

On Randomization of Affine Diffusion Processes with Application to Pricing of Options on VIX and S&P 500

LECH A. GRZELAK^{a,b,*}

^a*Mathematical Institute, Utrecht University, Utrecht, the Netherlands*

^b*Financial Engineering, Rabobank, Utrecht, the Netherlands*

Abstract

The class of Affine (Jump) Diffusion [8] (AD) has, due to its closed form characteristic function (ChF), gained tremendous popularity among practitioners and researchers. However, there is clear evidence that a linearity constraint is insufficient for precise and consistent option pricing. Any non-affine model must pass the strict requirement of quick calibration- which is often challenging. We focus here on Randomized AD (RAnD) models, i.e., we allow for exogenous stochasticity of the model parameters. Randomization of a pricing model occurs outside the affine model and, therefore, forms a generalization that relaxes the affinity constraints. The method is generic and can apply to any model parameter. It relies on the existence of moments of the so-called randomizer- a random variable for the stochastic parameter. The RAnD model allows flexibility while benefiting from fast calibration and well-established, large-step Monte Carlo simulation, often available for AD processes. The article will discuss theoretical and practical aspects of the RAnD method, like derivations of the corresponding ChF, simulation, and computations of sensitivities. We will also illustrate the advantages of the randomized stochastic volatility models in the consistent pricing of options on the S&P 500 and VIX.

Keywords: Randomized Affine-Diffusion, RAnD Method, Stochastic Parameters, Heston, Bates, VIX, S&P 500, Short Maturity Options, Stochastic Collocation, Quantization.

1. Introduction

The class of Affine Jump Diffusion, which we abbreviate by AD, asset price models constitutes a family of processes that, under certain linearity conditions, allow for a semi-closed form characteristic function (ChF) [8]. Once the ChF is derived, it can be used, via Fourier inversion, for pricing certain financial derivatives, thus facilitating fast asset price model calibration. Over many years, various affine models have been studied, leading to significant contributions like efficient large-step Monte Carlo simulations, PDE discretizations, quick calibrations or even analytical, perturbation-based approximations for implied volatilities. These accomplishments popularized the class of AD processes.

Although the high speed of “regular” option pricing is a necessary condition that allows for calibration, it is insufficient for a model to be used for exotic derivative pricing. The main limitation of classical dynamic stochastic volatility models driven solely by possibly correlated Brownian motion is the limited flexibility for short maturities, i.e., the slope of the skew is insufficient to fit the market data. By adding jumps, this effect can, to some extent, be mitigated; however, it is often unsatisfactory, especially in a volatile market. The class of affine models is inherently inflexible, with limited possibilities for model extensions. In particular, with the rise in popularity of products such as options on volatility, it is evident that the standard affine, stochastic volatility models are insufficient for accurate and consistent option

*Corresponding author at Mathematical Institute, Utrecht University, Utrecht, the Netherlands.

Email address: L.A.Grzelak@uu.nl (LECH A. GRZELAK)

The views expressed in this paper are the personal views of the authors and do not necessarily reflect the views or policies of their current or past employers. The authors have no competing interests.

pricing. This observation has initiated a *gold rush* to discover new classes of models that would address this pricing problem.

One possible route to extend the existing affine models was suggested by P. Carr and L. Wu in [7], where the problem of insufficient skew was reported, and the remedy of the form of parameter randomization was suggested: “Starting from the jump-diffusion stochastic volatility model of Bates [2], it would be tempting to try to capture stochastic skewness by randomizing the mean jump size parameter or the correlation parameter (...) However, randomizing either parameter is not amenable to analytic solution techniques that greatly aid econometric estimation.” Although the authors considered randomization as a technique to incorporate more flexibility into a stochastic model, thus a way to improve the calibration quality, the concept of randomizing is more fundamental, i.e., it represents the incorporation of the uncertainty of potentially hidden states that are not adequately captured by deterministic parameters. In this article, we focus on the randomization of the class of affine models and allow for additional flexibility by letting the model parameters be stochastic. In particular, we will randomize the parameters of the Bates model and examine their impact on the shapes of implied volatilities.

Attempts to randomize individual model parameters are already known in the literature. Starting from the simplest model, the Black-Scholes model with the randomized volatility coefficient- based on a discrete distribution- was presented in [4]. An extension to continuous distributions was analyzed in [19]. In their work, the authors discussed various stochastic variants for the volatility and their level of *explosiveness*- this led to several valuable expansions for short maturity options. In [20, 25], for example, it was shown that randomization of the initial variance of the Heston model significantly increased the steepness of the implied volatility, primarily when the randomizer is defined in an unbounded domain. This idea was further studied in [24] where the term *stationary*, to describe random initial parameter, was used. In [27] the concept of Recursive Marginal Quantization for one-dimensional diffusion discretization schemes was discussed, with the extension to higher order schemes in [29] and in [6] a method to discretize a stochastic process via quantization optimally was given. Another randomization attempt was presented in [10], where the volatility-of-volatility parameter in the Heston model was assumed to be stochastic, and approximations, based on perturbation, were derived. In the case of a randomizer with a finite number of states, one may think of a regime-switching model [13]. Such a setup was discussed [28], where the hidden regime continuous-time Markov chain was considered.

Customarily, an extension with a stochastic parameter would require that the model satisfies the linearity conditions- this, however, does not need to be the case under the framework proposed here. Instead, we only require the affinity to hold for a particular realization of a stochastic parameter. We build an external layer over the class of AD models and allow the parameter to be stochastic in that layer. In a standard approach, this would imply an infinite number of affine models, making this method impractical. The solution to this problem is at the heart of this article, in which we develop a method for constructing an appropriate ChF based on only a few, *critical* parameter realizations. Selecting these points is based on a quadrature rule based on the moments of the stochastic parameter. The procedure of building a randomized ChF, we call the RAnD method, which corresponds to “Randomized Affine-Diffusion”.

The presented structure is generic, and we believe it may be applied in various asset classes. To highlight its flexibility in modern finance, we apply this technique to the challenging problem of consistent pricing options on S&P 500 and VIX. This pricing problem is not new, and many innovative solutions have already been proposed.

The major challenge regarding this pricing task is to find a stochastic model that can sufficiently explain both the negative skew for equity options on the S&P 500 index and the positive skew for options on the VIX. In addition, different regime settings for implied volatility cause a calibration dilemma where the volatility of the volatility parameter under the Heston model is either too large for VIX or too small for S&P 500 (or both). The first attempts to address this pricing problem aimed to introduce additional degrees of freedom in multidimensional models. In [30], for example, the variance process in the Bates model was enriched with an additional jump process. In [1], the so-called 3/2 process and inclusion of jumps for the stock process could capture the observed upward skew for VIX, although the model was unable to calibrate to both derivatives simultaneously. More recently, in [23], another attempt to include jumps in both

returns and volatility added to stochastic volatility was discussed. The authors concluded that more flexibility, especially with a relaxed Feller condition, is necessary to provide a sufficient fit. A promising extension through the Hawkes process for the jumps was discussed [22]- however, additional degrees of freedom are insufficient for calibration purposes. A breakthrough in volatility modelling came with the so-called “*rough volatility*” models. In [21], the asymptotic expansions of the rough Bergomi [3] model to jointly describe the VIX and the S&P 500 were presented. It was further extended in [18] with the application of modulated Volterra processes. Recently, in [11], Gatheral et al. have introduced and successfully calibrated the so-called *quadratic rough Heston*. Another angle to address this intriguing pricing problem was proposed in [16], where a discrete-time dispersion-constrained martingale transport problem was applied.

The references enumerated above unquestionably address the problem of consistent options pricing on an index and its corresponding volatility. However, in our view, this is insufficient to claim victory- many of the proposed solutions only work for short maturities or lack effective numerical methods to be perceived as practical. We will show that by randomizing AD models, one can benefit from well-established numerical techniques and the ability to price long-term options while consistently pricing volatility products.

The article is organized as follows: in Section 2, we start with the definition of the class of AD and give an integral form for the randomized ChF. In Section 2.1, the RAnD method is introduced, and extensions for bivariate and multivariate parameters cases are presented. Section 2.3 presents an illustrative example of the randomization of the Black-Scholes model. We analyze here the computational aspects of the method and the evolution of implied volatilities. Section 3 focuses on the randomized Bates model; here, we start with the analysis of randomization and its impact on the implied volatilities; also, in this section, we address what was mentioned earlier in the introduction- the randomization of jump sizes. Further, in Section 3.1, we provide the pricing details for VIX options, and in Section 3.2, we use derived formulae to perform simultaneous model calibration to S&P and VIX options. Discussion about hedging under the RAnD framework is covered in Section 3.3 and in Sections 3.4 and 3.5, the details regarding Monte Carlo simulation with possible improvements are given. Section 4 concludes.

2. Affine (Jump) Diffusion processes and randomization

The class of stochastic AD processes for the asset dynamics refers to a fixed probability space $(\Omega, \mathcal{F}(t), \mathbb{Q})$ and a Markovian n -dimensional affine process $\mathbf{X}(t) = [X_1(t), \dots, X_n(t)]^T$ in some space $\mathbb{R} \subset \mathbb{R}^n$.

The stochastic model of interest can be expressed by the following stochastic differential form:

$$d\mathbf{X}(t) = \bar{\boldsymbol{\mu}}(t, \mathbf{X}(t))dt + \bar{\boldsymbol{\sigma}}(t, \mathbf{X}(t))d\widetilde{\mathbf{W}}(t) + \mathbf{J}(t)^T d\mathbf{X}_{\mathcal{P}}(t), \quad (2.1)$$

where $\widetilde{\mathbf{W}}(t)$ is an $\mathcal{F}(t)$ -standard column vector of *independent* Brownian motions in \mathbb{R}^n , $\bar{\boldsymbol{\mu}}(t, \mathbf{X}(t)) : \mathbb{R} \rightarrow \mathbb{R}^n$, $\bar{\boldsymbol{\sigma}}(t, \mathbf{X}(t)) : \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$, and $\mathbf{X}_{\mathcal{P}}(t) \in \mathbb{R}^n$ is a vector of orthogonal Poisson processes, characterized by an intensity vector $\bar{\boldsymbol{\xi}}(t, \mathbf{X}(t)) \in \mathbb{R}^n$.

$\mathbf{J}(t) \in \mathbb{R}^n$ is a vector governing the amplitudes of the jumps and is assumed to be a matrix of correlated random variables, that are independent of the state vector $\mathbf{X}(t)$ and of $\mathbf{X}_{\mathcal{P}}(t)$.

Additionally, we consider an orthogonal vector $\Theta = [\vartheta_1, \dots, \vartheta_n]^T$, $n \in \mathbb{N}$, where each ϑ_i is an independent, time-invariant, random variable¹. A realization of ϑ_i we indicate by θ_i , $\vartheta(\omega) = \theta$, *per se* the realization for Θ is indicated by $\theta = [\theta_1, \dots, \theta_n]^T$.

