{-\kappa T}\right)}\right]},
\end{align*}
$$

where $I_{\nu}(\cdot)$ is the modified Bessel function of the first kind, and $\eta(u)=\sqrt{(\kappa+u \sigma \rho)^{2}-\sigma^{2} u(u+1)}$.

## 3 Review on Broadie and Kaya Method

Before going into our exact simulation method, we briefly review the Broadie and Kaya's exact simulation method [7] of the Heston model. The simulation method, they devised, motivated our research and it contains important idea of Fourier inversion techniques.

Since the system (10) of the stochastic differential equations does not admit a closed-form solution and there is no elementary way to simulate $X_{T}$ and $Y_{T}$ exactly. Broadie and Kaya proposed an exact sampling scheme which uses the Fourier inversion techniques [7]. From (10),

$$
\begin{aligned}
Y_{T} & =y+r T-\frac{1}{2} \int_{0}^{T} X_{s} d s+\sqrt{1-\rho^{2}} \int_{0}^{T} \sqrt{X_{s}} d Z_{s}+\rho \int_{0}^{T} \sqrt{X_{s}} d W_{s} \\
& =y+r T-\frac{1}{2} \int_{0}^{T} X_{s} d s+\sqrt{1-\rho^{2}} \int_{0}^{T} \sqrt{X_{s}} d Z_{s}+\frac{\rho}{\sigma}\left(X_{T}-x-\kappa \theta T+\kappa \int_{0}^{T} X_{s} d s\right) \\
& =y+\frac{\rho}{\sigma}\left(X_{T}-x\right)+\left(r-\frac{\kappa \theta \rho}{\sigma}\right) T+\left(\frac{\rho \kappa}{\sigma}-\frac{1}{2}\right) \int_{0}^{T} X_{s} d s+\sqrt{1-\rho^{2}} \int_{0}^{T} \sqrt{X_{s}} d Z_{s}
\end{aligned}
$$

Recall that $Z$ is independent of $W$. Consequently, $Z$ is independent of the sigma field $\mathcal{F}_{T}^{X}=$ $\sigma\left(X_{t}: 0 \leq t \leq T\right)$. So, $\mathcal{F}_{T}^{X}$-conditional distribution of $\int_{0}^{T} \sqrt{X_{s}} d Z_{s}$ is normal with mean 0 . And its conditional variance can be computed by the Itô's isometry:

$$
\operatorname{var}_{x, y}\left(\int_{0}^{T} \sqrt{X_{s}} d Z_{s} \mid \mathcal{F}_{T}^{X}\right)=\mathbb{E}_{x, y}\left[\left(\int_{0}^{T} \sqrt{X_{s}} d Z_{s}\right)^{2} \mid \mathcal{F}_{T}^{X}\right]=\mathbb{E}_{x, y}\left[\int_{0}^{T} X_{s} d s \mid \mathcal{F}_{T}^{X}\right]=\int_{0}^{T} X_{s} d s
$$

These observations gives the following exact sampling scheme of the state variables $\left(X_{T}, Y_{T}\right)$ :
Algorithm 1 (Broadie and Kaya [7]). This algorithm generates the state variables $X_{T}$ and $Y_{T}$ of the Heston model.

Step (1) Generate a sample from the distribution of $X_{T}$
Step (2) Generate a sample from the conditional distribution of $I=\int_{0}^{T} X_{t} d t$ given $X_{T}$
Step (3) Generate a standard normal random number $Z$
Step (4) Set $Y_{T}=y+\frac{\rho}{\sigma}\left(X_{T}-x\right)+\left(r-\frac{\kappa \theta \rho}{\sigma}\right) T+\left(\frac{\rho \kappa}{\sigma}-\frac{1}{2}\right) I+\sqrt{\left(1-\rho^{2}\right) I} Z$.
For the step (1), one need to generate a sample from the distribution of $X_{T}$. Fortunately, the distribution of $X_{T}$ is well-known in the literature (e.g. see Glasserman [20]). The law of $X_{T}$ can be expressed as

$$
X_{T}=\frac{\sigma^{2}\left(1-e^{-\kappa T}\right)}{4 \kappa} \chi_{\delta}^{\prime 2}\left(\frac{4 \kappa e^{-\kappa T}}{\sigma^{2}\left(1-e^{-\kappa T}\right)} x\right)
$$

where $\delta=\frac{4 \kappa \theta}{\sigma^{2}}, \chi_{\delta}^{\prime 2}(\lambda)$ denotes the noncentral chi-squared random variable with $\delta$ degrees of freedom, and a noncentrality parameter $\lambda$. For the sampling of noncentral chi-squared random variable, refer to [20].

The Broadie and Kaya's idea of Fourier inversion techniques comes into play at the step (21). They showed that the conditional characteristic function $\varphi\left(\cdot \mid x, x_{T}\right)$ of $\int_{0}^{T} X_{t} d t$ given $X_{T}=x_{T}$ can be written as

$$
\begin{aligned}
& \varphi\left(\lambda \mid x, x_{T}\right)=\mathbb{E}_{x, y}\left[\exp \left(i \lambda \int_{0}^{T} X_{t} d t\right) \mid X_{T}=x_{T}\right] \\
& =\frac{\gamma(\lambda) e^{-(1 / 2)(\gamma(\lambda)-\kappa) T}\left(1-e^{-\kappa T}\right)}{\kappa\left(1-e^{-\gamma(\lambda) T}\right)} \\
& \times \exp \left\{\frac{x+x_{T}}{\sigma^{2}}\left[\frac{\kappa\left(1+e^{-\kappa T}\right)}{1-e^{-\kappa T}}-\frac{\gamma(\lambda)\left(1+e^{-\gamma(\lambda) T}\right)}{1-e^{-\gamma(\lambda) T}}\right]\right\} \frac{I_{0.5 \delta-1}\left[\sqrt{x x_{T}} \frac{4 \gamma(\lambda) e^{-0.5 \gamma(\lambda) T}}{\sigma^{2}\left(1-e^{-\gamma(\lambda) T}\right)}\right]}{I_{0.5 \delta-1}\left[\sqrt{x x_{T}} \frac{4 \kappa e^{-0.5 \kappa T}}{\sigma^{2}\left(1-e^{-\kappa T}\right)}\right]},
\end{aligned}
$$

where $\gamma(\lambda)=\sqrt{\kappa^{2}-2 \sigma^{2} i \lambda}$. Then they numerically inverted the conditional characteristic function $\varphi\left(\cdot \mid x, x_{T}\right)$ to have an approximation of the conditional distribution,

$$
\begin{equation*}
\mathbb{P}_{x}\left(\int_{0}^{T} X_{t} d t \leq v \mid X_{T}=x_{T}\right) \approx F_{h, N}\left(v \mid x, x_{T}\right)=\frac{h v}{\pi}+\frac{2}{\pi} \sum_{n=1}^{N} \frac{\sin (h n v)}{n} \operatorname{Re}\left[\varphi\left(h n \mid x, x_{T}\right)\right] \tag{13}
\end{equation*}
$$

where $h>0$ is the discretization step size. Finally, they applied the inverse transform method to $F_{h, N}\left(\cdot \mid x, x_{T}\right)$ to simulate the integral $\int_{0}^{T} X_{t} d t$, i.e., they generated a uniform random number $U$ and solved the following equation for $\int_{0}^{T} X_{t} d t$ numerically

$$
F_{h, N}\left(\int_{0}^{T} X_{t} d t \mid x, x_{T}\right)=U .
$$

## 4 Exact Simulation Method

In this section, we present in detail our exact sampling algorithm of the WMSV model. Since the process $(X, Y)$ is a time-homogeneous Markov process, the exact simulation method of $\left(X_{T}, Y_{T}\right)$ given an initial value $\left(X_{0}, Y_{0}\right)=(x, y)$ can be extended to the exact simulation method of ( $X_{t_{2}}, Y_{t_{2}}$ ) given ( $X_{t_{1}}, Y_{t_{2}}$ ) for arbitrary $t_{1}<t_{2}$ in a straightforward manner. Therefore we only consider the simulation of $\left(X_{T}, Y_{T}\right)$ given an initial value $\left(X_{0}, Y_{0}\right)=(x, y)$.

For the WMSV model, a naive extension of Broadie and Kaya method is hardly accomplishable. The difficulty comes from the dimensionality of the volatility factor process $X$. In the Broadie and Kaya method, one need to generate a sample from the conditional (univariate) distribution of $\int_{0}^{T} X_{t} d t$ given $X_{T}=x_{T}$, and it can be achieved by Fourier inversion techniques. In contrast, the integrated volatility factor $\int_{0}^{T} X_{t} d t$ of the WMSV model has a $d \times d$ matrix-variate distribution, which makes it almost impossible to generate a sample from the distribution by Fourier inversion techniques. Hence, instead of following the Broadie and Kaya's approach, we take a rather direct way to achieve the goal through Theorem [2.3. Roughly, our method consists of the following two step.
Algorithm 2. This algorithm generates $L$ pairs of the state variables $\left(X_{T}, Y_{T}\right)$ of the WMSV model.

Step (1) Generate $L$ samples $X_{T}^{(1)}, \cdots, X_{T}^{(L)}$ from the distribution of $X_{T}$
Step (2) For each $l=1, \cdots, L$, generate a sample $Y_{T}^{(l)}$ from the conditional distribution of $Y_{T}$ given $X_{T}=X_{T}^{(l)}$

In the following subsections, we go through the details of these two steps.

### 4.1 Sampling from the Distribution of $X_{T}$

As indicated in the proof of Theorem [2.3, $X_{T}$ has noncentral Wishart distribution with degrees of freedom $\delta$, covariance matrix $V(0,0)$, and matrix of noncentrality parameter $V(0,0)^{-1} \Psi(0,0)^{\top}$ $x \Psi(0,0)$. The Wishart distributions with integer degrees of freedom (i.e., $\delta \in \mathbb{N}$ ) are extensively studied in the literature of the multivariate statistical analysis (e.g. see [21, 27, 29]).

In case that $\delta$ is an integer which is greater than or equal to $d$, one way of exact simulation is squaring a normal random matrix [8]. Let $N_{t}$ be a solution of the following equation

$$
\begin{equation*}
d N_{t}=N_{t} H^{\top} d t+d B_{t} \Sigma, \quad \text { with } \quad N_{0}^{\top} N_{0}=x, \tag{14}
\end{equation*}
$$

where $B$ is a standard $\delta \times d$ matrix Brownian motion. One can easily check that $N^{\top} N$ satisfies the stochastic differential equation (2). The equation (14) admits a closed form solution

$$
N_{t}=\left(N_{0}+\int_{0}^{t} d B_{s} \Sigma e^{-s H}\right) e^{t H}, \quad \text { for } \quad 0 \leq t<\infty
$$

Notice that the rows of $N_{T}$ are independent normal random vectors with common covariance matrix $V(0,0)$ and the mean of $N_{T}$ is $N_{0} e^{T H}$. Therefore, for integer degrees $\delta$ of freedom, the exact sampling of $X_{T}$ can be achieved by sampling $\delta \times d$ i.i.d. normal random variables. And there is also a more sophisticated but efficient way of simulating noncentral Wishart distributions with an integer degrees of freedom 21].