For processes in the AD class, it is required that for a particular parameter realization, θ , drift $\bar{\boldsymbol{\mu}}(t, \mathbf{X}(t))$, interest rate component $\bar{r}(t, \mathbf{X}(t))$, and the covariance matrix $\bar{\boldsymbol{\sigma}}(t, \mathbf{X}(t))\bar{\boldsymbol{\sigma}}(t, \mathbf{X}(t))^T$ are of the *affine form*, i.e.

$$\bar{\boldsymbol{\mu}}(t, \mathbf{X}(t)) = a_0(\theta) + a_1(\theta)\mathbf{X}(t), \text{ for any } (a_0(\theta), a_1(\theta)) \in \mathbb{R}^n \times \mathbb{R}^{n \times n} \quad (2.2)$$

$$\bar{r}(t, \mathbf{X}(t)) = r_0(\theta) + r_1(\theta)^T \mathbf{X}(t), \text{ for } (r_0(\theta), r_1(\theta)) \in \mathbb{R} \times \mathbb{R}^n, \quad (2.3)$$

$$(\bar{\boldsymbol{\sigma}}(t, \mathbf{X}(t))\bar{\boldsymbol{\sigma}}(t, \mathbf{X}(t))^T)_{i,j} = (c_0(\theta))_{i,j} + (c_1(\theta))_{i,j}^T \mathbf{X}_j(t), \quad (2.4)$$

$$\bar{\boldsymbol{\xi}}(t, \mathbf{X}(t)) = l_0(\theta) + l_1(\theta)\mathbf{X}(t), \text{ with } (l_0(\theta), l_1(\theta)) \in \mathbb{R}^n \times \mathbb{R}^n, \quad (2.5)$$

¹We consider here $n \in \mathbb{N}$ stochastic parameters, this is however not a necessary constraint.

with $(c_0(\theta), c_1(\theta)) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n \times n}$, and where in (2.4) it is meant that each element in the matrix $\bar{\sigma}(t, \mathbf{X}(t))\bar{\sigma}(t, \mathbf{X}(t))^T$ should be affine, as well as each vector element in the drift and interest rate vectors.

2.1. The RAnD method: details on construction of the characteristic function

For a given realization of Θ , θ , we consider $\mathbf{X}_\theta(t) := \mathbf{X}(t)|\Theta = \theta$, $\mathbf{J}_\theta(t) := \mathbf{J}(t)|\Theta = \theta$. It is shown [8] that in this class, for a state vector $\mathbf{X}_\theta(t)$, the discounted characteristic function is also of the following form:

$$\phi_{\mathbf{X}_\theta}(\mathbf{u}; t, T) = \mathbb{E}_t \left[e^{-\int_t^T r(s)ds + i\mathbf{u}^T \mathbf{X}_\theta(T)} \right] = e^{\bar{A}(\mathbf{u}; \tau, \theta) + \bar{\mathbf{B}}^T(\mathbf{u}; \tau, \theta) \mathbf{X}_\theta(t)}, \quad (2.6)$$

with the expectation under risk-neutral measure \mathbb{Q} for $\tau = T - t$.

The coefficients $\bar{A} := \bar{A}(\mathbf{u}; \tau, \theta)$ and $\bar{\mathbf{B}} := \bar{\mathbf{B}}^T(\mathbf{u}; \tau, \theta)$ have to satisfy the following complex-valued *Riccati* ODEs, see the work by Duffie-Pan-Singleton [8]:

$$\begin{aligned} \frac{d\bar{A}}{d\tau} &= -r_0(\theta) + \bar{\mathbf{B}}^T a_0(\theta) + \frac{1}{2} \bar{\mathbf{B}}^T c_0(\theta) \bar{\mathbf{B}} + l_0^T \mathbb{E} \left[e^{\mathbf{J}_\theta(\tau) \bar{\mathbf{B}}} - 1 \right], \\ \frac{d\bar{\mathbf{B}}}{d\tau} &= -r_1(\theta) + a_1(\theta)^T \bar{\mathbf{B}} + \frac{1}{2} \bar{\mathbf{B}}^T c_1(\theta) \bar{\mathbf{B}} + l_1(\theta)^T \mathbb{E} \left[e^{\mathbf{J}_\theta(\tau) \bar{\mathbf{B}}} - 1 \right], \end{aligned} \quad (2.7)$$

where the expectation, $\mathbb{E}[\cdot]$ in (2.7), is taken with respect to the jump amplitude $\mathbf{J}_\theta(t)$.

The dimension of the (complex-valued) ODEs for $\bar{\mathbf{B}}(\mathbf{u}; \tau, \theta)$ corresponds to the dimension of the state vector $\mathbf{X}(t)$. Then, for stochastic parameter ϑ , the ChF is given by:

$$\phi_{\mathbf{X}}(\mathbf{u}; t, T) := \mathbb{E}_t \left[e^{-\int_t^T r(s)ds + i\mathbf{u}^T \mathbf{X}(T)} \right] = \mathbb{E}_t \left[\mathbb{E}_t \left[e^{-\int_t^T r(s)ds + i\mathbf{u}^T \mathbf{X}_\theta(T)} \mid \Theta = \theta \right] \right].$$

The inner expectation can be recognized as the conditional ChF in (2.6); thus, by definition of the ChF and integration over all the parameter space, we find,

$$\phi_{\mathbf{X}}(\mathbf{u}; t, T) = \mathbb{E}_t \left[\phi_{\mathbf{X}|\Theta}(\mathbf{u}; t, T) \right] = \int_{\mathbb{R}^n} \phi_{\mathbf{X}|\Theta=\theta}(\mathbf{u}; t, T) f_\Theta(\theta) d\theta. \quad (2.8)$$

We aim to provide numerically efficient methods for computation of ChF in (2.8).

Remark 1 (Simplified notation). *The representation above is rather generic in terms of multidimensional ChF and possibly a multidimensional set of stochastic parameters. However, such a generic case is rather exceptional, as, in a typical pricing application, a payoff function will only depend on a single, marginal distribution (even so, possibly, driven by the multidimensional system of the SDEs); therefore, for the sake of simplicity, from now on, we focus on the derivations of the Chf for $\mathbf{u} = [u, 0, \dots]$, and we denote it by u . Moreover, although the presented framework allows for simultaneous randomization of multiple parameters, we will consider a single stochastic parameter ϑ with θ indicating its realization. In the following sections, we consider a finite number of realizations of ϑ ; therefore, we will denote them as $\theta_1, \dots, \theta_N$, for some $N \in \mathbb{N}$. These “specific” realizations we will interchangeably call either “collocation” [15] or “quadrature” points. Finally, by $\vartheta(\hat{a}, \hat{b}, \cdot)$ we denote that θ is a random variable driven by parameters, \hat{a}, \hat{b} .*

To determine the ChF of an affine model with a *randomized parameter*, one needs to integrate the parameter’s probability density function over the whole domain (2.8). This can be avoided, i.e., the complicated integrand can be factored into a set of pairs $\{\{\omega_1, \theta_1\}, \dots, \{\omega_N, \theta_N\}\}$, $N \in \mathbb{N}$, with a nonnegative “weights” function, $\omega_i \geq 0$, such that $\sum_{n=1}^N \omega_n = 1$ and specific, collocation, points θ_n . Once the number of evaluations, N , is low, we can significantly reduce the computational cost. The key element here, however, is that the pairs, $\{\omega_n, \theta_n\}_{n=1}^N$, cannot be chosen arbitrarily but need to be computed based on the parameter’s distribution, ϑ . We consider a random parameter ϑ with its PDF, $f_\vartheta(\cdot)$, such that for a fixed number $N \in \mathbb{N}$, moments are finite, i.e., $\mathbb{E}[\vartheta^{2N}] < \infty$. Then we are able to determine a mapping, often called *quadrature rule*, function $\zeta(\vartheta) : \mathbb{R} \rightarrow \{\omega_n, \theta_n\}_{n=1}^N$. We follow the approach presented by G.H. Golub and J.H. Welsch in [12] where ω_n are the quadrature weights determined based on the

moments of the random parameter, ϑ . As explained in [12], one can calculate pairs $\{\omega_n, \theta_n\}_{n=1}^N$ based on a three-term recurrence relation which is well-known for orthogonal polynomials generated by $f_\vartheta(x)$, or equivalently, by the moments of ϑ . Although the technique to compute these pairs of points is well-founded, for the readers' convenience, the theoretical aspects and the algorithmic details ² are included in [Appendix A.1](#). Throughout the article, the method to compute these points will be denoted as $\zeta(\vartheta) : \mathbb{R} \rightarrow \{\omega_n, \theta_n\}_{n=1}^N$.

The RAnD method for determining the ChF with a randomized model parameter is presented in the theorem below.

Theorem 2.1 (Characteristic Function for Randomized Affine Jump Diffusion Processes). *Consider a random variable ϑ defined on some finite domain $D_\vartheta := [a, b]$, with its PDF, $f_\vartheta(x)$, CDF, $F_\vartheta(x)$ and a realization θ , $\vartheta(\omega) = \theta$ such that for some $N \in \mathbb{N}$ the moments are finite, $\mathbb{E}[\vartheta^{2N}] < \infty$. Let $\mathbf{X}(t)$ represent an affine state vector with some constant parameter θ . Assuming that the corresponding ChF, $\phi_{\mathbf{X}|\vartheta=\theta}(\cdot)$, is well defined and $2N$ times differentiable w.r.t θ , the unconditional ChF for the randomized \mathbf{X} , exists and is given by:*

$$\phi_{\mathbf{X}}(\mathbf{u}; t, T) = \sum_{n=1}^N \omega_n \phi_{\mathbf{X}|\vartheta=\theta_n}(\mathbf{u}; t, T) + \epsilon_N = \sum_{n=1}^N \omega_n e^{\bar{A}(\mathbf{u}; \tau, \theta_n) + \bar{\mathbf{B}}^T(\mathbf{u}; \tau, \theta_n) \mathbf{X}(t)} + \epsilon_N, \quad (2.9)$$

with $\bar{A}(\mathbf{u}; \tau, \theta_n)$, $\bar{\mathbf{B}}^T(\mathbf{u}; \tau, \theta_n)$ are defined in (2.7) and where

$$\epsilon_N = \frac{1}{(2N)!} \frac{\partial^{2N}}{\partial \xi^{2N}} \phi_{\mathbf{X}|\vartheta=\xi}(\mathbf{u}; t, T), \quad (2.10)$$

for $a < \xi < b$ and where the pairs $\{\omega_n, \theta_n\}_{n=1}^N$ are the Gauss-quadrature weights and the nodes based on the parameter distribution, $f_\vartheta(\cdot)$, determined by $\zeta(\vartheta) : \mathbb{R} \rightarrow \{\omega_n, \theta_n\}_{n=1}^N$ defined in [Appendix A.1](#).

Proof. Starting with the definition of the ChF and conditioning on a parameter ϑ , we find:

$$\phi_{\mathbf{X}}(\mathbf{u}; t, T) := \mathbb{E}_t \left[e^{-\int_t^T r(s) ds + i\mathbf{u}^T \mathbf{X}(T)} \right] = \mathbb{E}_t \left[\mathbb{E}_t \left[e^{-\int_t^T r(s) ds + i\mathbf{u}^T \mathbf{X}_\theta(T)} \mid \vartheta = \theta \right] \right].$$

Then, by definition of the ChF,

$$\begin{aligned} \phi_{\mathbf{X}}(\mathbf{u}; t, T) &= \mathbb{E}_t \left[\phi_{\mathbf{X}|\vartheta=\theta}(\mathbf{u}; t, T) \right] = \int_{\mathbb{R}} \phi_{\mathbf{X}|\vartheta=\theta}(\mathbf{u}; t, T) f_\vartheta(x) dx \\ &= \sum_{n=1}^N \omega_n \phi_{\mathbf{X}|\vartheta=\theta_n}(\mathbf{u}; t, T) + \epsilon_N. \end{aligned}$$

First, we prove that a convex linear combination of characteristic functions, $\sum_{n=1}^N \omega_n \phi_{\mathbf{X}|\vartheta=\theta_n}(\mathbf{u}; t, T)$, and $\sum_{n=1}^N \omega_n = 1$ with $\omega_n > 0$ of a finite or a countable number of characteristic function is the characteristic function.

Since ω_n can be associated with probabilities, we can define a discrete random variable $\mathbb{P}[\vartheta = \theta_n] = \omega_n$, so we have:

$$\sum_{n=1}^N \omega_n \phi_{\mathbf{X}|\vartheta=\theta_n}(\mathbf{u}; t, T) = \sum_{n=1}^N \mathbb{E}[1_{\vartheta=\theta_n}] \mathbb{E}_t \left[e^{i\mathbf{u}^T \mathbf{X}_{\theta_n}(T)} \right] = \sum_{n=1}^N \mathbb{E}[1_{\vartheta=\theta_n} e^{i\mathbf{u}^T \mathbf{X}_{\theta_n}(T)}],$$

since ϑ is independent of $\mathbf{X}(t)$. For Fubini's theorem to hold, the boundness condition needs to be satisfied,

$$\sum_{n=1}^N \mathbb{E} \left[\left| 1_{\vartheta=\theta_n} e^{i\mathbf{u}^T \mathbf{X}_{\theta_n}(T)} \right| \right] = \sum_{n=1}^N \mathbb{E} \left[\left| 1_{\vartheta=\theta_n} \right| e^{i\mathbf{u}^T \mathbf{X}_{\theta_n}(T)} \right] \leq \sum_{n=1}^N \mathbb{P}[\vartheta = \theta_n] = 1 < \infty,$$

²The accompanying Python and MATLAB codes are available at <https://github.com/LechGrzelak/Randomization>

therefore:

$$\sum_{n=1}^N \omega_n \phi_{\mathbf{X}|\vartheta=\theta_n}(\mathbf{u}; t, T) = \mathbb{E}\left[\sum_{n=1}^N 1_{\vartheta=\theta_n} e^{i\mathbf{u}^T \mathbf{X}_{\theta_n}(T)}\right] = \mathbb{E}\left[\exp\left(i\mathbf{u}^T \sum_{n=1}^N 1_{\vartheta=\theta_n} \mathbf{X}_{\theta_n}(T)\right)\right] =: \phi_Y(\mathbf{u}),$$

with $\phi_Y(\cdot)$ being a ChF of $\sum_{n=1}^N 1_{\vartheta=\theta_n} \mathbf{X}_{\theta_n}(T)$.

Using the fact that the collocation points θ_n for $n = 1, \dots, N$ are optimal and correspond to the zeros of orthogonal polynomials, we relate the collocation method to Gaussian quadrature where the integral of a complex-valued function, $\phi_{\mathbf{X}|\vartheta=\theta}(\mathbf{u}; t, T)$, can be approximated by a polynomial:

$$\int_{\mathbb{R}} \phi_{\mathbf{X}|\vartheta=x}(\mathbf{u}; t, T) f_{\vartheta}(x) dx - \sum_{n=1}^N \phi_{\mathbf{X}|\vartheta=\theta_n}(\mathbf{u}; t, T) \omega_n = \frac{1}{(2N)!} \frac{\partial^{2N}}{\partial \xi^{2N}} \phi_{\mathbf{X}|\vartheta=\xi}(\mathbf{u}; t, T) =: \epsilon_N,$$

with $f_{\vartheta}(x)$ the weight function, and ω_n the quadrature weights. The error is given by ϵ_N , for $a < \xi < b$, and the pair $\{\omega_n, \theta_n\}_{n=1}^N$ are the Gauss-quadrature weights and the nodes based on the parameter distribution as proven in [5] (p.180, Theorem 3.6.24). ■

Theorem 2.1 illustrates that the ChF of the randomized AD model is a weighted sum of a set of conditional ChFs evaluated at certain realizations, θ_n , of the underlying stochastic parameter ϑ . The theorem also shows, in Equation (2.10), the exponential decay of the error in terms of N associated with the quadrature approximation- suggesting that only a few terms will be needed to reach high-precision.

The key feature of the RAnD method is its ability to facilitate fast evaluation of the randomized ChF (2.9). The construction of the ChF is based on the randomized variable, ϑ , via its $2N$ moments, i.e., the availability of the moments allows us to determine $\zeta(\vartheta)$ which provides the quadrature points $\{\omega_n, \theta_n\}_{n=1}^N$. It is, therefore, essential that the randomizing variables enable the closed-form moment calculation. Moreover, when specifying a randomizing variable, its support must correspond to the physical range of the randomized parameter; for example, some of the model parameters may be only valid for specific ranges, such as correlation or volatility coefficients. In Table 1, a few possible distributions with their corresponding moments are provided. When analytical moments are available, the computation of the corresponding points only requires the computation of a Cholesky decomposition and certain eigenvalues (see Appendix A.1); it is therefore computationally cheap.

Table 1: Selected distributions for the stochastic parameters. For the normal random variable with some mean, μ , and variance, σ^2 , it is sufficient to consider standard normal distribution, $\mathcal{N}(0, 1)$, and properly scale the θ_n points, obtained from Algorithm 1.

name	raw moment	domain
$\vartheta \sim \mathcal{U}([\hat{a}, \hat{b}])$	$\mathbb{E}[\vartheta^n] = \frac{\hat{b}^{n+1} - \hat{a}^{n+1}}{(n+1)(\hat{b} - \hat{a})}$	$[\hat{a}, \hat{b}]$
$\vartheta \sim \exp(\hat{a})$	$\mathbb{E}[\vartheta^n] = \frac{n!}{\hat{a}^n}$	\mathbb{R}^+
$\vartheta \sim \mathcal{N}(0, 1)$	$\mathbb{E}[\vartheta^n] = (n-1)!!$ if n even; 0 otherwise	\mathbb{R}
$\vartheta \sim \Gamma(\hat{a}, \hat{b})$	$\mathbb{E}[\vartheta^n] = \hat{b}^n \Gamma(n + \hat{a}) / \Gamma(\hat{a})$	\mathbb{R}^+
$\vartheta \sim \chi^2(\hat{a}, \hat{b})$	$\mathbb{E}[\vartheta^n] = 2^{n-1} (n-1)! (\hat{a} + n\hat{b}) + \sum_{j=1}^{n-1} \frac{(n-1)! 2^{j-1}}{(n-j)!} (\hat{a} + j\hat{b}) \mathbb{E}[\vartheta^{n-j}]$	$\mathbb{R}^+ \cup \{0\}$

2.2. Pricing with the RAnD method

We will focus on pricing two different plain vanilla options, namely European options on an asset, $S(t)$, and options on the associated volatility, VIX. The pricing will rely on a Fourier inversion method, namely the COS method [26], which, by a change of the pricing coefficients, H_k and the corresponding ChF allows for pricing of both contracts. The generic pricing equation is given by:

$$V(t_0) = e^{-r(T-t_0)} \sum_{k=0}^{N_c-1} \Re \left[\phi_{\mathbf{X}} \left(\frac{k\pi}{b-a}; t_0, T \right) \exp \left(-ik\pi \frac{a}{b-a} \right) \right] \cdot H_k + \epsilon_{c_1}, \quad (2.11)$$

where N_c represents the number of expansion terms, $\Re(\cdot)$ is the real part and where H_k for $k \geq 0$ are known in closed-form coefficients corresponding to the payoff function. We will use the H_k coefficients derived for European-style call/put options (see [26] for details) and for options on VIX (presented later, in Section 3, Lemma 3.2). Parameters, a and b are the *tuning* parameters used to determine the integration range; while the error ϵ_{c_1} , is exponentially decaying in N_c .

Remark 2 (Inversion of the randomized ChF). *The application of Fourier inversion to the randomized ChF in (2.9) yields,*

$$f_{\mathbf{X}}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \sum_{n=1}^N \omega_n \phi_{\mathbf{X}|\vartheta=\theta}(u; t, T) du = \sum_{n=1}^N \omega_n f_{\mathbf{X}|\vartheta=\theta}(x). \quad (2.12)$$

Since $\omega_1 + \dots + \omega_N = 1$, $\omega_n \geq 0$, for $n = 1, \dots, N$, which implies that the density of the affine, randomized, system of SDEs, $\mathbf{X}(t)$ can be expressed as a, possibly multi-modal, mixture distribution. Although the exact representation for the density $f_{\mathbf{X}}(\cdot)$ may be challenging to derive analytically, one can use Fourier inversion to recover the corresponding density and utilize it for swift option pricing.

A natural extension of the RAnD method is to consider two parameters that are stochastic and follow a bivariate distribution. Such an extension would require that the conditional moments are known explicitly- which is somewhat troublesome. To this day, only a few distributions allow for exact moments. However, if we stay, for example, within the Gaussian world, such an extension to a 2D case is trivial. Corollary 2.1 gives us the general details.

Corollary 2.1 (Random parameters with bivariate distribution). *Under a bivariate distribution $\Theta = [\vartheta_1, \vartheta_2]$ with $\zeta(\vartheta_1) = \{\omega_{1,n}, \theta_{1,n}\}_{n=1}^N$ and conditioned on $\zeta(\vartheta_2|\vartheta_1) = \{\omega_{2,n}, \theta_{2,n}\}_{n=1}^N$, the randomized ChF is given by:*

$$\phi_{\mathbf{X}}(u; t, T) = \sum_{n_1=1}^N \omega_{n_1} \sum_{n_2=1}^N \omega_{n_2} \phi_{\mathbf{X}|\vartheta_1=\theta_{n_1}, \vartheta_2=\theta_{n_2}}(u; t, T) + \tilde{\epsilon}_N, \quad (2.13)$$

where N indicates the number of expansion terms, $\vartheta_2|\vartheta_1$ indicates a conditional random variable, and the remaining specification follows Theorem 2.1.

2.3. Illustrative case: the RAnD Black-Scholes model

As the first example for the RAnD method, let us consider the randomized Black-Scholes (RAnD BS) model. Given the probability space $(\Omega, \mathcal{F}(t), \mathbb{Q})$, we define a stock $S(t) := \{S(t, \omega) : t \in T\}$ and the randomizing volatility $\sigma := \{\sigma(\omega) : \Omega \rightarrow \mathbb{R}^+\}$, with some constant interest rate, r . We consider three different, continuous, randomizers for σ , i.e., we assume the volatility parameter, σ , to either follow a uniform distribution, $\sigma \sim \mathcal{U}([\cdot, \cdot])$, gamma distribution, $\sigma \sim \Gamma(\cdot, \cdot)$, or the non-central chi-squared distribution, $\sigma \sim \chi^2(\cdot, \cdot)$; thus, also mimicking a similar structure as in the Heston model where the CIR dynamics drive the variance process (see Equations (3.1) and (3.2) for the exact relation).

Formally, the RAnD BS model is given by the following dynamics:

$$\begin{aligned} dS(t) &= rS(t)dt + \sigma S(t)dW(t), \quad S(t_0) > 0, \\ \sigma &\sim \mathcal{U}([\hat{a}, \hat{b}]), \quad \text{or } \sigma \sim \Gamma(\hat{a}, \hat{b}), \quad \text{or } \sigma \sim \hat{a}\chi^2(\hat{b}, \hat{c}), \end{aligned}$$

for $\hat{a}, \hat{b}, \hat{c} \in \mathbb{R}^+$. As it will be presented later, the randomization of the BS model is elegant as it allows for a generation of the European option prices in closed form (see [4] where discrete randomizers for the BS model were covered). However, since the randomized volatility is time-invariant, or *stationary*, such a model cannot compete with, e.g. the model of Heston, where the volatility has a term structure, and it is driven by a correlated stochastic process. Contrary to the Heston model, however, the RAnD BS model does not require Fourier transformation to price options.

We have two techniques to compute option prices of the randomized Black-Scholes model. First, a straightforward approach is based on the closed-form solution of the “unrandomized”

model. The second method is based on randomization at the ChF level. By application of the RAnD method and conditioning on the volatility realizations, the price of a European call option, for $t_0 = 0$, reads:

$$\begin{aligned} V(t_0) &= e^{-rT} \mathbb{E}_{t_0} [\max(S(T) - K, 0)] \\ &= e^{-rT} \mathbb{E}_{t_0} [\mathbb{E}_{t_0} [\max(S(T) - K, 0) | \sigma = \sigma_n]] = e^{-rT} \mathbb{E}_{t_0} [V_c(t_0, T, K, S(t_0), r, \sigma_n)], \end{aligned}$$

with $\mathbb{E}_{t_0}[\cdot]$ being the expectation operator with filtration $\mathcal{F}(t_0)$, where $V_c(t_0, T, K, S(t_0), r, \sigma_n)$ represents a European call option price with strike K , expiry T and the corresponding interest rate r , and it is defined in (2.15) below. The price $V(t_0)$ is now given by:

$$\begin{aligned} V(t_0) &= e^{-rT} \int_{\mathbb{R}^+} V_c(t_0, T, K, S(t_0), r, x) f_\sigma(x) dx \\ &= e^{-rT} \sum_{n=1}^N \omega_n V_c(t_0, T, K, S(t_0), r, \sigma_n) + \hat{\epsilon}_N, \end{aligned} \quad (2.14)$$

where pairs $\{\omega_n, \sigma_n\}$ are the weights and the quadrature points, determined based on the distribution of $\sigma(\omega)$ and are computed using Algorithm 1 in Appendix A.1, $\hat{\epsilon}_N$ is the quadrature error defined in (2.9).