As far as we know, an exact sampling scheme for a noncentral Wishart distribution with non-integer valued degrees of freedom has only recently been devised by Ahdida and Alfonsi [2]. They have used a splitting method of the infinitesimal generator of Wishart processes. Their exact sampling scheme requires sampling of at most $d(d-1) / 2$ i.i.d. normal random variables and $d$ noncentral chi-square random variables. In the numerical experiments of this paper, we used the exact sampling method of Ahdida and Alfonsi [2].

### 4.2 Sampling from the Conditional Distribution of $Y_{T}$ Given $X_{T}$

This subsection is devoted to the step of sampling from the conditional distribution of $Y_{T}$ given $X_{T}=x_{T}$. This step is the most technical and time consuming step in our method. As explained in Section 3, Broadie and Kaya adopted Fourier inversion techniques to invert the conditional characteristic function of $\int_{0}^{T} X_{t} d t$ given $X_{T}=x_{T}$. We follow a similar approach, but we invert directly the conditional characteristic function of $Y_{T}$ given $X_{T}=x_{T}$ to avoid the difficulty in converting the characteristic function of matrix-variate random variable.

Let $\varphi\left(\cdot ; x, y, x_{T}\right)$ be the conditional characteristic function, i.e.,

$$
\varphi\left(\lambda ; x, y, x_{T}\right)=\mathbb{E}_{x, y}\left[e^{i \lambda Y_{T}} \mid X_{T}=x_{T}\right], \quad \text { for } \quad \lambda \in \mathbb{R}
$$

And let $F\left(\cdot ; x, y, x_{T}\right)$ be the corresponding distribution function:

$$
F\left(v ; x, y, x_{T}\right)=\mathbb{P}_{x, y}\left(Y_{T} \leq v \mid X_{T}=x_{T}\right)
$$

By Levy's inversion formula, the distribution can be recovered from the characteristic function, i.e., for $-\infty<l_{\epsilon}<v$, we have

$$
\begin{aligned}
F\left(v ; x, y, x_{T}\right) & =F\left(l_{\epsilon} ; x, y, x_{T}\right)+\frac{1}{2 \pi} \int_{-\infty}^{\infty} \varphi\left(\lambda ; x, y, x_{T}\right) \frac{e^{-i \lambda l_{\epsilon}}-e^{-i \lambda v}}{i \lambda} d \lambda \\
& =F\left(l_{\epsilon} ; x, y, x_{T}\right)+\frac{1}{\pi} \int_{0}^{\infty} \operatorname{Re}\left[\varphi\left(\lambda ; x, y, x_{T}\right) \frac{e^{-i \lambda l_{\epsilon}}-e^{-i \lambda v}}{i \lambda}\right] d \lambda \\
& =F\left(l_{\epsilon} ; x, y, x_{T}\right)+\frac{1}{\pi} \int_{0}^{\infty} \operatorname{Im}\left[\varphi\left(\lambda ; x, y, x_{T}\right)\left(e^{-i \lambda l_{\epsilon}}-e^{-i \lambda v}\right)\right] \frac{d \lambda}{\lambda}
\end{aligned}
$$

Notice that we can make $F\left(l_{\epsilon} ; x, y, x_{T}\right)$ small enough to ignore by taking a small $l_{\epsilon}$. So the integral terms are dominant in the above expressions. The integral above can be approximated by a numerical integration method. We use the trapezoidal rule to compute the distribution function numerically. It is known that the trapezoidal rule works well for oscillating integrands because the errors tend to be cancelled (see section 4 of Abate and Whitt [1). An integral on
the whole positive real line can be approximated in the following way. For notational simplicity, we write the integrand as $g(\lambda)$.

$$
\int_{0}^{\infty} g(\lambda) d \lambda=\frac{g(0+)}{2} h+h \sum_{n=1}^{\infty} g(n h)-e_{d}(h)=\frac{g(0+)}{2} h+h \sum_{n=1}^{N} g(n h)-e_{d}(h)-e_{t}(N),
$$

where $e_{d}(h)$ and $e_{t}(N)$ are errors due to the discretization of the continuous variable and the truncation of the infinite sum, respectively. Each of these errors can be made arbitrarily small by taking sufficiently small $h$ and sufficiently large $N$, and we give a detailed discussion on the control of these errors in Section 4.3. One can easily find that $g(0+)=v-l_{\epsilon}$. Therefore, we approximate the distribution function by the following finite trigonometric series:

$$
\begin{align*}
F_{h, N}\left(v ; x, y, x_{T}\right)= & \frac{h\left(v-l_{\epsilon}\right)}{2 \pi}+\frac{1}{\pi} \sum_{n=1}^{N} \operatorname{Im}\left[\frac{\varphi\left(n h ; x, y, x_{T}\right)}{n}\left(e^{-i n h l_{\epsilon}}-e^{-i n h v}\right)\right] \\
= & \frac{h\left(v-l_{\epsilon}\right)}{2 \pi}+\frac{1}{\pi} \sum_{n=1}^{N}\left(\frac{\operatorname{Re}\left[\varphi\left(n h ; x, y, x_{T}\right)\right]}{n}\left(\sin (n h v)-\sin \left(n h l_{\epsilon}\right)\right)\right)  \tag{15}\\
& -\frac{1}{\pi} \sum_{n=1}^{N}\left(\frac{\operatorname{Im}\left[\varphi\left(n h ; x, y, x_{T}\right)\right]}{n}\left(\cos (n h v)-\cos \left(n h l_{\epsilon}\right)\right)\right)
\end{align*}
$$

We address how to evaluate the conditional characteristic function $\varphi\left(\lambda ; x, y, x_{T}\right)$. The conditional characteristic function can be obtained by taking $u=-i \lambda$ in the formula (77). The formula (7) involves the solutions $\psi, \phi, \Psi$, and $V$ of the system (8) of ordinary differential equations. As we have shown in Section [2.3, the system (8) admits a closed-form solution for the Heston model. In general, the system (8) does not admit a closed-form solution because of non-commutativity of matrix multiplications. But the system (8) can be efficiently solved by numerical methods. In particular, we used the MATLAB function ode45 for solving equations in our experiments 1 It should be noted that the functions $\psi, \phi, \Psi$, and $V$ are needed to be evaluated only at $(0,0),(0,-i h), \cdots,(0,-i N h)$, and such evaluation points can be chosen uniformly across samples $X_{T}^{(1)}, \cdots, X_{T}^{(L)}$. Therefore, it is enough to solve the system (8) at those grid points only once for all simulation runs, and this numerical step does not cause a computational burden if the number $L$ of simulation runs is relatively larger than $N$.

The expression (7) requires calculations of the power of complex numbers, which is, in general, a multi-valued function:

$$
(z)^{\nu}=\left(|z| e^{i \operatorname{Arg}(z)}\right)^{\nu}=\left(|z| e^{i(\operatorname{Arg}(z)+2 m \pi)}\right)^{\nu}=|z|^{\nu} e^{i \nu \operatorname{Arg}(z)+2 m \nu \pi}, \quad m \in \mathbb{Z}
$$

with principal argument $-\pi<\operatorname{Arg}(z) \leq \pi$. If $\lambda \mapsto z(\lambda)$ is a complex-valued continuous function on $\mathbb{R}$ which does not attains 0 , then there exists a unique continuous function $\lambda \mapsto \arg (z(\lambda))$ such that $-\pi<\arg (z(0)) \leq \pi$. Such a continuation can be easily constructed by tracing the principal argument $\operatorname{Arg}(z(\lambda))$ and adding or subtracting $2 \pi$ at the discontinuous points. And then, we can use it to construct a continuous version of the power function:

$$
(z(\lambda))^{\nu}=|z(\lambda)|^{\nu} e^{i \nu \arg (z(\lambda))} .
$$

It is obvious that $\lambda \mapsto \operatorname{det} V(0,-i \lambda)$ is a continuous function which does not attains 0 , and we can apply the above observation to calculate $(\operatorname{det} V(0,-i \lambda))^{\delta / 2}$.

[^1]The conditional characteristic function (7) involves a hypergeometric function of matrix argument. The hypergeometric function of matrix argument is defined by the following power series

$$
\begin{equation*}
{ }_{0} F_{1}(b ; y)=\sum_{k=0}^{\infty} \sum_{|\iota|=k} \frac{1}{(b)_{\iota}} \frac{C_{\iota}(y)}{k!} . \tag{16}
\end{equation*}
$$

For the definitions of $|\iota|$ and $(b)_{\iota}$, see Section A.1. The zonal polynomials $C_{\iota}(y)$ are not polynomials of the matrix $y$, but polynomials of its eigenvalues. So, we sometimes write the hypergeometric functions and the zonal polynomials in the following way:

$$
C_{\iota}(y)=C_{\iota}\left(\alpha_{1}, \cdots, \alpha_{d}\right), \quad{ }_{0} F_{1}(b ; y)={ }_{0} F_{1}\left(b ; \alpha_{1}, \cdots, \alpha_{d}\right),
$$

where $\alpha_{1}, \cdots, \alpha_{d}$ are eigenvalues of $y$. Koev and Edelman [26] exploited the combinatorial properties of zonal polynomials and the generalized hypergeometric coefficients to develop an algorithm for computing a truncated version of (16):

$$
{ }_{0}^{m} F_{1}\left(b ; \alpha_{1}, \cdots, \alpha_{d}\right)=\sum_{k=0}^{m} \sum_{|\iota|=k} \frac{1}{(b)_{\iota}} \frac{C_{\iota}\left(\alpha_{1}, \cdots, \alpha_{d}\right)}{k!} .
$$

Since the denominator of the series (16) grows faster than factorial order, the series converges quickly and the truncated version gives a good approximation. We provide a detailed error analysis of the truncated series in Section 4.3, MATLAB implementations of Koev and Edelman's algorithms are available in the author's homepage [25]. But the routine gets only real eigenvalues as input arguments. So we modified the codes applicable for complex input arguments.

To sample the $\log$-price $Y_{T}$, we use the inverse transform method. We generate a uniform random number $U$ and we numerically solve the equation for $Y_{T}$

$$
\begin{equation*}
F_{h, N}\left(Y_{T} ; x, y, x_{T}\right)=U . \tag{17}
\end{equation*}
$$

The equation can be solved efficiently by numerical method, e.g., Newton's method, because the function $F_{h, N}$ is a strictly increasing function. We provide details in the following subsection.

### 4.3 Some Implementation Issues

In order to implement our method, we need to resolve some issues. We have to determine appropriate values of $l_{\epsilon}, h$, and $N$ in (15), and decide the number of summands to approximate the infinite series (16). We also need to address how to solve the equation (17).

A careful choice of values of $l_{\epsilon}, h$, and $N$ is crucial in our method, because they determine the points where the characteristic function $\varphi\left(\cdot ; x, y, x_{T}\right)$ is evaluated. Too few evaluation points might introduce a large bias of the Monte Carlo simulation due to the approximation error of the conditional distribution function. Too many points might make our method too slow, because the evaluation of the characteristic function $\varphi\left(\cdot ; x, y, x_{T}\right)$ is the most time-consuming step of our method.