At each realization σ_n , the pricing is known in closed form and is given the Black-Scholes formula:

$$V_c(t_0, T, K, S(t_0), r, \sigma_n) = S(t) F_{\mathcal{N}(0,1)}(d_1) - K F_{\mathcal{N}(0,1)}(d_2) e^{-r(T-t_0)}, \quad (2.15)$$

with $d_1 = \frac{1}{\sqrt{\sigma_n^2(T-t_0)}} [\log S(t_0)/K + (r + 1/2\sigma_n^2)(T - t_0)]$, $d_2 = d_1 - \sqrt{\sigma_n^2(T - t_0)}$.

Remark 3 (Continuous randomizers for the BS model- previous attempts). *Early attempts to randomize the Black-Scholes model using continuous random variables are already known in the literature. In [19], the authors have randomized the variance of the Black-Scholes model by a CEV-generated distribution. The primary aim of such randomization was to achieve the so-called moderately explosive distribution allowing for enhanced skew. The presented framework, however, although via the asymptotic expansion allowed for short-maturity proxies for the implied volatilities, is very much problem and randomizer specific. Moreover, due to the nature of the expansion, it performs well only for short-expiry derivatives, contrary to the RAnD method discussed in this article.*

2.3.1. Randomization of the ChF under the Black-Scholes model

The presented randomization technique is not limited to the Black-Scholes model, where a closed-form solution exists for each volatility realization, σ_n . In order to achieve more flexibility and be able to price challenging derivatives, the stochastic volatility models, like the Heston and Bates models, will be considered. A numerical routine, like Fourier inversion, needs to be used. Although such inversions are swift due to exponentially convergent methods, like the COS method [26], it is important to keep the number of inversions as low as possible. This requirement suggests that a preferred way to compute a randomized option price is via randomization at the ChF level, introduced in (2.9), which for the RAnD BS model is given by the following expression,

$$\phi_X(u; t_0, T) = \sum_{n=1}^N \omega_n \exp\left(iuX(t_0) + \left(r - \frac{1}{2}\sigma_n^2\right)iu(T - t_0) - \frac{1}{2}\sigma_n^2u^2(T - t_0)\right) + \epsilon_N, \quad (2.16)$$

with $X(t_0) = \log S(t_0)$, the error ϵ_N defined in (2.9).

The pricing of European-style options can be performed using the expansion formula in (2.11).

At this point, one can still employ the pricing equation (2.11) to (2.14); however, it would require N evaluations of the pricing equation in (2.11). Since the strength of the COS method is its ability to compute option prices for a range of strikes K , it is important to keep the number of vector-matrix multiplications in (2.11) to a minimum; it is, therefore, beneficial

to evaluate the pricing only once- thus the randomization should take place at the ChF level (as a ChF does not depend on strikes). An additional element that needs to be considered is the generated error, i.e., depending on the smoothness of the option value (2.14) or the corresponding ChF (2.16), error $\hat{\epsilon}_N$ may differ from the ϵ_N in (2.14).

2.3.2. RAnD BS model: convergence and implied volatilities

In order to assess the convergence of the RAnD method and the impact of the randomization on implied volatilities, we consider a numerical experiment, where two pricing approaches, presented in the previous section, will be compared. In particular, we are interested in the implied volatility behaviour for short maturity options ($T = 0.1$), as this is often considered a challenging case. We start by specifying the parameters for the randomizing random variables (see Table 2). With the distributions given, we are able to determine the corresponding pairs

Table 2: Parameters used in the RAnD BS model simulation. For the non-central chi-square distributions, the parameters were chosen based on the Heston model (see Dynamics (3.1)), with: $\gamma = 1.9$, $\kappa = 0.5$, $\bar{v} = 0.3$, $v_0 = 0.3$, $T = 0.1$.

	\hat{a}	\hat{b}	\hat{c}
$\sigma \sim \mathcal{U}([\hat{a}, \hat{b}])$	0.1	0.45	-
$\sigma \sim \Gamma(\hat{a}, \hat{b})$	2.55	0.1	-
$\sigma \sim \hat{a}\chi^2(\hat{b}, \hat{c})$	0.088	0.1662	3.2417

$\{\omega_n, \sigma_n\}$, for $n = 1, \dots, N$. These points are based on the moments of the randomizing random variable (see Table 1) and utilization of Algorithm 1. Figure 1 (left) illustrates the method's convergence regarding the number of collocation points N . Surprisingly, for the randomized Black-Scholes model, it does not matter whether the randomization takes place at the price level or the corresponding ChF. However, the convergence rate is affected by the associated randomizer. We report the highest convergence speed for the uniform distribution, lower for gamma, and the slowest convergence rate for the non-central chi-square randomizer. The phenomenon can be explained by the *amount* of randomness introduced to the pricing equation.

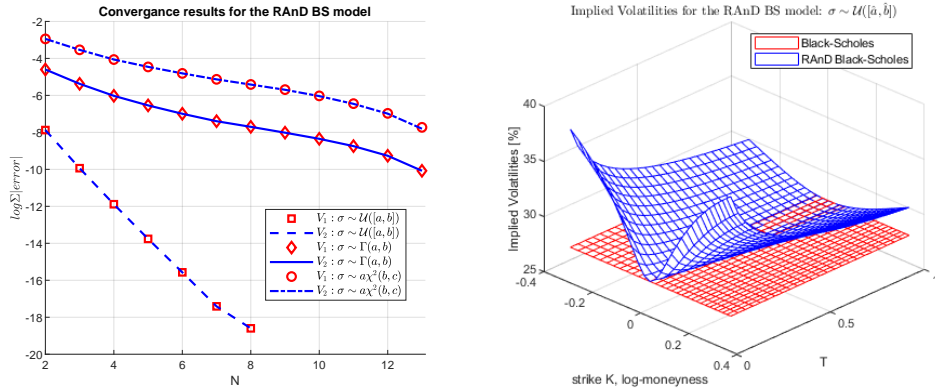


Figure 1: Left: The RAnD method convergence for the number of collocation points N . Red markers correspond to the results obtained based on analytical expression (2.14), and blue lines correspond to Fourier expansion (2.16). The “error” is the difference between option prices against the reference, fetched with 5 million paths. The summation is over the entire range of option strikes. Right: Implied volatility surface for the RAnD BS model for $\sigma \sim \mathcal{U}([\hat{a}, \hat{b}])$.

In Figure 1 (right), the impact of $\sigma \sim \mathcal{U}([\hat{a}, \hat{b}])$, on the implied volatilities is illustrated (the remaining cases are illustrated in Appendix A.2 in Figure A.10). As expected, the RAnD BS model can generate an implied volatility smile. However, because of the independence between the stock's Brownian motion and the randomizer, the skew control is limited, irrespective of the randomizing distribution. It is important to note that by introducing stochastic but time-invariant volatility, we can generate an implied volatility surface that develops in time from

a very peaked, *almost explosive*, to a flattening shape. The implied volatility surface pattern is very similar in all three cases. The RAnD BS model could be further extended, with, for example, a displacement parameter that would introduce skewness in the model, thus further extending the applicability of the Black-Scholes model's standard.

In the final experiment for the RAnD BS model, we report the convergence rate in terms of the implied volatilities in Table 3. We see that the convergence is not affected by the maturity time T . We expect that, with stochastic volatility models, the convergence will depend on the time-to-expiry. The results are excellent; for $N = 6$ points, the maximum absolute error is about 0.05 – 0.07%, which is a convincing result and implies that the method can be used for accurate model calibration and pricing. As expected, the reported errors depend on the randomizing variable (compare Table 3 against Table A.5 in Appendix A.5).

Table 3: Maximum implied volatility, $IV(\cdot)$, error computed against Monte Carlo results with 10 million samples, with $K = S_0 e^{0.1\sqrt{T}\delta}$, $\delta = [-3, -2, -1, -0.5, 0, 0.5, 1, 2, 3]$.

$\sigma \sim \Gamma(a, b)$	error = $\max_i IV(K_i) - IV_{\text{ref}}(K_i) $							
	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$
$T = 1d$	0.56 %	0.26 %	0.15 %	0.10 %	0.07 %	0.06 %	0.05 %	0.04 %
$T = 1w$	0.56 %	0.26 %	0.15 %	0.10 %	0.07 %	0.06 %	0.04 %	0.04 %
$T = 2w$	0.55 %	0.25 %	0.14 %	0.09 %	0.06 %	0.04 %	0.03 %	0.03 %
$T = 1m$	0.54 %	0.24 %	0.13 %	0.08 %	0.06 %	0.04 %	0.03 %	0.03 %
$T = 3m$	0.54 %	0.24 %	0.13 %	0.08 %	0.05 %	0.04 %	0.03 %	0.02 %
$T = 6m$	0.55 %	0.25 %	0.14 %	0.09 %	0.07 %	0.05 %	0.04 %	0.03 %
$T = 12m$	0.56 %	0.26 %	0.15 %	0.10 %	0.07 %	0.05 %	0.04 %	0.04 %

Further analysis regarding randomization and implied volatility shapes will be discussed later in the context of the Heston and the Bates models.

3. The RAnD framework and pricing of plain vanilla and VIX options

This section focuses on applying the RAnD method to stochastic volatility and jump models. In particular, we will consider the Bates [2] model, which forms an elegant extension of the model of Heston [17]. From the pricing perspective, the characteristic function of both models is known analytically. Although some may argue that managing jumps in the pricing framework is of limited value, especially for hedging, it is well known that *pure* stochastic volatility, affine, models cannot generate sufficient skew for short maturity options. Adding jumps to calibrate short-maturity options is often a preferred choice. However, it is often insufficient. Our goal is to use the RAnD method and apply it to the challenging task of simultaneous pricing of plain-vanilla and VIX options. In particular, we illustrate the ability to fit both option markets simultaneously by randomizing some Bates model parameters.

Randomization of individual Heston model parameters is known in the literature. Initial random volatility was discussed in [24, 20, 25], and in [10], the perturbation-based approximations for stochastic *vol-of-vol* parameter were derived.

We will focus on the Bates model: the corresponding ChF, pricing of plain-vanilla and VIX options and calibration, under the RAnD method, to market data.