As indicated in the previous subsection, the approximation (15) involves three different kinds of errors: the discretization error $e_{d}(h)$, the truncation error $e_{t}(N)$, and the left tail probability $F\left(l_{\epsilon} ; x, y, x_{T}\right)$. We suggest a way to control these errors based on the conditional mean and standard deviation of $Y_{T}$ given $X_{T}=x_{T}$. Recall that the mean and standard deviation can be easily found by differentiating the characteristic function $\varphi\left(\cdot ; x, y, x_{T}\right)$ :

$$
\begin{aligned}
\mu\left(x_{T}\right) & =\mathbb{E}_{x, y}\left[Y_{T} \mid X_{T}=x_{T}\right]=\operatorname{Im}\left(\varphi^{\prime}\left(0 ; x, y, x_{T}\right)\right), \\
\sigma\left(x_{T}\right)^{2} & =\mathbb{E}_{x, y}\left[\left(Y_{T}-\mu\right)^{2} \mid X_{T}=x_{T}\right]=-\operatorname{Re}\left(\varphi^{\prime \prime}\left(0 ; x, y, x_{T}\right)\right)-\left(\operatorname{Im}\left(\varphi^{\prime}\left(0 ; x, y, x_{T}\right)\right)\right)^{2} .
\end{aligned}
$$

In the section 5 of Abate and Whitt [1] they used Poisson summation formula to prove that the discretization error is bounded by two simple tail probabilities

$$
0 \leq e_{d}(h) \leq\left(1-F\left(2 \pi / h+l_{\epsilon} ; x, y, x_{T}\right)\right)+F\left(-2 \pi / h+v ; x, y, x_{T}\right)
$$

for $v-l_{\epsilon}<2 \pi / h$. Suppose that we want to calculate $F\left(v ; x, y, x_{T}\right)$ with an error less than $\epsilon$. Let $l_{\epsilon}$ and $u_{\epsilon}$ such that $F\left(l_{\epsilon} ; x, y, x_{T}\right) \leq \epsilon / 4$ and $F\left(u_{\epsilon} ; x, y, x_{T}\right) \geq 1-\epsilon / 4$. Take $h=\frac{2 \pi}{u_{\epsilon}-l_{\epsilon}}$. Then, for $l_{\epsilon} \leq v \leq u_{\epsilon}$,

$$
\frac{2 \pi}{h}+l_{\epsilon}=u_{\epsilon}, \quad \text { and } \quad-\frac{2 \pi}{h}+v \leq-\frac{2 \pi}{h}+u_{\epsilon}=l_{\epsilon}
$$

so that

$$
\left|F\left(l_{\epsilon} ; x, y, x_{T}\right)-e_{d}(h)\right| \leq \max \left\{F\left(l_{\epsilon} ; x, y, x_{T}\right), e_{d}(h)\right\} \leq \epsilon / 2 .
$$

Then we turn to the truncation error $e_{t}(N)$. Since the summands in (15) are oscillating, the absolute value of the last term gives an estimate for the truncation error:

$$
\left|e_{t}(N)\right| \approx \frac{1}{\pi}\left|\operatorname{lm}\left[\frac{\varphi\left(N h ; x, y, x_{T}\right)}{N}\left(e^{-i N h l_{\epsilon}}-e^{-i N h v}\right)\right]\right| \leq \frac{2\left|\varphi\left(N h ; x, y, x_{T}\right)\right|}{\pi}
$$

So, we may choose a large $N$ so that $\left|\varphi\left(N h ; x, y, x_{T}\right)\right| \leq \epsilon \pi / 4$ to make the truncation error approximately less than $\epsilon / 2$.

These error bounds and estimates are theoretically appealing, but they are not suitable for practical use. For example, we can compute $\varphi\left(N h ; x, y, x_{T}\right)$, but it is very time-consuming. Furthermore, we need to evaluate $F\left(\cdot ; x, y, x_{T}\right)$, which is not known in advance. To overcome these difficulties, we exploit a heuristic idea, which is obtained from numerical experiments. The conditional distribution of $Y_{T}$ given $X_{T}=x_{T}$ is roughly similar to the normal distribution, because the terminal value $X_{T}$ of volatility factor is already known and the main source of randomness is the stochastic integral with respect to a Brownian motion. But they show some differences as well. The normal distribution is symmetric with respect to mean, but the conditional distribution of $Y_{T}$ is asymmetric, and the decay rate of the conditional characteristic function of $Y_{T}$ gets slower as the correlation parameter $R$ increases. So, we start with $l_{\epsilon}, h$, and $N$ which are suggested by the normal distribution, and then modify them to take such effects into account. The normal distribution suggests us to take $l_{\epsilon}, h$, and $N$ of the following form:

$$
\left\{\begin{aligned}
l_{\epsilon} & =\Phi^{-1}\left(\epsilon / 4 ; \mu\left(x_{T}\right), \sigma\left(x_{T}\right)^{2}\right)=\mu\left(x_{T}\right)+\sigma\left(x_{T}\right) \Phi^{-1}(\epsilon / 4), \\
h & =2 \pi\left(\Phi^{-1}\left(1-\epsilon / 4 ; \mu\left(x_{T}\right), \sigma\left(x_{T}\right)^{2}\right)-\Phi^{-1}\left(\epsilon / 4 ; \mu\left(x_{T}\right), \sigma\left(x_{T}\right)^{2}\right)\right)^{-1} \\
& =-\pi\left(\sigma\left(x_{T}\right) \Phi^{-1}(\epsilon / 4)\right)^{-1} \\
N & =\left\lceil\frac{1}{h \sigma\left(x_{T}\right)} \sqrt{-2 \log (\pi \epsilon / 4)}\right\rceil,
\end{aligned}\right.
$$

where $\Phi^{-1}\left(\cdot ; \mu, \sigma^{2}\right)$ and $\Phi^{-1}(\cdot)$ are the inverse functions of the normal distribution $\mathcal{N}\left(\mu, \sigma^{2}\right)$ and the standard normal distribution, respectively. And $\lceil a\rceil$ denotes the smallest integer greater than $a$. Then we modify them with the mean $\mu\left(x_{T}\right)$ and the correlation parameter $R$. Since we want to choose the points $u=-i n h, n=1, \cdots, N$ uniformly across samples $X_{T}^{(1)}, \cdots, X_{T}^{(L)}$, we take $l_{\epsilon}, h$, and $N$ in a conservative way:

$$
\left\{\begin{align*}
l_{\epsilon} & =\min _{l=1, \cdots, L}\left\{\mu\left(x_{T}^{(l)}\right)+\sigma\left(X_{T}^{(l)}\right) \Phi^{-1}(\epsilon / 4)\right\}  \tag{18}\\
h & =-\pi \times \min _{l=1, \cdots, L}\left(\left(c_{1}\left|\mu\left(X_{T}^{(l)}\right)\right|+\sigma\left(X_{T}^{(l)}\right)\right) \Phi^{-1}(\epsilon / 4)\right)^{-1} \\
N & =\max _{l=1, \cdots, L}\left[\frac{1+c_{2} \sqrt{\operatorname{tr}\left[\left[^{\top} R\right]\right.}}{h \sigma\left(X_{T}^{(l)}\right)} \sqrt{-2 \log (\pi \epsilon / 4)}\right]
\end{align*}\right.
$$

for appropriate constants $c_{1} \geq 0$ and $c_{2} \geq 0$. In particular, we take $c_{1}=0.1$ and $c_{2}=0.5$ in our experiments.

Now, we consider the problem of determining where to truncate the hypergeometric series (16). The partitions of integers can be ordered lexicographically, i.e., if $\iota=\left(k_{1}, \cdots, k_{d}\right)$ and $\tilde{\iota}=\left(l_{1}, \cdots, l_{d}\right)$ are two partitions of integers we will write $\iota>\tilde{\iota}$ if $k_{i}>l_{i}$ for the first index $i$ at which the parts become unequal. With this order relation, we give an upper bound for the truncation error of the hypergeometric functions. The proof is given in Appendix A.4.

Proposition 4.1. For $b>\frac{1}{2}(d-1)$ and an integer $m \geq d-1$, we have for $\alpha_{1}, \cdots, \alpha_{d} \in \mathbb{C}$

$$
\begin{equation*}
\left|{ }_{0} F_{1}\left(b ; \alpha_{1}, \cdots, \alpha_{d}\right)-{ }_{0}^{m} F_{1}\left(b ; \alpha_{1}, \cdots, \alpha_{d}\right)\right| \leq \frac{\left(\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|\right)^{m+1}}{(m+1)!(b)_{\hat{\iota}(m+1)}} e^{\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|} \tag{19}
\end{equation*}
$$

where $\hat{\iota}(m+1)$ is the smallest partition among the partitions of $m+1$ into not more than $d$ parts.

Notice that the smallest partition $\hat{\iota}(k)=\left(k_{1}, \cdots, k_{d}\right)$ among the partitions of $k$ into not more than $d$ parts is the most evenly distributed partition, and the smallest partition $\hat{\iota}(k+1)$ of $k+1$ is the partition which is obtained by adding 1 to $k_{1}$ if all the $k_{i}$ 's are equal, or adding 1 to $k_{j}$ at which $k_{j}$ differs from $k_{j-1}$ otherwise. For example, $(2,2,2,2,1),(2,2,2,2,2)$, and $(3,2,2,2,2)$ are the smallest partitions of 9,10 , and 11 into 5 parts, respectively. With this observation, we can compute the hypergeometric coefficients $(b)_{\hat{\iota}(k)}$ as follows: $(b)_{(1)}=b$ and

$$
(b)_{\hat{\iota}(k+1)}= \begin{cases}(b)_{\hat{\iota}(k)}\left(b+k_{1}\right) & \text { if } k_{1}=\cdots=k_{d}  \tag{20}\\ (b)_{\hat{\iota}(k)}\left(b+k_{j}-\frac{1}{2}(j-1)\right) & \text { if } k_{1}=\cdots=k_{j-1} \neq k_{j}=\cdots=k_{d}\end{cases}
$$

where $\hat{\iota}(k)=\left(k_{1}, \cdots, k_{d}\right)$.
We use the bound (19) as a criterion for truncating the series (16). Notice that the parameter $b$ is a constant $\frac{1}{2} \delta$, which satisfies the assumption of Proposition 4.1. Since it is a constant, we need to compute the hypergeometric coefficients of the smallest partitions only once in all the simulation runs, and it does not add computational burden to our method. So the upper bound (19) can be calculated with minor additional computational budget.

In order to solve the equation (17), we use the Newton's method. A careful choice of the initial guess is necessary to enhance the convergence. As explained before, the conditional distribution $F\left(v ; x, y, x_{T}\right)$ is approximated as the normal one with mean $\mu\left(x_{T}\right)$ and variance $\sigma\left(x_{T}\right)^{2}$. Therefore, we take the initial guess $v_{0}$ as the solution of the equation

$$
\Phi\left(v_{0} ; \mu\left(x_{T}\right), \sigma\left(x_{T}\right)^{2}\right)=U, \quad \text { or } \quad v_{0}=\Phi^{-1}\left(U ; \mu\left(x_{T}\right), \sigma\left(x_{T}\right)^{2}\right)
$$

Then we apply the Newton's method to the function $F_{h, N}\left(v ; x, y, x_{T}\right)$ to generate a sequence which approximates the solution of the equation (17):

$$
v_{k+1}=v_{k}-\frac{F_{h, N}\left(v_{k} ; x, y, x_{T}\right)}{F_{h, N}^{\prime}\left(v_{k} ; x, y, x_{T}\right)}
$$

We iterate the loop for a fixed number of times, usually 4 or 5 times, and then use the bisection search method if the sequence, generated by Newton's method, fails to converges within the predetermined number of iterations.

We summarize the method of generating a sample from the conditional distribution of $Y_{T}$ given $X_{T}=x_{T}$.
Algorithm 3. This algorithm generates $L$ samples $Y_{T}^{(1)}, \cdots, Y_{T}^{(L)}$ from the conditional distribution of $Y_{T}$ given $X_{T}=X_{T}^{(1)}, \cdots, X_{T}^{(L)}$, respectively.

Step (1) Determine $l_{\epsilon}, h$, and $N$ according to (18),
Step (2) Evaluate $\psi(0, u), \phi(0, u), \Psi(0, u)$, and $V(0, u)$ at $u=-i n h, n=1, \cdots, N$
Step (3) Evaluate $\varphi\left(n h ; x, y, X_{T}^{(l)}\right)$ for $n=1, \cdots, N$ and $l=1, \cdots, L$
Step (4) Generate IID uniform random numbers $U^{(1)}, \cdots, U^{(L)}$
Step (5) For each $l=1, \cdots, L$, solve $F_{h, N}\left(Y_{T}^{(l)} ; x, y, X_{T}^{(l)}\right)=U^{(l)}$ for $Y_{T}^{(l)}$

## 5 Numerical Results

In this section, we compare numerically our exact simulation method with other existing methods: the Euler discretization scheme for the WMSV model and the Broadie-Kaya method for the Heston model.

### 5.1 Comparison between the Exact Sampling and the Euler Scheme

The system of stochastic differential equations (2) and (3) does not admit a closed form solution. In such a case, one way to simulate the model is to discretize the time interval and simulate another process, which approximates the model, on these discrete time grids. Euler discretization is a basic discretization scheme. Let $0=t_{0}<t_{1}<\cdots<t_{N}=T$ be a partition of the time interval $[0, T]$ into $N$ equal subintervals, i.e., $\Delta t=t_{i}-t_{i-1}=T / N, i=1, \cdots, N$. We discretize the Wishart process (2) by setting $\hat{X}_{t_{0}}=x$, and

$$
\hat{X}_{t_{i}}=\left(\hat{X}_{t_{i-1}}+\left(\delta \Sigma^{\top} \Sigma+H \hat{X}_{t_{i-1}}+\hat{X}_{t_{i-1}} H^{\top}\right) \Delta t+\sqrt{\hat{X}_{t_{i-1}}} \Delta W_{t_{i}} \Sigma+\Sigma^{\top}\left(\Delta W_{t_{i}}\right)^{\top} \sqrt{\hat{X}_{t_{i-1}}}\right)^{+}
$$

where $\Delta W_{t_{i}}=W_{t_{i}}-W_{t_{i-1}}$. Here, $A^{+}$denotes the positive part of a symmetric matrix $A$ : we set $A^{+}=O \operatorname{diag}\left(\lambda_{1}^{+}, \cdots, \lambda_{d}^{+}\right) O^{\top}$ if $A=O \operatorname{diag}\left(\lambda_{1}, \cdots, \lambda_{d}\right) O^{\top}$ and $O$ an orthogonal matrix. In order to make $\hat{X}_{t_{i}}$ positive semidefinite, we take the positive part at each time grid. The discretization of the log-price process (3) is

$$
\hat{Y}_{t_{i}}=\hat{Y}_{t_{i-1}}+\left(r-\frac{1}{2} \operatorname{tr}\left[\hat{X}_{t_{i-1}}\right]\right) \Delta t+\operatorname{tr}\left[\sqrt{\hat{X}_{t_{i-1}}}\left(\Delta W_{t_{i}} R^{\top}+\Delta Z_{t_{i}} \sqrt{I_{d}-R R^{\top}}\right)\right]
$$

where $\Delta Z_{t_{i}}=Z_{t_{i}}-Z_{t_{i-1}}$. Euler discretization scheme is easy to understand and implement. But it has serious disadvantages: the distribution of the samples drawn from Euler scheme is different from the true distribution, and it may require very fine discretization to reduce the bias as small as acceptable. These will be illustrated by numerical results.

We will compare the exact simulation method with the Euler discretization scheme in the terms of distributions, convergence and performance. The simulation experiments in this paper were performed on a desktop PC with an Intel Core2 Quad 3.00 GHz processor and 3.25 GB RAM, running Windows XP Professional. We used the MATLAB in the version R2009a.

In order to demonstrate the difference of empirical distributions of exact simulation method and Euler discretization method, we generated $10^{6}$ samples and estimated the density functions of $Y_{T}$ using those samples. To estimate density functions, we used the MATLAB function ksdensity which computes a density estimate using kernel density estimation. The theoretical density function was obtained by a numerical inversion of the characteristic function (4). Figure 1 shows the density functions of $Y_{T}$ for the parameter set:

$$
\delta=1.1, \quad r=0, \quad y=0, \quad T=1.0
$$



Figure 1: Density estimates of the log-price: The upper left panel shows the overall shapes of density estimates, and the other three panels magnify different parts of the upper left one.

$$
\begin{aligned}
& x=\left[\begin{array}{ll}
0.0298 & 0.0119 \\
0.0119 & 0.0108
\end{array}\right], \quad H=\left[\begin{array}{cc}
-1.2479 & -0.8985 \\
-0.0820 & -1.1433
\end{array}\right], \\
& \Sigma=\left[\begin{array}{ll}
0.3417 & 0.3493 \\
0.1848 & 0.3090
\end{array}\right], \quad R=\left[\begin{array}{ll}
-0.2243 & -0.1244 \\
-0.2545 & -0.7230
\end{array}\right] .
\end{aligned}
$$

The set of parameters except $\delta$ for the model is taken from Da Fonseca and Grassell: 2 [13]. The number of time steps for the Euler method is set to 25 . From Figure 11, we observe that the distribution of the samples drawn from Euler method apparently differs from the true distribution.

The errors in density estimates result from the discretization bias of the Euler scheme. Even worse, there is no appropriate way to measure the bias errors. In contrast, the error of exact simulations comes mostly from the variance, and it can be easily measured by the sample variance. To illustrate such a difference, we present European call option price estimates obtained by two simulation methods. The model parameters are the same as above except $\delta=3.2$. We consider an at-the-money call option, i.e., the strike price $K$ is set to equal to

[^2]| Methods | $\begin{gathered} \text { No. of } \\ \text { time steps } \end{gathered}$ | No. of simulation runs | MC estimates | std. errors | $\begin{aligned} & \text { Time } \\ & (\mathrm{sec}) \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Exact | N/A | 50000 | 0.192415 | $0.1430 \times 10^{-2}$ | 12.14 |
|  |  | 100000 | 0.191451 | $0.1009 \times 10^{-2}$ | 24.28 |
|  |  | 500000 | 0.192143 | $0.4535 \times 10^{-3}$ | 121.4 |
|  |  | 1000000 | 0.191513 | $0.3202 \times 10^{-3}$ | 242.8 |
| Euler | 50 | 50000 | $0.195202^{*}$ | $0.1456 \times 10^{-2}$ | 179.4 |
|  |  | 100000 | 0.194829* | $0.1034 \times 10^{-2}$ | 358.7 |
|  |  | 500000 | $0.193827^{*}$ | $0.4603 \times 10^{-3}$ | 1793.5 |
|  |  | 1000000 | $0.194197^{*}$ | $0.3259 \times 10^{-3}$ | 3587.0 |
|  | 100 | 50000 | 0.194016 | $0.1461 \times 10^{-2}$ | 358.4 |
|  |  |  | 0.193264 | $0.1027 \times 10^{-2}$ | 716.8 |
|  |  | 500000 | $0.193008^{*}$ | $0.4570 \times 10^{-3}$ | 3584.2 |
|  |  | 1000000 | $0.193073^{*}$ | $0.3234 \times 10^{-3}$ | 7168.5 |

Table 1: The Monte-Carlo estimates of the call option price with $\delta=3.2$ : The theoretical option price is 0.191575 . The asterisked numbers are those for which the theoretical price lies outside of the $95 \%$ confidence interval.
$S_{0}=e^{Y_{0}}=1$. We computed the theoretical price of the option by applying the Carr and Madan approach [9 to the Laplace transform formula (4). Table 1 shows the Monte-Carlo estimates of the option price. The Euler method gives unreliable results : the theoretical price lies in only two (out of eight) $95 \%$ confidence intervals built by the Euler method. Furthermore, the computation times spent by the Euler method are much longer than those for the exact simulation method.

The accuracy of the Euler method gets worse as $\delta$ decreases. For a small $\delta$, the discretized Wishart process crosses more frequently the boundary of the cone of symmetric positive definite matrices. At each time it passes the boundary, its negative part is truncated to make it positive semidefinite, and such truncations might cause serious bias error. To illustrate such case, we set $\delta=1.1$, and all other parameters are set as the same as above. Table 2 reveals that the Euler discretization method hardly gives reliable estimates for $\delta=1.1$. The theoretical price never lies in the $95 \%$ confidence intervals built by the Euler method, but it always lies in the $95 \%$ confidence intervals from the exact simulation.

### 5.2 Comparison with Broadie and Kaya Method

To apply our method to the Heston model, we used the explicit formula (12) instead of (77). With (12), we do not need to solve the ordinary differential equations, and we can use the built-in MATLAB function besseli ${ }^{3}$ instead of the algorithm of Koev and Edelman [26] for the computation of the hypergeometric function of matrix arguments.

We pointed out in Section 4.3 that a careful choice of the grid size $h$ and the truncation number $N$ is necessary in our method. The same is true for any Fourier inversion based simulation method, e.g., Broadie and Kaya's method [7]. In the paper of Broadie and Kaya, they did not give any guideline on how to choose $h$ and $N$, and they argued that appropriate $h$ and $N$ can be found by trial and error. But it is difficult to find appropriate $h$ and $N$, and it will be illustrated by numerical experiments.