The Bates model, under the \mathbb{Q} measure, is described by the following system of SDEs:

$$\begin{aligned} dS(t)/S(t) &= (r - \lambda\mathbb{E}[e^J - 1]) dt + \sqrt{v(t)}dW_x(t) + (e^J - 1) dX_{\mathcal{P}}(t), \\ dv(t) &= \kappa(\bar{v} - v(t)) dt + \gamma\sqrt{v(t)}dW_v(t), \end{aligned} \tag{3.1}$$

with Poisson process $X_{\mathcal{P}}(t)$, intensity λ , and normally distributed jump sizes, $J \sim \mathcal{N}(\mu_j, \sigma_j^2)$, with $\mathbb{E}[e^J] = e^{\mu_j + \frac{1}{2}\sigma_j^2}$. $X_{\mathcal{P}}(t)$ is assumed to be independent of the Brownian motions and the jump sizes. There is a correlation ρ between the governing Brownian motions, $\rho dt = dW_x(t)dW_v(t)$. Under this model the variance process follows the non-central chi-square distribution, $\chi^2(\delta, \bar{\kappa}(\cdot, \cdot))$ with δ degrees of freedom and non centrality parameter $\bar{\kappa}(t_0, t)$,

$$v(t)|v(t_0) \sim \bar{c}(t_0, t)\chi^2(\delta, \bar{\kappa}(t_0, t)), \tag{3.2}$$

where

$$\bar{c}(t_0, t) = \frac{\gamma^2}{4\kappa}(1 - e^{-\kappa(t-t_0)}), \quad \delta = \frac{4\kappa\bar{v}}{\gamma^2}, \quad \bar{\kappa}(t_0, t) = \frac{4\kappa e^{-\kappa(t-t_0)}v(t_0)}{\gamma^2(1 - e^{-\kappa(t-t_0)})}. \quad (3.3)$$

Under a log transformation, the dynamics of the stock, $X(t) = \log S(t)$, belongs to the class of affine jump-diffusion processes, and it reads:

$$dX(t) = \left(r - \frac{1}{2}v(t) - \lambda\mathbb{E}[e^J - 1]\right) dt + \sqrt{v(t)}dW_x(t) + JdX_{\mathcal{P}}(t),$$

so that, given Theorem 2.1, we can derive the corresponding, randomized, ChF, which is given by the following expression:

$$\phi_X(u; t, T) = \sum_{n=1}^N \omega_n e^{iuX(t) + \bar{C}(u, \tau; \theta_n)v(t) + \bar{A}_B(u, \tau; \theta_n)} + \epsilon_N, \quad (3.4)$$

for $\tau = T - t$, with complex valued functions $\bar{A}_B(u; \tau, \theta_n)$ and $\bar{C}(u; \tau, \theta_m)$, given in Appendix A.3. The third argument θ_n represents a particular realization of model parameter, $\vartheta \in \{v_0, \lambda, \gamma, \rho, \mu_J, \sigma_J, \kappa, \bar{v}\}$, that we wish to randomize.

Once the RAnD ChF is derived, it is worth checking the convergence concerning the number of summation terms, N , in (3.4). As an example, in Figure 2 the convergence for randomization of the vol-vol parameter, γ , with two distributions, uniform, \mathcal{U} , and gamma distribution, Γ , is depicted. The convergence rate depends on the distribution itself but also on its parameters. In both cases, the convergence rate is excellent; however, as expected, it is much faster for a uniform distribution than for the gamma distribution. In both cases, satisfactory results will be accomplished for $N = 5$. Randomization of other model parameters has shown equivalent convergence patterns.

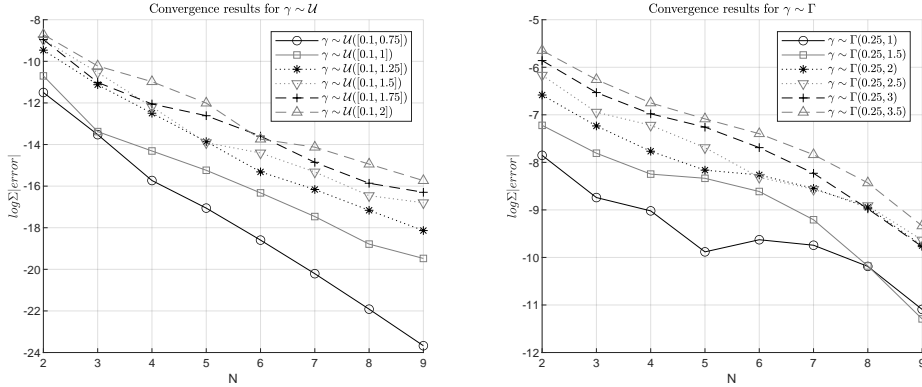


Figure 2: Convergence results for randomization of vol-of-vol parameter, γ under the Bates model with: $T = 31/365$, $r = 0$, $\mu_J = -0.25$, $\sigma_J = 0.05$, $\lambda = 0.1$, $\kappa = 0.5$, $\gamma = 0.72$, $\bar{v} = 0.1$, $\rho = -0.85$ and $v_0 = 0.0625$. Left: $\gamma \sim \mathcal{U}(\hat{a}, \hat{b})$ for varying \hat{a} and \hat{b} ; Right: $\gamma \sim \Gamma(\hat{a}, \hat{b})$ for varying \hat{a} and \hat{b} .

The impact of the randomization for the Bates model can be analyzed with the ChF in (3.4). In particular, the time evolution of the implied volatility surface for randomized vol-vol, γ , and the jump amplitude parameter, μ_J , is presented in Figure 3. The results are insightful, showing that the RAnD method can magnify the skewness of short-maturity options. Intriguingly, the evolution of the skew differs for both parameters, i.e., randomized γ allows for slow skew decay towards the benchmark- the basic Bates model, while the randomization of μ_J generates much more rapid decay. Depending on the market data, both shapes may be desired.

Because of the flexibility of the RAnD method, the randomization of all the Bates model parameters can be analyzed. As discussed earlier, one can specify a variety of distributions (see Table 1) for the model parameters. In Figure 4 and Figure 5, the impact of randomization with different distributions and parameters is compared against the benchmark. Randomization

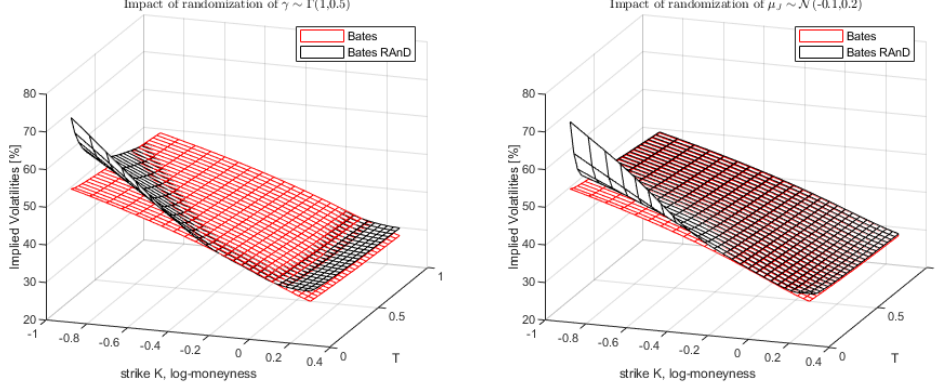


Figure 3: Implied volatility surface for RAnD Bates model. Left panel: randomized *vol-vol*, $\gamma \sim \Gamma(1,0.5)$. Right panel: randomized jump's mean, $\mu_J \sim \mathcal{N}(-0.1,0.2)$. Other model parameters are $r = 0$, $\mu_J = -0.1$, $\sigma_J = 0.06$, $\lambda = 0.08$, $\kappa = 0.5$, $\gamma = 0.5$, $\bar{v} = 0.13$, $\rho = -0.7$, $T = 1/12$, and $v_0 = 0.13$.

enables flexible control of the implied volatility shapes. In this experiment we choose the reference model parameters ($r = 0.0$, $\mu_J = -0.1$, $\sigma_J = 0.06$, $\lambda = 0.08$, $\kappa = 0.5$, $\gamma = 0.5$, $\bar{v} = 0.13$, $\rho = -0.7$, $v_0 = 0.13$ and $T = 1/12$), and perform one-by-one randomization.

Starting with γ , in the left panel of Figure 4, we report a significant impact on the curvature and the skew. Randomization of v_0 (right panel of Figure 4) mainly affects the implied volatility level; however, some curvature effect is also present. The randomization of the correlation (left panel of Figure 5) generates the rotation of the implied volatilities around the ATM level. Another striking effect is reported for μ_J where only the volatilities far from the ATM level are affected, i.e., one can control either the left or the right tail while keeping the ATM level fixed. The remaining parameters are presented in Figure A.11 and Figure A.12 in Appendix A.6.

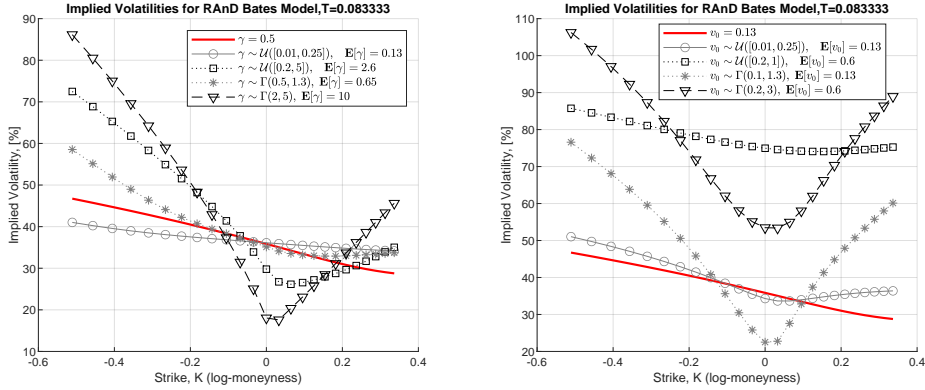


Figure 4: Impact of randomized parameters on implied volatilities. Left: randomized *vol-vol*, γ . Right: randomized *initial vol*, v_0 .

3.1. Vix option pricing under the randomized Heston model with jumps

For a given fixed time-horizon $[t, T]$, the volatility index of an asset $S(t)$, denoted as $\text{vix}(t, T)$ is defined as:

$$\overline{\text{vix}}^2(t, T) = 100^2 \times \frac{-2}{T-t} \mathbb{E}_t \left[\log \frac{S(T)}{S(t)} \right], \quad (3.5)$$

where $\mathbb{E}_t[\cdot]$ indicates the expectation taken under the risk-neutral measure \mathbb{Q} and the natural filtration $\mathcal{F}(t)$. The VIX formulation in (3.5) represents the annualized square root of

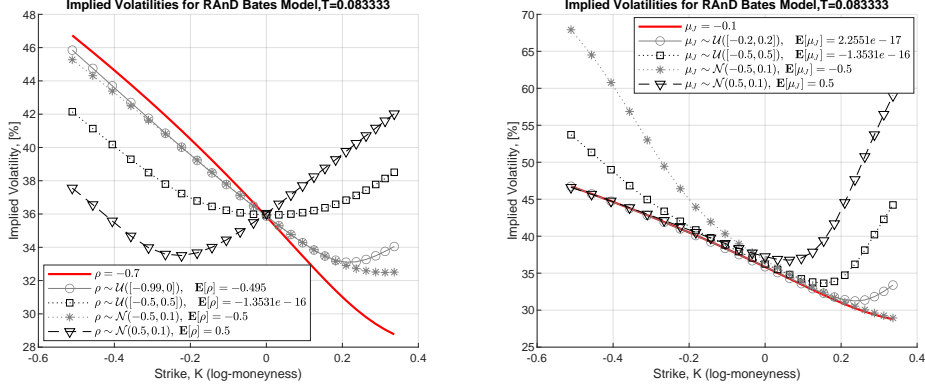


Figure 5: Impact of randomized parameters on implied volatilities. Left: randomized correlation, ρ . Right: randomized jump's mean, μ_J .

the price of a contract, where the most liquid options are for $T - t = 30$ days. The scaling with 100 relates to a percentage representation of the volatility.

Depending on the dynamics chosen for stock $S(t)$, the expression in (3.5) may be derived in closed-form. In particular, for the case of the Bates model: the closed-form expression for the expectation and the probability density function, $f_{\text{vix}}(x; t, T)$, are known explicitly and are presented in Lemma 3.1 below.