[^3]| Methods | No. of time steps | No. of simulation runs | MC estimates | std. errors | $\begin{aligned} & \hline \text { Time } \\ & (\mathrm{sec}) \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Exact | N/A | 50000 | 0.114129 | $0.7381 \times 10^{-3}$ | 15.34 |
|  |  | 100000 | 0.113630 | $0.5215 \times 10^{-3}$ | 30.68 |
|  |  | 500000 | 0.112916 | $0.2320 \times 10^{-3}$ | 153.40 |
|  |  | 1000000 | 0.113085 | $0.1642 \times 10^{-3}$ | 306.80 |
| Euler | 50 | 50000 | $0.117013^{*}$ | $0.7654 \times 10^{-3}$ | 180.33 |
|  |  | 100000 | $0.117020^{*}$ | $0.5441 \times 10^{-3}$ | 360.67 |
|  |  | 500000 | 0.116698* | $0.2428 \times 10^{-3}$ | 1803.34 |
|  |  | 1000000 | 0.116940* | $0.1718 \times 10^{-3}$ | 3606.67 |
|  | 100 | 50000 | $0.116940^{*}$ | $0.7541 \times 10^{-3}$ | 360.54 |
|  |  | 100000 | 0.115963* | $0.5321 \times 10^{-3}$ | 721.09 |
|  |  | 500000 | 0.115333* | $0.2383 \times 10^{-3}$ | 3605.45 |
|  |  | 1000000 | $0.115363{ }^{*}$ | $0.1684 \times 10^{-3}$ | 7210.90 |

Table 2: The Monte-Carlo estimates of the call option price with $\delta=1.1$ : The theoretical option price is 0.113000 . The asterisked numbers are those for which the theoretical price lies outside of the $95 \%$ confidence interval.

This section is designed to show how the various choices of $h$ and $N$ in (13) affect the accuracy and performance of the Broadie and Kaya's method, and to compare performance of their method with ours. For the purpose, we generate $10^{6}$ samples of the Heston model using our method, and we also generate the same number of samples using Broadie and Kaya's method for various combinations of $h$ and $N$. We used them to obtain Monte-Carlo estimates of the European call option price. The model parameter is set to

$$
\begin{aligned}
& x=0.010201, \quad y=\log (100), \quad \kappa=6.21, \quad \theta=0.019 \\
& \sigma=0.61, \quad \rho=-0.7, \quad r=0.0319, \quad T=1.0, \quad K=100
\end{aligned}
$$

This set of parameters is taken from Duffie et al. [16]: it was calibrated to the option prices for S\&P 500 on November 2, 1993, and used in Broadie and Kaya 7 as well. Table 3 shows the simulation results $\sqrt[4]{4}$ From the table, the pair of $h=32$ and $N=25$ seems to be the best choice of $h$ and $N$ among outputs of the experiments. But, even with that choice of $h$ and $N$, the Broadie and Kaya's method is slower than our method.

Through this experiment, we found that the Broadie-Kaya method requires many trials to find the optimal choice of $h$ and $N$. It is even worse that the optimal choice of $h$ and $N$ is sensitive to the change of the model parameters. To demonstrate the sensitivity of their method, we give another experimental results with the same parameter except $T=0.25$. Table 4 shows that the true option price never lies in the confidence intervals of their methods for the same choices of $h$ and $N$. In contrast, our automatic choice of $h$ and $N$ according to (18) makes our method adaptive to the parameter changes.

## 6 Concluding Remarks

We proposed a method for the exact simulation of the asset price and volatility factor of Wishart multidimensional stochastic volatility model(WMSV). Our simulation method is inspired by the Fourier inversion techniques of Broadie and Kaya [7], and based on the analysis of the conditional

[^4]| Methods | $h$ | $N$ | Nh | MC estimates | std. errors | Time(sec) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| KK | N/A | N/A | N/A | 6.8125 | $7.4302 \times 10^{-3}$ | 59.77 |
| BK | 8 | 100 | 800 | 6.8276* | $7.4377 \times 10^{-3}$ | 280.4 |
|  |  | 200 | 1600 | 6.8059 | $7.4206 \times 10^{-3}$ | 591.1 |
|  |  | 400 | 3200 | 6.7993 | $7.4208 \times 10^{-3}$ | 1172.0 |
|  | 16 | 50 | 800 | 6.8162 | $7.4299 \times 10^{-3}$ | 123.4 |
|  |  | 100 | 1600 | 6.8025 | $7.4220 \times 10^{-3}$ | 297.0 |
|  |  | 200 | 3200 | 6.8211* | $7.4257 \times 10^{-3}$ | 584.8 |
|  | 32 | 25 | 800 | 6.8098 | $7.4302 \times 10^{-3}$ | 68.0 |
|  |  | 50 | 1600 | 6.8011 | $7.4144 \times 10^{-3}$ | 132.6 |
|  |  | 100 | 3200 | 6.8191 | $7.4286 \times 10^{-3}$ | 293.6 |
|  | 64 | 13 | 832 | $6.8346{ }^{*}$ | $7.4252 \times 10^{-3}$ | 41.16 |
|  |  | 25 | 1600 | $6.8222^{*}$ | $7.4173 \times 10^{-3}$ | 72.52 |
|  |  | 50 | 3200 | 6.8184 | $7.4137 \times 10^{-3}$ | 130.8 |

Table 3: The Monte-Carlo estimates of the call option price with $T=1.0$ : The theoretical option price is 6.8061 . The asterisker numbers are those for which the theoretical price lies outside of the $95 \%$ confidence interval. KK refers to our method and BK refers to the method of Broadie and Kaya.

| Methods | $h$ | $N$ | Nh | MC estimates | std. errors | Time(sec) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| KK | N/A | N/A | N/A | 2.6720 | $2.9214 \times 10^{-3}$ | 78.18 |
| BK | 8 | 100 | 800 | $2.7883^{*}$ | $2.8438 \times 10^{-3}$ | 279.5 |
|  |  | 200 | 1600 | 2.7597* | $2.8631 \times 10^{-3}$ | 608.0 |
|  |  | 400 | 3200 | 2.7463* | $2.8705 \times 10^{-3}$ | 1290.2 |
|  | 16 | 50 | 800 | $2.7853^{*}$ | $2.8435 \times 10^{-3}$ | 124.0 |
|  |  | 100 | 1600 | 2.7587* | $2.8664 \times 10^{-3}$ | 306.6 |
|  |  | 200 | 3200 | 2.7483* | $2.8796 \times 10^{-3}$ | 647.7 |
|  | 32 | 25 | 800 | $2.7802^{*}$ | $2.8408 \times 10^{-3}$ | 67.3 |
|  |  | 50 | 1600 | $2.7610^{*}$ | $2.8678 \times 10^{-3}$ | 138.7 |
|  |  | 100 | 3200 | 2.7488* | $2.8798 \times 10^{-3}$ | 327.9 |
|  | 64 | 13 | 832 | $2.7845{ }^{*}$ | $2.8433 \times 10^{-3}$ | 40.63 |
|  |  | 25 | 1600 | 2.7651* | $2.8695 \times 10^{-3}$ | 75.19 |
|  |  | 50 | 3200 | $2.7445 *$ | $2.8736 \times 10^{-3}$ | 150.9 |

Table 4: The Monte-Carlo estimates of the call option price with $T=0.25$ : The theoretical option price is 2.6709 . The asterisked numbers are those for which the theoretical price lies outside of the $95 \%$ confidence interval. KK refers to our method and BK refers to the method of Broadie and Kaya.

Laplace transform of the asset price given a volatility level. The proposed simulation method can be used to generate unbiased price estimators for path-independent or mildly path-dependent options such as Bermudan options and American call options with discrete dividends.

We illustrated the accuracy and speed of the exact simulation method by numerical experiments. The numerical results confirmed that the exact simulation method gives accurate simulation results and it is faster than the crude Euler discretization method. The numerical comparison on the Heston model revealed that our method is more adaptive than the original exact simulation method of Broadie and Kaya.

In the paper of Broadie and Kaya [7], the authors extended the exact simulation method for the Heston model to other affine jump diffusion models. With the same approach as theirs, the exact simulation method for the WMSV model can be extended to single asset matrix affine jump diffusion models [28] provided that the jumps in the model has the constant intensity and the jump size can be sampled exactly. These extensions are straightforward and explained well in Broadie and Kaya [7].

## A Appendix

## A. 1 Noncentral Wishart Distributions \& Special Functions

This section is intended to recall the definitions of noncentral Wishart distributions and some multivariate special functions. For a detailed discussion on multivariate distributions and special functions, refer to Muirhead [29].

The probability density function of the noncentral Wishart distribution involves two special functions: the multivariate gamma function and the hypergeometric function of matrix arguments. We start with the multivariate gamma function, which is defined in terms of an integral over the cone $S_{d}^{++}$of symmetric positive definite $d \times d$ matrices.

Definition A.1. The multivariate gamma function, denoted by $\Gamma_{d}(a)$, is defined to be

$$
\Gamma_{d}(a)=\int_{S_{d}^{++}} \exp (-\operatorname{tr}[y])(\operatorname{det}[y])^{a-(d+1) / 2}(d y), \quad \text { for } \quad \operatorname{Re}(a)>\frac{1}{2}(d-1) .
$$

Note that when $d=1$, the multivariate gamma function becomes the usual gamma function, $\Gamma_{1}(a)=\Gamma(a)$.

In order to give the definition of the hypergeometric function of matrix arguments, we need to introduce the zonal polynomials, which is defined in terms of partitions of positive integers. Let $k$ be a positive integer. A partition $\iota$ of $k$ is written as $\iota=\left(k_{1}, k_{2}, \cdots\right)$, where $\sum_{j} k_{j}=k$ and $k_{1} \geq k_{2} \geq \cdots \geq 0$. We order the partitions lexicographically: let $\iota=\left(k_{1}, k_{2}, \cdots\right)$ and $\tilde{\iota}=\left(l_{1}, l_{2}, \cdots\right)$ be two partitions, we write $\iota>\hat{\iota}$ if $k_{j}>l_{j}$ for the first index $j$ at which two parts become unequal. In case $\iota=\left(k_{1}, \cdots, k_{d}\right)$ and $\tilde{\iota}=\left(l_{1}, \cdots, l_{d}\right)$ are two partitions with $\iota>\tilde{\iota}$, we say the monomial $\alpha_{1}^{k_{1}} \cdots \alpha_{d}^{k_{d}}$ is of higher weight than the monomial $\alpha_{1}^{l_{1}} \cdots \alpha_{d}^{l_{d}}$.

Definition A.2. Let $y$ be a $d \times d$ complex symmetric matrix with eigenvalues $\alpha_{1}, \cdots, \alpha_{d}$ and let $\iota$ be a partition of $k$ into not more than $d$ parts. The zonal polynomial of $y$ corresponding to $\iota=\left(k_{1}, \cdots, k_{d}\right)$, denoted by $C_{\iota}(y)$, is a symmetric, homogeneous polynomial of degree $k$ in the eigenvalues $\alpha_{1}, \cdots, \alpha_{d}$ such that
(i) The term of highest weight in $C_{\iota}(y)$ is $\alpha_{1}^{k_{1}} \cdots \alpha_{d}^{k_{d}}$, i.e.,

$$
C_{\iota}(y)=c_{\iota} \alpha_{1}^{k_{1}} \cdots \alpha_{d}^{k_{d}}+\text { terms of lower weight },
$$

where $c_{\iota}$ is a constant.
(ii) $C_{\iota}(y)$ is an eigenfunction of the differential operator $\Delta_{y}$ given by

$$
\Delta_{y}=\sum_{j=1}^{d} \alpha_{j}^{2} \frac{\partial^{2}}{\partial \alpha_{j}^{2}}+\sum_{j=1}^{d} \sum_{\substack{i=1 \\ i \neq j}}^{d} \frac{\alpha_{j}^{2}}{\alpha_{j}-\alpha_{i}} \frac{\partial}{\partial \alpha_{j}}
$$

In other words, $C_{\iota}(y)$ satisfies $\Delta_{y} C_{\iota}(y)=\lambda C_{\iota}(y)$ for some $\lambda$.
(iii) As $\iota$ varies over all partitions of $k$, the zonal polynomials have unit coefficients in the expansion of $(\operatorname{tr}[y])^{k}$, i.e.,

$$
\sum_{|c|=k} C_{\iota}(y)=(\operatorname{tr}[y])^{k}=\left(\alpha_{1}+\cdots+\alpha_{d}\right)^{k}
$$

where $|\iota|$ denotes the sum of parts of $\iota$, i.e. $|\iota|=k_{1}+\cdots+k_{d}$.
Definition A.3. The hypergeometric functions of matrix argument are given by

$$
{ }_{p} F_{q}\left(a_{1}, \cdots, a_{p} ; b_{1}, \cdots, b_{q} ; y\right)=\sum_{k=0}^{\infty} \sum_{|\iota|=k} \frac{\left(a_{1}\right)_{\iota} \cdots\left(a_{p}\right)_{\iota}}{\left(b_{1}\right)_{\iota} \cdots\left(b_{p}\right)_{\iota}} \frac{C_{\iota}(y)}{k!}
$$

where the generalized hypergeometric coefficient $(a)_{\iota}$ for a partition $\iota=\left(k_{1}, \cdots, k_{d}\right)$ is given by

$$
(a)_{\iota}=\prod_{j=1}^{d}\left(a-\frac{1}{2}(j-1)\right)_{k_{j}}
$$

and $(a)_{k}=a(a+1) \cdots(a+k-1),(a)_{0}=1$. Here, the argument of the function is a complex symmetric $d \times d$ matrix and the parameters $a_{j}, b_{j}$ are arbitrary complex numbers. No denominator parameter $b_{j}$ is allowed to be zero or an integer or half-integer $\leq \frac{1}{2}(d-1)$, otherwise some of the denominators in the series will vanish.

Note that in case $d=1$, there is only one partition of $k$ into not more than $d$, namely $(k)$. Therefore, the hypergeometric function of matrix arguments becomes to the usual hypergeometric function of real variables.

Likewise the noncentral chi-square distributions, the noncentral Wishart distribution with integer degrees of freedom is defined by multiplying normal random vectors with themselves. Consider $\delta(\in \mathbb{N})$ independent random vectors of $\mathbb{R}^{d}$, denoted by $Z_{1}, \cdots, Z_{\delta}$, with multivariate normal distributions with means $\mu_{1}, \cdots, \mu_{\delta}$ and the common nonsingular covariance matrix $C$. The distribution of the random matrix

$$
W=\sum_{k=1}^{\delta} Z_{k} Z_{k}^{\top}
$$

is called the noncentral Wishart distribution with $\delta$ degrees of freedom, covariance matrix $C$, and matrix of noncentrality parameter $\Omega=C^{-1} \sum_{k=1}^{\delta} \mu_{k} \mu_{k}^{\top}$. If $\delta \geq d$, then $W$ has a probability density function which is of the form:

$$
\begin{equation*}
\frac{(\operatorname{det}[y])^{(\delta-d-1) / 2}}{2^{d \delta / 2} \Gamma_{d}\left(\frac{1}{2} \delta\right)(\operatorname{det}[C])^{\delta / 2}} \exp \left\{-\frac{1}{2} \operatorname{tr}\left[C^{-1} y+\Omega\right]\right\}_{0} F_{1}\left(\frac{1}{2} \delta ; \frac{1}{4} \Omega C^{-1} y\right), \quad y \in S_{d}^{++} \tag{21}
\end{equation*}
$$

The function (21) is still a density function when $\delta$ is any real number greater than $d-1$. This observation make it possible to extend the notion of the noncentral Wishart distribution to non-integer degrees of freedom.

Definition A.4. Suppose that $\delta>d-1, C \in S_{d}^{++}$, and $\Omega$ is a $d \times d$ matrix such that $C \Omega$ is symmetric positive semidefinite. If a $d \times d$ symmetric positive definite random matrix $W$ has the probability density function (21), then $W$ is said to have the noncentral Wishart distribution with $\delta$ degrees of freedom, covariance matrix $C$, and matrix of noncentrality parameter $\Omega$. We will say $W$ has $\mathcal{W}_{d}(\delta, C, \Omega)$ distribution.
Proposition A. 5 (Theorem 3.5.3 in Gupta and Nagar [23]). Suppose $W$ has noncentral Wishart distribution $\mathcal{W}_{d}(\delta, C, \Omega)$. Then its Laplace transform is given by

$$
\mathbb{E}\left[e^{-\operatorname{tr}(\vartheta W)}\right]=\operatorname{det}\left[I_{d}+2 \vartheta C\right]^{-\delta / 2} \exp \left\{-\operatorname{tr}\left[\vartheta\left(I_{d}+2 C \vartheta\right)^{-1} C \Omega\right]\right\},
$$

for any symmetric positive semidefinite matrix $\vartheta$.

## A. 2 Wishart Processes with Time-varying Linear Drift

In this section, we slightly extend the notion of Wishart processes in order to compute the conditional Laplace transform of log-price given volatility level.

A symmetric positive semidefinite matrix valued stochastic process $X$ is called a Wishart process with time-varying linear drift if it is a weak solution of the following stochastic differential equation

$$
\begin{equation*}
d X_{t}=\left(\delta \Sigma^{\top} \Sigma+H(t) X_{t}+X_{t} H(t)^{\top}\right) d t+\sqrt{X_{t}} d W_{t} \Sigma+\Sigma^{\top} d W_{t}^{\top} \sqrt{X_{t}}, \text { with } X_{0}=x \tag{22}
\end{equation*}
$$

where $\delta \geq d-1, \Sigma$ is a $d \times d$ matrix, $x$ is a symmetric positive semidefinite matrix, $H(\cdot)$ is a $d \times d$ matrix valued continuous function, and $W$ is a standard $d \times d$ matrix Brownian motion. Wishart process with time-varying linear drift has noncentral Wishart marginal distributions.
Proposition A.6. Let $X$ be a Wishart process with time-varying linear drift which solves (22). Then $X_{T}$ has noncentral Wishart distribution $\mathcal{W}_{d}\left(\delta, V(0), V(0)^{-1} \Psi(0)^{\top} x \Psi(0)\right)$, where $V(t)$ and $\Psi(t)$ are solutions of following system of ordinary differential equations

$$
\left\{\begin{aligned}
\frac{d}{d t} \Psi(t) & =-H(t)^{\top} \Psi(t), \\
\frac{d}{d t} V(t) & =-\Psi(t)^{\top} \Sigma^{\top} \Sigma \Psi(t),
\end{aligned}\right.
$$

with the terminal value $\Psi(T)=I_{d}$ and $V(T)=0$.
Proof. Using standard argument(e.g., see Appendix B of Gourieroux and Sufana [22]), one may prove that for a symmetric positive semidefinite matrix $\vartheta$,

$$
\mathbb{E}_{x}\left[e^{-\operatorname{tr}\left(\vartheta X_{T}\right)}\right]=e^{-\hat{\phi}(0, \vartheta)-\operatorname{tr}(\hat{\psi}(0, \vartheta) x)},
$$

where $\hat{\phi}$ and $\hat{\psi}$ are the solution of following equations

$$
\left\{\begin{array}{l}
\partial_{t} \hat{\psi}(t, \vartheta)=2 \hat{\psi}(t, \vartheta) \Sigma^{\top} \Sigma \hat{\psi}(t, \vartheta)-H(t)^{\top} \hat{\psi}(t, \vartheta)-\hat{\psi}(t, \vartheta) H(t), \\
\partial_{t} \hat{\phi}(t, \vartheta)=-\delta \operatorname{tr}\left[\hat{\psi}(t, \vartheta) \Sigma^{\top} \Sigma\right],
\end{array}\right.
$$

with terminal values $\hat{\phi}(T, \vartheta)=0$ and $\hat{\psi}(T, \vartheta)=\vartheta$. Using differentiation rules $\frac{d}{d t} A(t)^{-1}=$ $-A(t)^{-1} \frac{d}{d t} A(t) A(t)^{-1}$ and $\frac{d}{d t} \ln \operatorname{det}[A(t)]=\operatorname{tr}\left[A(t)^{-1} \frac{d}{d t} A(t)\right]$, one may check that

$$
\left\{\begin{array}{l}
\hat{\psi}(t, \vartheta)=\Psi(t) \vartheta\left(I_{d}+2 V(t) \vartheta\right)^{-1} \Psi(t)^{\top}, \\
\hat{\phi}(t, \vartheta)=\frac{\delta}{2} \ln \operatorname{det}\left[I_{d}+2 \vartheta V(t)\right],
\end{array}\right.
$$

solves the above system of differential equations. Therefore,

$$
\mathbb{E}_{x}\left[e^{-\operatorname{tr}\left(\vartheta X_{T}\right)}\right]=\operatorname{det}\left[I_{d}+2 \vartheta V(0)\right]^{-\delta / 2} \exp \left\{-\operatorname{tr}\left[\vartheta\left(I_{d}+2 V(0) \vartheta\right)^{-1} \Psi(0)^{\top} x \Psi(0)\right]\right\}
$$

Since Laplace transform uniquely characterizes a distribution, $X_{T}$ has noncentral Wishart distribution $\mathcal{W}_{d}\left(\delta, V(0), V(0)^{-1} \Psi(0)^{\top} x \Psi(0)\right)$ by Proposition A.5.