Lemma 3.1. *Under the Bates model (3.1), the VIX defined in (3.5) is expressed by:*

$$\overline{\text{vix}}^2(t, T) = 100^2 \times \text{vix}^2(t, T), \quad (3.6)$$

$$\text{vix}^2(t, T) = a(t, T)v(t) + b(t, T) + c, \quad (3.7)$$

where

$$a(t, T) = \frac{1 - e^{-\kappa(T-t)}}{\kappa(T-t)}, \quad b(t, T) = \bar{v}(1 - a(t, T)), \quad c = 2\lambda \left(e^{\mu_J + \frac{1}{2}\sigma_J^2} - \mu_J - 1 \right),$$

and $v(t)$ follows a non-central chi-square distribution, defined by the CIR process in (3.1). λ represents the Poisson intensity, and μ_J and variance σ_J^2 are the parameters for the jumps magnitude, $J \sim \mathcal{N}(\mu_J, \sigma_J^2)$. The PDF of vix in (3.7), is given by:

$$f_{\text{vix}}(x; t, T) = 2\alpha_1 x f_{\chi^2}(\delta, \bar{\kappa}(t_0, t))(\alpha_1(x^2 - \alpha_2)), \quad (3.8)$$

where $\alpha_1 = \frac{1}{a(t, T)\bar{c}(t_0, t)}$ and $\alpha_2 = b(t, T) + c$ with δ , $\bar{\kappa}(\cdot, \cdot)$ and $\bar{c}(\cdot, \cdot)$ defined in (3.3). Finally, the ChF for $\text{vix}^2(t, T)$ is given by:

$$\phi_{\text{vix}^2(t, T)}(u) = e^{iu(b(t, T)+c)} \left(\frac{\alpha_1}{\alpha_1 - 2iu} \right)^{\frac{1}{2}\delta} \exp \left(\frac{iu a(t, T)\bar{c}(t_0, t)\bar{\kappa}(t_0, t)}{1 - 2iu a(t, T)\bar{c}(t_0, t)} \right). \quad (3.9)$$

Proof. The proof can be found in Appendix A.4. ■

Given the results from Lemma 3.1 and the availability of the ChF (3.9) for VIX defined in (3.5), we consider a European option with strike price K and expiry T , whose terminal payoff is described as $\max(\text{vix}(T, T + \delta T) - K, 0)$, with δT is equal to 30 days. According to the risk-neutral valuation theory, the option price at time t , denoted by $V_{\text{vix}}(t)$, can be expressed as the discounted conditional expectation of the terminal payoff under the risk-neutral measure, \mathbb{Q} :

$$\begin{aligned} V_{\text{vix}}(t) &= e^{-r(T-t)} \mathbb{E}_t \left[\max(\overline{\text{vix}}(T, T + \delta T) - \bar{K}, 0) \right] \\ &= 100 \times e^{-r(T-t)} \int_{\mathbb{R}^+} \max(\sqrt{v} - K, 0) f_{\text{vix}^2}(v; T, T + \delta T) dv, \end{aligned} \quad (3.10)$$

for $\bar{K} = 100 \times K$ and $\overline{\text{vix}}(\cdot, \cdot)$ defined in (3.6). Since the ChF for $\text{vix}^2(\cdot, \cdot)$ is explicitly known (see Equation (3.9)), we utilize the Fourier expansion from (2.11). The missing ingredients are the payoff-dependent coefficients H_k . These coefficients are provided by Lemma 3.2.

Lemma 3.2. *Assuming, $a < K^2$, the H_k coefficients, defined in (2.11) are given by:*

$$\begin{aligned} H_k &= \frac{2}{b-a} \int_a^b \max(\sqrt{y} - K, 0) \cos\left(k\pi \frac{y-a}{b-a}\right) dy \\ &=: \frac{2}{b-a} I_1(k) - K \frac{2}{b-a} I_2(k), \end{aligned}$$

where two integral terms $I_1(k)$ and $I_2(k)$ are given by:

$$I_1(k) := \begin{cases} a_0 \left[\sin(k_2) (C_F(a_1) - C_F(a_2)) + \cos(k_2) (S_F(a_2) - S_F(a_1)) \right] \\ \quad + a_5 \left[\sqrt{b} \sin(a_3) - K \sin(a_4) \right], & k \neq 0, \\ \frac{2}{3} (b^{\frac{3}{2}} - K^3), & k = 0, \end{cases}$$

and

$$I_2(k) := \begin{cases} a_5 [\sin(a_3) + \sin(-a_4)], & k \neq 0, \\ (b - K^2), & k = 0, \end{cases}$$

with $k_1 = \frac{k\pi}{b-a}$, $k_2 = \frac{ak\pi}{b-a}$, $a_0 = \frac{\sqrt{\frac{\pi}{2}}}{k_1^{3/2}}$, $a_1 = \sqrt{bk_1 \frac{2}{\pi}}$, $a_2 = K \sqrt{k_1 \frac{2}{\pi}}$, $a_3 = bk_1 - k_2$, $a_4 = K^2 k_1 - k_2$, $a_5 = \frac{1}{k_1}$ and where $C_F(\cdot)$ and $S_F(\cdot)$ are the so-called Fresnel integrals defined as:

$$C_F(x) = \int_0^x \cos(t^2) dt, \quad S_F(x) = \int_0^x \sin(t^2) dt.$$

Proof. After the change of variables, $x = \sqrt{y}$, the integration is straightforward. ■

By semi-analytic expressions for pricing options on the underlying $S(t)$, via Equation (2.11) and on the VIX, in (3.10), we can calibrate the randomized AD models. The pricing formulae depend on the integration ranges $[a, b]$ that needs to be chosen carefully. It is advised to choose it based on the cumulants, which have been derived for VIX derivatives in [22]. We report, however, that calibration of VIX derivatives may result in rather extreme parameters giving rise to numerical instabilities. Since we can utilize the analytically known distribution for $\overline{\text{vix}}$, the pricing may also be performed by directly integrating the payoff function and employing the PDF in (3.8):

$$V_{\text{vix}}(t) = 100 \times 2\alpha_1 e^{-r(T-t)} \int_{\bar{K}} x(x - \bar{K}) f_{\chi^2(\delta, \bar{\kappa}(t, T))}(\alpha_1(x^2 - \alpha_2)) dx,$$

with α_1 and α_2 in (3.8) and δ , $\bar{\kappa}(t, T)$ defined in (3.3). With one of the model parameters stochastic, the RAnD pricing equation reads:

$$V_{\text{vix}}(t) = 100 \times 2\alpha_1 e^{-r(T-t)} \sum_{n=1}^N \omega_n \int_{\bar{K}} x(x - \bar{K}) f_{\chi^2(\delta, \bar{\kappa}(t, T))}(\alpha_1(x^2 - \alpha_2); \theta_n) dx,$$

where θ_n in $f_{\chi^2(\cdot, \cdot)}(\cdot; \theta_n)$ indicates a particular realization of the model parameter and ω_n corresponds to its weight.

3.2. Market data experiment: calibration to S&P500 and VIX options

This section illustrates how the randomized Bates (RAnD Bates) model can fit implied volatilities for S&P 500 and VIX. In the experiment, we consider three randomly selected dates: 02/02/2022, 13/05/2022 and 14/07/2022, for which the calibration exercise was performed. In addition, we have considered short expiry options of one month in all the cases- a challenging task for conventional stochastic volatility models. The calibration was performed in two stages;

in the first step, the *pure* Bates model was calibrated to European S&P 500 options, and such attained parameters form an initial guess for the RAnD Bates model. Then, the optimization was performed with the *target* function, including implied volatilities on S&P 500 and VIX. In the calibration, an additional weight for the ATM level was used. In this calibration stage, we have used the randomized vol-vol parameter γ and randomized it with a uniformly distributed random variable, $\gamma \sim \mathcal{U}([\hat{a}, \hat{b}])$.

The calibration results are presented in Figures 6,7,8. For all three cases, the results were compared against the Bates model (indicated by the red line). We conclude that jumps in the Bates model allow for a good fit for European options on the index (illustrated in the left panels). However, it is also clear that the Bates model does not have sufficient flexibility to fit VIX options (right panels). The standard Bates model gives results significantly outside the bid-ask spreads for all three studied cases.

The calibrated parameters for the RAnD Bates model are reported in Table 4. The results are excellent, i.e., for all three cases, we can calibrate to both products simultaneously, and the implied volatilities for the RAnD Bates model for VIX were well within the bid-ask spreads.

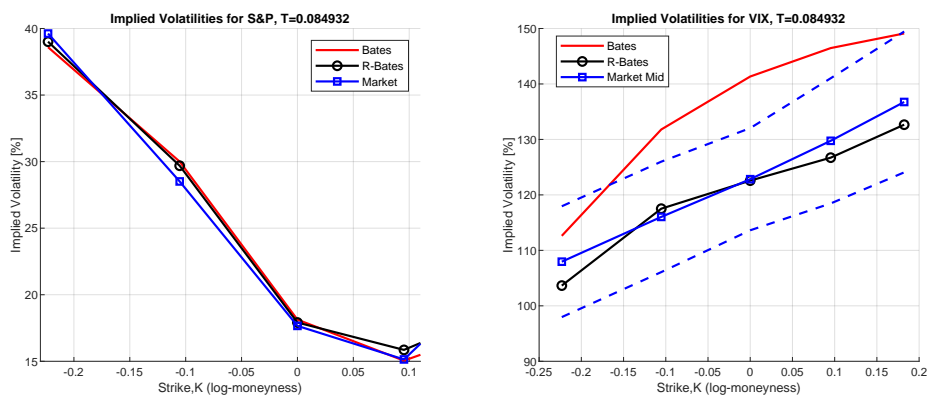


Figure 6: Calibration results of the RAnD Bates model. The implied volatilities for S&P and ViX were obtained on 02/02/2022. Dotted lines indicate bid-ask spreads. Left: S&P, Right: VIX.

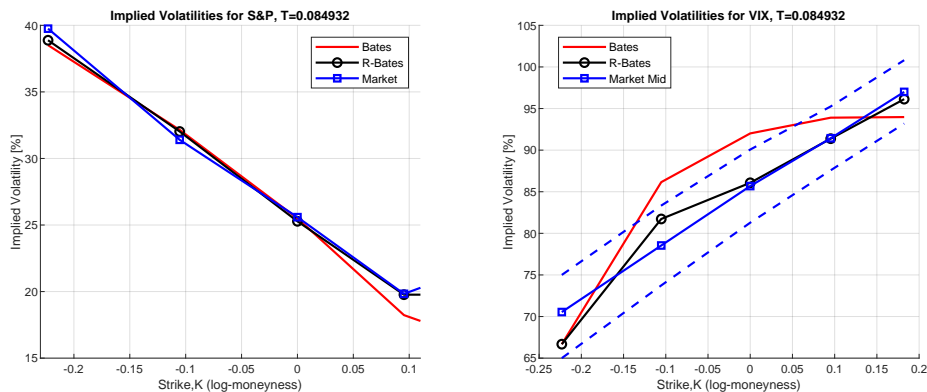


Figure 7: Calibration results of the RAnD Bates model. The implied volatilities for S&P and ViX were obtained on 14/07/2022. Dotted lines indicate bid-ask spreads. Left: S&P, Right: VIX.

It is important to remember that the calibration of the randomized models, especially for models like the Bates model with multiple parameters for stochastic volatility and jumps, is not trivial. For that reason, we have used the standard Bates model parameters as the initial guess for the RAnD framework- otherwise, one may encounter the well-known problem of the possibility of ending up in a local minimum.

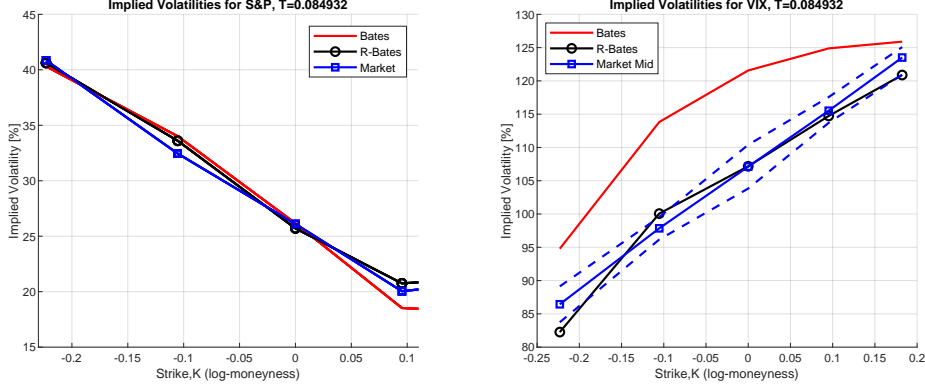


Figure 8: Calibration results of the RAnD Bates model. The implied volatilities for S&P and ViX were obtained on 13/05/2022. Dotted lines indicate bid-ask spreads. Left: S&P, Right: VIX.