## A. 3 Details of calculation of (12)

In this section, we provide details of the calculation of the conditional characteristic function (12). The first equation in the system (11) is the classical Riccati equation, and its closed-form solution is well-known (e.g., see Section 10.7.2 of Filipović [17]). In particular,

$$
\begin{aligned}
\psi(t, u) & =-\frac{u(u+1)\left(e^{\eta(u)(T-t)}-1\right)}{\eta(u)\left(e^{\eta(u)(T-t)}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u)(T-t)}-1\right)} \\
\int_{t}^{T} \psi(s, u) d s & =-\frac{2}{\sigma^{2}} \log \left(\frac{2 \eta(u) e^{\frac{1}{2}(\eta(u)+\kappa+u \sigma \rho)(T-t)}}{\eta(u)\left(e^{\eta(u)(T-t)}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u)(T-t)}-1\right)}\right)
\end{aligned}
$$

where $\eta(u)=\sqrt{(\kappa+u \sigma \rho)^{2}-\sigma^{2} u(u+1)}$. It follows that

$$
e^{-\phi(t, u)}=\left(\frac{2 \eta(u) e^{\frac{1}{2}(\eta(u)+\kappa+u \sigma \rho)(T-t)}}{\eta(u)\left(e^{\eta(u)(T-t)}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u)(T-t)}-1\right)}\right)^{\delta / 2} e^{-u r(T-t)}
$$

The solution of the third linear equation is

$$
\begin{align*}
\Psi(t, u) & =\exp \left\{-\frac{1}{2} \int_{t}^{T}\left(\kappa+u \sigma \rho+\sigma^{2} \psi(s, u)\right) d s\right\} \\
& =e^{-\frac{1}{2}(\kappa+u \sigma \rho)(T-t)} \exp \left\{-\frac{\sigma^{2}}{2} \int_{t}^{T} \psi(s, u) d s\right\} \\
& =e^{-\frac{1}{2}(\kappa+u \sigma \rho)(T-t)} \frac{2 \eta(u) e^{\frac{1}{2}(\eta(u)+\kappa+u \sigma \rho)(T-t)}}{\eta(u)\left(e^{\eta(u)(T-t)}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u)(T-t)}-1\right)} \\
& =\frac{2 \eta(u) e^{\frac{1}{2} \eta(u)(T-t)}}{\eta(u)\left(e^{\eta(u)(T-t)}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u)(T-t)}-1\right)} \tag{23}
\end{align*}
$$

A direct integration shows that

$$
\begin{equation*}
V(t, u)=\frac{1}{2} \frac{\sigma^{2}\left(e^{\eta(u)(T-t)}-1\right)}{\eta(u)\left(e^{\eta(u)(T-t)}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u)(T-t)}-1\right)} \tag{24}
\end{equation*}
$$

We divide (23) by (24) to have

$$
\frac{\Psi(0, u)}{V(0, u)}=\frac{4 \eta(u) e^{0.5 \eta(u) T}}{\sigma^{2}\left(e^{\eta(u) T}-1\right)}=\frac{4 \eta(u) e^{-0.5 \eta(u) T}}{\sigma^{2}\left(1-e^{-\eta(u) T}\right)}
$$

In particular,

$$
\frac{\Psi(0,0)}{V(0,0)}=\frac{4 \kappa e^{-0.5 \kappa T}}{\sigma^{2}\left(1-e^{-\kappa T}\right)}
$$

Observe that the relation between $\phi(t, u)$ and $\Psi(t, u)$ :

$$
e^{-\phi(t, u)}=(\Psi(t, u))^{\delta / 2} \exp \left\{\frac{\delta}{4}(\kappa+u \sigma \rho)(T-t)-u r(T-t)\right\}
$$

Remind the relationship between the hypergeometric functions and the modified Bessel functions:

$$
{ }_{0} F_{1}\left(\nu+1 ; \frac{1}{4} x^{2}\right)=(x / 2)^{-\nu} \Gamma(\nu+1) I_{\nu}(x) .
$$

Using the identities above, we found that

$$
\begin{aligned}
& \left(\frac{V(0,0)}{V(0, u)}\right)^{\delta / 2} \exp \{-\phi(0, u)\} \frac{{ }_{0} F_{1}\left(\frac{1}{2} \delta ; \frac{1}{4}\left(\frac{\Psi(0, u)}{V(0, u)} \sqrt{x x_{T}}\right)^{2}\right)}{F_{1}\left(\frac{1}{2} \delta ; \frac{1}{4}\left(\frac{\Psi(0,0)}{V(0,0)} \sqrt{x x_{T}}\right)^{2}\right)} \\
& =\left(\frac{V(0,0)}{V(0, u)}\right)^{\delta / 2}\left(\frac{\Psi(0, u)}{V(0, u)}\right)^{-\delta / 2+1}\left(\frac{\Psi(0,0)}{V(0,0)}\right)^{\delta / 2-1} \\
& \quad \times(\Psi(0, u))^{\delta / 2} \exp \left\{\frac{\delta}{4}(\kappa+u \sigma \rho) T-u r T\right\} \frac{I_{0.5 \delta-1}\left(\sqrt{x x_{T}} \frac{\Psi(0, u)}{V(0, u)}\right)}{I_{0.5 \delta-1}\left(\sqrt{x x_{T}} \frac{\Psi(0,0)}{V(0,0)}\right)} \\
& =V(0,0) \frac{\Psi(0, u)}{V(0, u)} \exp \left\{\frac{1}{2} \kappa T-u\left(r-\frac{\kappa \theta \rho}{\sigma}\right) T\right\} \frac{I_{0.5 \delta-1}\left(\sqrt{x x_{T}} \frac{\Psi(0, u)}{V(0, u)}\right)}{I_{0.5 \delta-1}\left(\sqrt{x x_{T}} \frac{\Psi(0,0)}{V(0,0)}\right)} \\
& =\frac{\eta(u)\left(1-e^{-\kappa T}\right)}{\kappa\left(1-e^{-\eta(u) T}\right)} \exp \left\{-u\left(r-\frac{\kappa \theta \rho}{\sigma}\right) T-\frac{1}{2}(\eta(u)-\kappa) T\right\} \frac{I_{0.5 \delta-1}\left[\sqrt{x x_{T}} \frac{4 \eta(u) e^{-0.5 \eta(u) T}}{\sigma^{2}\left(1-e^{-\eta(u) T)}\right]}\right.}{I_{0.5 \delta-1}\left[\sqrt{x x_{T}} \frac{4 e^{-0.5 \kappa T}}{\sigma^{2}\left(1-e^{-\kappa T}\right)}\right]} .
\end{aligned}
$$

Recall that $\eta(u)^{2}=(\kappa+u \sigma \rho)^{2}-\sigma^{2} u(u+1)$. Using this identity, we have

$$
\begin{aligned}
& 2 \psi(0, u)+\frac{\Psi(0, u)^{2}}{V(0, u)}=\frac{-2 u(u+1)\left(e^{\eta(u) T}-1\right)}{\eta(u)\left(e^{\eta(u) T}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u) T}-1\right)} \\
& \quad+\frac{4 \eta(u) e^{-0.5 \eta(u) T}}{\sigma^{2}\left(1-e^{-\eta(u) T}\right)} \frac{2 \eta(u) e^{0.5 \eta(u) T}}{\eta(u)\left(e^{\eta(u) T}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u) T}-1\right)} \\
& =\frac{-2 \sigma^{2} u(u+1)\left(e^{\eta(u) T}-1\right)\left(1-e^{-\eta(u) T}\right)+8 \eta(u)^{2}}{\sigma^{2}\left(1-e^{-\eta(u) T}\right)\left(\eta(u)\left(e^{\eta(u) T}+1\right)+(\kappa+u \sigma \rho)\left(e^{\eta(u) T}-1\right)\right)} \\
& =\frac{-2 \sigma^{2} u(u+1)\left(1-e^{-\eta(u) T}\right)^{2}+8 \eta(u)^{2} e^{-\eta(u) T}}{\sigma^{2}\left(1-e^{-\eta(u) T)\left(\eta(u)\left(1+e^{-\eta(u) T}\right)+(\kappa+u \sigma \rho)\left(1-e^{-\eta(u) T}\right)\right)}\right.} \\
& =\frac{-2 \sigma^{2} u(u+1)-4 \sigma^{2} u(u+1) e^{-\eta(u) T}-2 \sigma^{2} u(u+1) e^{-2 \eta(u) T}+8(\kappa+u \sigma \rho)^{2} e^{-\eta(u) T}}{\sigma^{2}\left(1-e^{-\eta(u) T}\right)\left(\eta(u)\left(1+e^{-\eta(u) T}\right)+(\kappa+u \sigma \rho)\left(1-e^{-\eta(u) T}\right)\right)} \\
& =\frac{-2 \sigma^{2} u(u+1)\left(1+e^{-\eta(u) T}\right)^{2}+8(\kappa+u \sigma \rho)^{2} e^{-\eta(u) T}}{\sigma^{2}\left(1-e^{-\eta(u) T}\right)\left(\eta(u)\left(1+e^{-\eta(u) T}\right)+(\kappa+u \sigma \rho)\left(1-e^{-\eta(u) T}\right)\right)} \\
& =\frac{2}{\sigma^{2}} \frac{\left(\eta(u)^{2}-(\kappa+u \sigma \rho)^{2}\right)\left(1+e^{-\eta(u) T}\right)^{2}+4(\kappa+u \sigma \rho)^{2} e^{-\eta(u) T}}{\left(1-e^{-\eta(u) T}\right)\left(\eta(u)\left(1+e^{-\eta(u) T}\right)+(\kappa+u \sigma \rho)\left(1-e^{-\eta(u) T}\right)\right)} \\
& =\frac{2}{\sigma^{2}} \frac{\eta(u)^{2}\left(1+e^{-\eta(u) T}\right)^{2}-(\kappa+u \sigma \rho)^{2}\left(1-e^{-\eta(u) T}\right)^{2}}{\left(1-e^{-\eta(u) T}\right)\left(\eta(u)\left(1+e^{-\eta(u) T}\right)+(\kappa+u \sigma \rho)\left(1-e^{-\eta(u) T}\right)\right)} \\
& =\frac{2}{\sigma^{2}} \frac{\eta(u)\left(1+e^{-\eta(u) T}\right)-(\kappa+u \sigma \rho)\left(1-e^{-\eta(u) T}\right) .}{1-e^{-\eta(u) T}} .
\end{aligned}
$$

And

$$
\frac{1}{V(0,0)}=\frac{2}{\sigma^{2}} \frac{2 \kappa}{1-e^{-\kappa T}} \quad \text { and } \quad \frac{\Psi(0,0)^{2}}{V(0,0)}=\frac{2}{\sigma^{2}} \frac{2 \kappa e^{-\kappa T}}{1-e^{-\kappa T}} .
$$

By substituting above quantities into the formula (7), we found the closed-form expression (12) for the conditional Laplace transform of the Heston model.

## A. 4 Proof of Proposition 4.1

We start with a simple lemma.

Lemma A.7. Let $b>\frac{1}{2}(d-1)$. Suppose $\hat{\imath}(k)$ is the smallest partition among the partitions of $k$ into not more than $d$ parts. Then

$$
0<(b)_{\hat{\iota}(k)} \leq(b)_{\iota}
$$

for all partitions $\iota$ of $k$ into not more than $d$ parts.
Proof. Fix $k$ and we simplify the notation $\hat{\iota}(k)$ as $\hat{\iota}$. The factors of the hypergeometric coefficients are of the form

$$
0<b-\frac{1}{2}(d-1) \leq b+l-\frac{1}{2}(j-1) \leq b+k-1,
$$

for $j=1, \cdots, d$, and $l=0, \cdots, k_{j}-1$. Thus, we have $(b)_{\hat{\imath}}>0$. For a partition $\iota=\left(k_{1}, \ldots, k_{d}\right)$, define $\alpha(\iota):=\max \left\{k_{i}-k_{j}: 1 \leq i<j \leq d\right\}$ and $\beta(\iota):=\min \left\{j-i: k_{i}-k_{j}=\alpha(\iota), i<j\right\}$. Assume that $\iota$ is not the smallest one. Then $\alpha(\iota) \geq 2$. Let $\alpha(\iota)=k_{i^{*}}-k_{j^{*}}$ and $\beta(\iota)=j^{*}-i^{*}$. We define a new partition $\iota^{\prime}=\left(k_{1}, \ldots, k_{i^{*}-1}, k_{i^{*}}-1, k_{i^{*}+1}, \ldots, k_{j^{*}-1}, k_{j^{*}}+1, k_{j^{*}+1}, \ldots, k_{d}\right)$. Then $(b)_{\iota^{\prime}}=(b)_{\iota} \times \frac{b+k_{j^{*}}-\left(j^{*}+1\right) / 2}{b+k_{i^{*}}-\left(i^{*}+1\right) / 2}<(b)_{\iota}$. For the new partition $\iota^{\prime}$, if we have $\alpha\left(\iota^{\prime}\right)=\alpha(\iota)$, then $\beta\left(\iota^{\prime}\right)=\beta(\iota)+2$. Since $\beta(\cdot) \leq d-1$, we should have a partition with decreased $\alpha$ value after some iterations. Hence, we obtain a partition $\hat{\iota}$ with $\alpha(\hat{\imath}) \leq 1$ in finite steps, which is the smallest one. Since the new partitions always have smaller hypergeometric coefficient values, we conclude $(b)_{\hat{\iota}}<(b)_{\iota}$ for any non-smallest partition $\iota$.