Table 4: Parameters determined in calibration of S&P and VIX

Calibrated RAnD Bates parameters								
date	κ	v_0	\bar{v}	ρ	μ_J	σ_J	λ	γ
02/02/2022	0.5	0.170^2	0.23	-0.65	-0.25	0.05	0.25	$\gamma \sim \mathcal{U}([0.01, 2.3])$
13/05/2022	0.14	0.267^2	0.28	-0.8	-0.25	0.02	0.1	$\gamma \sim \mathcal{U}([0.002, 2.1])$
14/07/2022	0.5	0.250^2	0.10	-0.85	-0.25	0.05	0.15	$\gamma \sim \mathcal{U}([0.05, 1.4])$

3.3. Hedging under the RAnD method

Using the pricing Equation (2.11), one can find the hedge parameters Δ and Γ explicitly [26]. However, it is a rather involved task to compute the sensitivity to the model parameters, or the parameters of the randomizer need to be computed.

Typically, when the market data changes, a model needs to be re-calibrated, requiring a trader to *hedge* each parameter. Applying the RAnD method allowed us to make use of randomized processes; it led us to a marginal density for the asset price as a convex combination of affine densities (see Remark 2). Conversely, the option price can be represented as the same convex combination among the corresponding option prices. Moreover, due to the linearity of the derivative operator, the same convex combination applies to all option Greeks. In particular, this ensures that starting from analytically tractable affine densities, one deals with a model that preserves the analytical tractability. Let us consider a scenario in which a model parameter θ follows a specific distribution with a parameter \hat{a} . Under such a randomized model, the sensitivity of the derivative price, $V(t_0)$, should be calculated to the parameter, \hat{a} , that drives the parameter's randomness. The RAnD framework uses the randomized ChF, which must be differentiated to compute the sensitivities. By differentiation of the pricing, Equation (2.11), for $\bar{A} := \bar{A}(u; \tau, \theta_n)$ and $\bar{B} := \bar{B}^T(u; \tau, \theta_n)$, we find:

$$\frac{\partial V(t_0)}{\partial \hat{a}} = e^{-r(T-t_0)} \sum_{k=0}^{N_c-1} \Re \left[\exp \left(-ik\pi \frac{a}{b-a} \right) \frac{\partial}{\partial \hat{a}} \phi_{\mathbf{X}} \left(\frac{k\pi}{b-a}; t_0, T \right) \right] \cdot H_k + \epsilon_{c_2}, \quad (3.11)$$

with:

$$\frac{\partial}{\partial \hat{a}} \phi_{\mathbf{X}}(u; t_0, T) = \sum_{n=1}^N \phi_{\mathbf{X}|_{\vartheta=\theta_n}}(u) \left[\frac{\partial \omega_n}{\partial \hat{a}} + \omega_n \left(\frac{\partial \bar{A}}{\partial \theta_n} + \mathbf{X}(t) \frac{\partial \bar{B}}{\partial \theta_n} \right) \frac{\partial \theta_n}{\partial \hat{a}} \right], \quad (3.12)$$

where $\phi_{\mathbf{X}|_{\vartheta=\theta_n}}(u) = e^{\bar{A}(\mathbf{u}; \tau, \theta_n) + \bar{B}^T(\mathbf{u}; \tau, \theta_n) \mathbf{X}(t)}$, with H_k the payoff coefficient (independent of the model parameters). Error ϵ_{c_1} is a function of ϵ_{c_1} in (2.11) and the quadrature error, ϵ_N , in (2.9). The expression in (3.12) involves derivatives of the ChF with respect to the model parameters $\partial \bar{A} / \partial \theta_n$ and $\partial \bar{B} / \partial \theta_n$, which may to be derived analytically, and the sensitivity to the quadrature pairs, $\{\omega_n, \theta_n\}$, $\partial \theta_n / \partial \hat{a}$ and $\partial \omega_n / \partial \hat{a}$. Since the computation of these pairs is

not analytic, it is recommended to compute these derivatives numerically, with, for example, finite differences:

$$\frac{\zeta(\vartheta(\hat{a} + \delta_{\hat{a}})) - \zeta(\vartheta(\hat{a} - \delta_{\hat{a}}))}{2\delta_{\hat{a}}} \approx \left\{ \frac{\partial \omega_n}{\partial \hat{a}}, \frac{\partial \theta_n}{\partial \hat{a}} \right\},$$

where $\vartheta(\hat{a})$ indicates the dependence of the random variable and parameter \hat{a} , $\delta_{\hat{a}}$ is the *shock* size and $\zeta(\vartheta) : \mathbb{R} \rightarrow \{\omega_n, \theta_n\}_{n=1}^N$ is defined in [Appendix A.1](#). Due to the applied finite difference shocks to \hat{a} , an additional bias will be introduced. We expect, however, this error to be of acceptable magnitude.

3.4. Convergence improvements for Monte Carlo simulation with randomized parameters

Under the RAnD framework, the randomness of the model parameters is not modelled by an AD SDE, but it is enforced “externally” and consequently affects the simulation with Monte Carlo methods and the convergence. In this section, we discuss how to perform Monte Carlo sampling and propose an efficient variance reduction technique that will improve the convergence of the method.

Monte Carlo simulation of the randomized model implies that every simulated path of the underlying process $\mathbf{X}(t)$ will be randomized in a sense that will depend on a realization of ϑ , θ_n . Suppose we consider a discretization of some process $\mathbf{X}(t)$ from (2.1) with M Monte Carlo paths given by:

$$\begin{aligned} n : x_n(t + \Delta t; \theta_n) &= x_n(t; \theta_n) + \int_t^{t+\Delta t} \bar{\mu}(s, x_n(s); \theta_n) ds + \int_t^{t+\Delta t} \bar{\sigma}(s, \mathbf{x}_n(s); \theta_n) d\tilde{w}_n(s) \\ &+ \int_t^{t+\Delta t} J_n(s; \theta_n)^T dx_{\mathcal{P},n}(s), \end{aligned} \quad (3.13)$$

with $\tilde{w}_n(s)$ and $x_{\mathcal{P},n}(s)$ representing the realization of the Brownian motion and Poisson process, respectively. Then, depending on the discretization method, the integrals above can be found explicitly or approximately by either Euler, Milstein or any other higher-order discretization scheme. Since the randomization of model parameters is time-invariant, this implies that every discretized Monte Carlo path of (3.13) will be driven by a different realization of the randomized parameter. Formally, considering a simulation with M paths, every n 'th, $n = 1, \dots, M$, path will depend on the n 'th sample from the parameter distribution,

The computational time can be reduced by following a similar strategy as in [Section 2.3](#). We have two possibilities for derivative pricing under the RAnD framework: directly using all the Monte Carlo paths and evaluate the pricing equation, $V_1(t_0)$ in (3.14), or we can employ the idea behind the RAnD method and divide the pricing problem into N sub-problems, where for each quadrature node $\{\theta_n\}_{n=1}^N$, a separate simulation is performed and weighted according to $\{\omega_n\}_{n=1}^N$ (see $V_2(t_0)$ in [Equation \(3.14\)](#)),

$$V_1(t_0) \approx \frac{e^{-rT}}{M} \sum_{i=1}^M \Pi(x_i(T), T), \quad V_2(t_0) \approx \frac{e^{-rT}}{M} \sum_{n=1}^N \omega_n \sum_{i=1}^M \Pi(x_i(T; \vartheta = \theta_n), T). \quad (3.14)$$

The division into sub-problems gives stability to the model; however, the cost is the quadrature error, affecting the overall pricing error.

In [Figure 9](#), the numerical results for the example case discussed in [Section 2.3](#) are presented. The left panel of [Figure 9](#) confirms that the stability is affected by the randomized volatility coefficient. In the right panel of [Figure 9](#), we see stability by dividing the problem into N sub-problems. Although the division process improves the convergence, it will negatively impact the computational cost (the pricing needs to be performed N times). Finally, a particular division strategy will also depend on the pricing problem.

3.5. RAnD method for piece-wise constant parameters

The methodology to obtain the randomized ChF can be extended to multiple or piece-wise parameters. In such a case, the approach could be considered a rough approximation for a parameter driven by a stochastic process. Using the insight that under the affine models functions $\bar{A}(u, \tau)$ and $\bar{B}(u; \tau)$ in (2.9) are the solution of the Riccati-type ODEs with nonzero initial

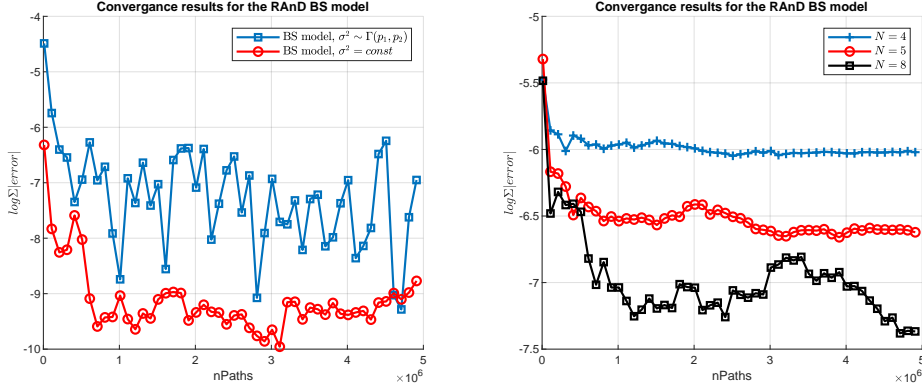


Figure 9: Convergence results for an increasing number of Monte Carlo paths. Left panel: impact of randomization on convergence rate. Right panel: RAnD method convergence with sub-divided pricing.

conditions $\bar{A}(u, 0)$ and $\bar{B}(u, 0)$, for a time-grid $\tau_{n_1} \leq \tau_{n_2} \cdots \leq \tau_{n_m} = \tau$, the model parameters can be evaluated at each interval determined by $T - \tau_i$. Piece-wise constant parameters imply that the characteristic function can be evaluated recursively, i.e. at the first interval from $[0, \tau_{n_1})$, we use the initial conditions $\bar{B}(u, 0) = 0$ and $\bar{A}(u, 0) = 0$. When the corresponding analytic solution is determined (for details see [26], p.521), we obtain two solutions, a_1 and c_1 . For interval $[\tau_1, \tau_2)$, we then assign the initial conditions $\bar{B}(u, \tau_1) = c_1$ and $\bar{A}(u, \tau_1) = a_1$. This procedure will be repeated until the last time step, where the initial values c_{n_m-1} and a_{n_m-1} are used to evaluate $\bar{B}(u, \tau_{n_m})$ and $\bar{A}(u, \tau_{n_m})$.

Piece-wise parameters divided over n_m intervals will require separate handling of each interval. Corollary 3.1 provides the details.

Corollary 3.1 (Piece-wise random parameters). *Consider n_m independent random variables $\vartheta_1, \dots, \vartheta_{n_m}$, and a time grid with n_m intervals of the time interval $[t, T]$ with $\tau = T - t$. Each of the variables corresponds to an interval. Then, the corresponding randomized ChF is given by:*

$$\phi_{\mathbf{X}}(u; t, T) = \sum_{n_1=1}^N \cdots \sum_{n_m=1}^N \omega_{n_1} \dots \omega_{n_m} e^{\bar{A}(u; \tau, \theta_{n_1}, \dots, \theta_{n_m}) + \bar{B}^T(u; \tau, \theta_{n_1}, \dots, \theta_{n_m}) \mathbf{X}(t)} + \bar{\epsilon}_N, \quad (3.15)$$

under the same assumption as specified in Theorem 2.1.