For reader's better understanding, we give an example which illustrates the idea of the above proof. Consider the case $d=5$ and $k=8$. In this case, the smallest partition is

$$
\hat{\iota}=\hat{\iota}(8)=(2,2,2,1,1) .
$$

We start with a partition which is not the smallest, say $\iota=(4,3,1,0,0)$. From this partition, we successively looking for a new partition with a smaller hypergeometric coefficients until we arrive at the smallest one:

$$
\begin{aligned}
& (4,3,1,0,0), \alpha(\iota)=4, \beta(\iota)=3 \quad(3,3,1,1,0), \alpha(\iota)=3, \beta(\iota)=3
\end{aligned}
$$

$$
\begin{aligned}
& (3,2,1,1,1), \alpha(\iota)=2, \beta(\iota)=2 \\
& (2,2,2,1,1), \alpha(\hat{\imath})=1, \beta(\hat{\imath})=1
\end{aligned}
$$

At each stage, the values of the hypergeometric coefficients decrease, and the smallest partition has the minimal hypergeometric coefficient. The next lemma is about comparisons between hypergeometric coefficients of partitions with different sizes.

Lemma A.8. Let $b>\frac{1}{2}(d-1)$ and $k \geq d$. Then

$$
(b)_{\hat{\iota}(k)}<(b)_{\hat{\iota}(k+1)} .
$$

Proof. This lemma is a simple consequence of the recurrence relation (20). We write $\hat{\iota}(k)=$ $\left(k_{1}, \cdots, k_{d}\right)$. Since $k \geq d$, we have $k_{j} \geq 1$ for all $j=1, \cdots, d$. Therefore,

$$
1 \leq k_{j}<b+k_{j}-\frac{1}{2}(d-1) \leq b+k_{j}-\frac{1}{2}(j-1)
$$

for all $j=1, \cdots, d$. This observation and the recurrence relation (20) complete the proof.
Now we consider the zonal polynomials. Zonal polynomials are polynomials of eigenvalues of the matrix, and their coefficients are all nonnegative (see the recurrence relation (14) in page 234 of Muirhead [29]). Therefore,

$$
\left|C_{\iota}\left(\alpha_{1}, \cdots, \alpha_{d}\right)\right| \leq C_{\iota}\left(\left|\alpha_{1}\right|, \cdots,\left|\alpha_{d}\right|\right)
$$

for all partitions $\iota$, and all complex numbers $\alpha_{1}, \cdots, \alpha_{d}$.
Proof of Proposition 4.1. The proof is a combination of the inequalities which are established in Lemma A. 7 and A.8. Observe that

$$
\begin{aligned}
& \left|{ }_{0} F_{1}(b ; y)-{ }_{0}^{m} F_{1}(b ; y)\right|=\left|\sum_{k=m+1}^{\infty} \sum_{|\iota|=k} \frac{1}{(b)_{\iota}} \frac{C_{\iota}(y)}{k!}\right| \leq \sum_{k=m+1}^{\infty} \sum_{|\iota|=k} \frac{1}{(b)_{\iota}} \frac{\left|C_{\iota}(y)\right|}{k!} \\
& \quad \leq \sum_{k=m+1}^{\infty} \frac{1}{k!(b)_{\hat{\iota}(k)}} \sum_{|\iota|=k} C_{\iota}\left(\left|\alpha_{1}\right|, \cdots,\left|\alpha_{d}\right|\right) \leq \frac{1}{(b)_{\hat{\iota}(m+1)}} \sum_{k=m+1}^{\infty} \frac{1}{k!} \sum_{|\iota|=k} C_{\iota}\left(\left|\alpha_{1}\right|, \cdots,\left|\alpha_{d}\right|\right) .
\end{aligned}
$$

By definition, zonal polynomials are normalized so that

$$
\sum_{|\iota|=k} C_{\iota}\left(\beta_{1}, \cdots, \beta_{d}\right)=\left(\beta_{1}+\cdots+\beta_{d}\right)^{k}
$$

for all complex numbers $\beta_{1}, \cdots, \beta_{d}$ (see Definition 7.2 .1 of Muirhead [29]). Therefore, we have

$$
\begin{aligned}
\left|{ }_{0} F_{1}(b ; y)-{ }_{0}^{m} F_{1}(b ; y)\right| & \leq \frac{1}{(b)_{\hat{\iota}(m+1)}} \sum_{k=m+1}^{\infty} \frac{1}{k!}\left(\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|\right)^{k} \\
& \leq \frac{1}{(b)_{\hat{\iota}(m+1)}} \sum_{k=0}^{\infty} \frac{1}{(k+m+1)!}\left(\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|\right)^{k+m+1} \\
& \leq \frac{\left(\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|\right)^{m+1}}{(m+1)!(b)_{\hat{\iota}(m+1)}} \sum_{k=0}^{\infty} \frac{1}{k!}\left(\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|\right)^{k} \\
& =\frac{\left(\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|\right)^{m+1}}{(m+1)!(b)_{\hat{\iota}(m+1)}} e^{\left|\alpha_{1}\right|+\cdots+\left|\alpha_{d}\right|}
\end{aligned}
$$

## References

[1] J. Abate and W. Whitt. The Fourier-series method for inverting transforms of probability distributions. Queueing Systems Theory Appl., 10(1-2):5-87, 1992.
[2] A. Ahdida and A. Alfonsi. Exact and high order discretization schemes for Wishart processes and their affine extensions. Ann. Appl. Probab., 23(3):1025-1073, 2013.
[3] D. E. Amos. Algorithm 644: A portable package for bessel functions of a complex argument and nonnegative order. ACM Trans. Math. Softw., 12(3):265-273, Sept. 1986.
[4] D. Bates. Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. Rev. Financ. Stud., 9(1):69-107, 1996.
[5] D. S. Bates. Post-' 87 crash fears in the s\&p 500 futures option market. Journal of Econometrics, 94(1-2):181-238, 2000.
[6] A. Benabid, H. Bensusan, N. El Karoui, et al. Wishart stochastic volatility: Asymptotic smile and numerical framework. preprint, 2008.
[7] M. Broadie and Ö. Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. Oper. Res., 54(2):217-231, 2006.
[8] M.-F. Bru. Wishart processes. J. Theoret. Probab., 4(4):725-751, 1991.
[9] P. Carr, D. B. Madan, and R. H. Smith. Option valuation using the fast fourier transform. Journal of Computational Finance, 2:61-73, 1999.
[10] P. Carr and L. Wu. Stochastic skew in currency options. Journal of Financial Economics, 86(1):213-247, 2007.
[11] P. Christoffersen, S. Heston, and K. Jacobs. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. Manage. Sci., 55(12):1914-1932, Dec. 2009.
[12] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann. Affine processes on positive semidefinite matrices. Ann. Appl. Probab., 21(2):397-463, 2011.
[13] J. da Fonseca and M. Grasselli. Riding on the smiles. Quant. Finance, 11(11):1609-1632, 2011.
[14] J. Da Fonseca, M. Grasselli, and C. Tebaldi. A multifactor volatility Heston model. Quant. Finance, 8(6):591-604, 2008.
[15] J. R. Dormand and P. J. Prince. A family of embedded Runge-Kutta formulae. J. Comput. Appl. Math., 6(1):19-26, 1980.
[16] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jumpdiffusions. Econometrica, 68(6):1343-1376, 2000.
[17] D. Filipović. Term-structure models. Springer Finance. Springer-Verlag, Berlin, 2009. A graduate course.
[18] P. Gauthier and D. Possamaï. Prices expansion in the wishart model. The IUP Journal of Computational Mathematics, 4(1):44-71, 2011.
[19] P. Gauthier and D. Possamaï. Efficient simulation of the wishart model. The IUP Journal of Computational Mathematics, 5(1):14-58, 2012.
[20] P. Glasserman. Monte Carlo methods in financial engineering, volume 53 of Applications of Mathematics (New York). Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.
[21] L. J. Gleser. A canonical representation for the noncentral Wishart distribution useful for simulation. J. Amer. Statist. Assoc., 71(355):690-695, 1976.
[22] C. Gourieroux and R. Sufana. Derivative pricing with Wishart multivariate stochastic volatility. J. Bus. Econom. Statist., 28(3):438-451, 2010.
[23] A. K. Gupta and D. K. Nagar. Matrix variate distributions, volume 104 of Chapman \& Hall/CRC Monographs and Surveys in Pure and Applied Mathematics. Chapman \& Hall/CRC, Boca Raton, FL, 2000.
[24] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies, 6(2):327-343, 1993.
[25] P. Koev. http://math.mit.edu/~plamen/software/.
[26] P. Koev and A. Edelman. The efficient evaluation of the hypergeometric function of a matrix argument. Math. Comp., 75(254):833-846, 2006.
[27] A. M. Kshirsagar. Bartlett decomposition and Wishart distribution. Ann. Math. Statist., 30:239-241, 1959.
[28] M. Leippold and F. Trojani. Asset pricing with matrix jump diffusions. Available at SSRN 1274482, 2008.
[29] R. J. Muirhead. Aspects of multivariate statistical theory. John Wiley \& Sons Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.
[30] D. Revuz and M. Yor. Continuous martingales and Brownian motion, volume 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, third edition, 1999.


[^0]:    *National Institute for Mathematical Sciences, Republic of Korea, E-mail: ckang@nims.re.kr
    ${ }^{\dagger}$ Department of Mathematical Sciences, KAIST, Republic of Korea, E-mail: wanmo.kang@kaist.edu

[^1]:    ${ }^{1}$ The MATLAB function ode45 is based on an explicit Runge-Kutta $(4,5)$ formula of Dormand and Prince 15 .

[^2]:    ${ }^{2}$ In their paper, the calibrated $\delta$ for DAX index is 0.5776 , and it violate the assumption $\delta>d-1$. Indeed, in that case, the Wishart process $X$ is no longer an affine process in the sense of Cuchiero et. al. [12].

[^3]:    ${ }^{3}$ The MATLAB funciton besseli computes the modified Bessel function of the first kind $I_{\nu}(\cdot)$ using the algorithm of Amos 3].

[^4]:    ${ }^{4}$ In our method, the $h$ and $N$ are automatically chosen by (18)