The representation above gives us an extension for n_m independent model parameters. The summation is over all parameters and their corresponding quadrature points; thus, it is subject to the so-called *curse of dimensionality*. Therefore, this expression is not desirable for many stochastic parameters. However, the representation in (3.15) can be reduced using the sparse grid approach (see [31, 14]). The significant advantage of such a technique is that the number of grid points does not grow exponentially with the dimension but only polynomially.

4. Conclusion

In this article, we have introduced the RAnD method for efficient computation of the affine models with random parameters. The proposed framework is generic and can be applied to any stochastic model, even outside the class of affine diffusions. As long as the randomizing random variable gives rise to finite, preferably closed-form, moments, one can price European-style options efficiently. The heart of the method is formed by a few *critical* collocation points to recover the characteristic function. Fast computation of the characteristic function is possible because the method converges exponentially in the number of expansion terms. We have shown that the randomization of stochastic models provides a breeze of fresh air to the class of affine models. Finally, we have applied the RAnD method to the Bates model and shown that randomization allows for simultaneous calibration to S&P and VIX options- a heavily desired feature of modern models.

References

- [1] J. Baldeaux and A. Badran. Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model. *Applied Mathematical Finance*, 21(4):299–312, 2014.
- [2] D. Bates. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *Review of Financial Studies*, 9(1):69–107, 1996.
- [3] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.
- [4] D. Brigo and F. Mercurio. Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, 5(04):427–446, 2002.
- [5] R. Bulirsch and J. Stoer. *Introduction to numerical analysis*, volume 3. Springer, 2002.
- [6] G. Callegaro, L. Fiorin, and M. Grasselli. Quantization meets Fourier: a new technology for pricing options. *Annals of Operations Research*, 282(1):59–86, 2019.
- [7] P. Carr and L. Wu. Stochastic skew in currency options. *Journal of Financial Economics*, 86(1):213–247, 2007.
- [8] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68:1343–1376, 2000.
- [9] J. Favard. Sur les polynomes de Tchebicheff. *C. R. Acad. Sci., Paris*, 200:2052–2053, 1935. ISSN 0001-4036.
- [10] J.-P. Fouque and Y. F. Saporito. Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. *Quantitative Finance*, 18(6):1003–1016, 2018.
- [11] J. Gatheral, P. Jusselin, and M. Rosenbaum. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *arXiv preprint arXiv:2001.01789*, 2020.
- [12] G. H. Golub and J. H. Welsch. Calculation of Gauss quadrature rules. *Mathematics of Computation*, 23(106):221–230, 1969.
- [13] S. Goutte, A. Ismail, and H. Pham. Regime-switching stochastic volatility model: estimation and calibration to vix options. *Applied Mathematical Finance*, 24(1):38–75, 2017.
- [14] L. A. Grzelak. Sparse grid method for highly efficient computation of exposures for xVA. *Applied Mathematics and Computation*, 434:127446, 2022.
- [15] L. A. Grzelak, J. Witteveen, M. Suarez-Taboada, and C. W. Oosterlee. The stochastic collocation Monte Carlo sampler: highly efficient sampling from expensive distributions. *Quantitative Finance*, 19(2):339–356, 2019.
- [16] J. Guyon. The joint S&P 500/VIX smile calibration puzzle solved. *Risk*, April, 2020.
- [17] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Fin. Studies*, 6:327–343, 1993.
- [18] B. Horvath, A. Jacquier, and P. Tankov. Volatility options in rough volatility models. *SIAM Journal on Financial Mathematics*, 11(2):437–469, 2020.
- [19] A. Jacquier and P. Roome. Black-Scholes in a CEV random environment. *arXiv preprint arXiv:1503.08082*, 2015.
- [20] A. Jacquier and F. Shi. The randomized Heston model. *SIAM Journal on Financial Mathematics*, 10(1): 89–129, 2019.
- [21] A. Jacquier, C. Martini, and A. Muguruza. On VIX futures in the rough Bergomi model. *Quantitative Finance*, 18(1):45–61, 2018.
- [22] B. Jing, S. Li, and Y. Ma. Consistent pricing of VIX options with the Hawkes jump-diffusion model. *The North American Journal of Economics and Finance*, 56:101326, 2021.
- [23] T. Kokholm and M. Stisen. Joint pricing of VIX and SPX options with stochastic volatility and jump models. *The Journal of Risk Finance*, 2015.
- [24] V. Lemaire, T. Montes, and G. Pagès. Stationary Heston model: calibration and pricing of exotics using product recursive quantization. *Quantitative Finance*, 22(4):611–629, 2022.
- [25] S. Mechkov. “Hot-Start” initialization of the Heston model. *Risk.net*, 2016, Available at SSRN 2674858, 2016.
- [26] C. W. Oosterlee and L. A. Grzelak. *Mathematical modeling and computation in finance*. World Scientific, 2019.
- [27] G. Pagès and A. Sagna. Recursive marginal quantization of the Euler scheme of a diffusion process. *Applied Mathematical Finance*, 22(5):463–498, 2015.
- [28] A. Papanicolaou and R. Sircar. A regime-switching Heston model for VIX and S&P 500 implied volatilities. *Quantitative Finance*, 14(10):1811–1827, 2014.
- [29] R. Rudd, T. A. McWalter, J. Kienitz, and E. Platen. Fast quantization of stochastic volatility models. *arXiv preprint arXiv:1704.06388*, 2017.
- [30] A. Sepp. Pricing options on realized variance in the Heston model with jumps in returns and volatility. *Journal of Computational Finance*, 11(4):33–70, 2008.
- [31] S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, 4:240–243, 1963.

Appendix A. Additional materials and proofs

Appendix A.1. Moments and optimal collocation pairs: $\zeta(\vartheta) : \mathbb{R} \rightarrow \{\omega_n, \theta_n\}_{n=1}^N$

This section discusses the mechanism for the construction of orthogonal polynomials. Later, this methodology is applied to establish the $\zeta(\vartheta) : \mathbb{R} \rightarrow \{\omega_n, \theta_n\}_{n=1}^N$ that are used in the RAnD method.

A sequence of orthogonal polynomials $\{p_i\}_{i=0}^N$, with $\deg(p_i) = i$, is said to be orthogonal in L^2 with respect to PDF $f_\vartheta(\vartheta)$ of ϑ , if the following holds,

$$\mathbb{E}[p_i(\vartheta)p_j(\vartheta)] = \int_{\mathbb{R}} p_i(x)p_j(x)f_\vartheta(x)dx = \delta_{i,j}\mathbb{E}[p_i^2(\vartheta)], \quad i, j = 0, \dots, N, \quad (\text{A.1})$$

with \mathbb{R} the support of ϑ , $\delta_{i,j}$ the Kronecker delta.

An important property of orthogonal polynomials is their definition in terms of a recurrence relation, given in the theorem below.

Theorem Appendix A.1 (Recurrence in orthogonal polynomials). *For any given density function $f_\vartheta(\cdot)$, a unique sequence of monic orthogonal polynomials $p_i(x)$ exists, with $\deg(p_i(x)) = i$, which can be constructed as follows,*

$$p_{i+1}(x) = (x - \alpha_i)p_i(x) - \beta_i p_{i-1}(x), \quad i = 0, \dots, N-1, \quad (\text{A.2})$$

where $p_{-1}(x) \equiv 0$, $p_0(x) \equiv 1$ and where α_i and β_i is the recurrence coefficients,

$$\alpha_i = \frac{\mathbb{E}[\vartheta p_i^2(\vartheta)]}{\mathbb{E}[p_i^2(\vartheta)]}, \quad \text{for } i = 0, \dots, N-1, \quad \beta_i = \frac{\mathbb{E}[p_i^2(\vartheta)]}{\mathbb{E}[p_{i-1}^2(\vartheta)]}, \quad \text{for } i = 1, \dots, N-1, \quad (\text{A.3})$$

with $\beta_0 = 0$.

Proof. The proof can be found in [9]. ■

Parameters α_i and β_i are entirely determined in terms of the moments of a random variable ϑ . For many densities (weight functions in the integration in (A.4)) the three-term recurrence relation in (A.2) of the corresponding orthogonal polynomials has been determined. In a scenario when the probability density function $f_\vartheta(\cdot)$ is not known explicitly, or its evaluation is computationally expensive. In such circumstances, it is desirable to express α_i and β_i in (A.3) in terms of the moments of ϑ [12]. Let us consider the monomials $m_i(\vartheta) = \vartheta^i$, and define $\mu_{i,j}$ as follows,

$$\mu_{i,j} = \mathbb{E}[m_i(\vartheta)m_j(\vartheta)] = \int_{\mathbb{R}} x^{i+j} f_\vartheta(x)dx = \mathbb{E}[\vartheta^{i+j}], \quad i, j = 0, \dots, N. \quad (\text{A.4})$$

From all moments $\mu_{i,j}$ we construct the so-called Gram matrix $M = \{\mu_{i,j}\}_{i,j=0}^N$, which is symmetric and contains all moments $\{1, \mathbb{E}[\vartheta^1], \dots, \mathbb{E}[\vartheta^{2N}]\}$. Since matrix M is, by definition, positive definite [12], we decompose $M = R^T R$, by the Cholesky decomposition of M .

The following step relates the Cholesky upper triangular matrix R to the orthogonal polynomials. This relationship has been established in [12] and is given by,

$$\alpha_j = \frac{r_{j,j+1}}{r_{j,j}} - \frac{r_{j-1,j}}{r_{j-1,j-1}}, \quad j = 1, \dots, N, \quad \text{and} \quad \beta_j = \left(\frac{r_{j+1,j+1}}{r_{j,j}} \right)^2, \quad j = 1, \dots, N-1, \quad (\text{A.5})$$

with $r_{0,0} = 1$ and $r_{0,1} = 0$ and where $r_{i,j}$ is the (i,j) -th element of matrix R . This relation gives us the manifestations for α_j and β_j when the matrix of moments has been computed.

The recurrence coefficients α_i and β_i are often called the Darboux coefficients. The choice of these coefficients guarantees that the generated polynomials $p_n(\vartheta)$ are orthogonal.

The next step is to relate the coefficients α_i and β_i to the zeros of the orthogonal polynomials $p_n(\vartheta)$, $n = 0, \dots, N$, utilizing the so-called the *eigenvalue method*, presented in the theorem below.

Theorem Appendix A.2 (Eigenvalue method). *The zeros θ_n , $n = 1, \dots, N$, of the orthogonal polynomial $p_N(\vartheta)$ are the eigenvalues of the symmetric tridiagonal matrix,*

$$\hat{J} := \begin{pmatrix} \alpha_1 & \sqrt{\beta_1} & 0 & 0 & 0 \\ \sqrt{\beta_1} & \alpha_2 & \sqrt{\beta_2} & 0 & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & 0 & \sqrt{\beta_{N-2}} & \alpha_{N-1} & \sqrt{\beta_{N-1}} \\ 0 & 0 & 0 & \sqrt{\beta_{N-1}} & \alpha_N \end{pmatrix}, \quad (\text{A.6})$$

i.e., $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)^T$ is a vector of eigenvalues satisfying $\hat{J}\mathbf{v} = \theta_i\mathbf{v}$ with $i = 1, \dots, N$ for any real vector \mathbf{v} , with α_i and β_i being the coefficients of the three-term recurrence relation (A.2).

Proof. The proof can be found in [12]. ■

Based on the coefficients α_i and β_i , the collocation points θ_n are the eigenvalues of matrix (A.6). Eigenvalue calculation for a tridiagonal matrix in Theorem Appendix A.2 is performed by, e.g. the Lanczos algorithm.

The final step is to determine the corresponding weights, ω_n . This can be done based on the Theorem Appendix A.3 below,

Theorem Appendix A.3 (Optimal weights ω_n). *Let $\mathbf{v}^{(n)} := [v_1^{(n)}, \dots, v_N^{(n)}]^T$ be an eigenvector of \hat{J} in (A.6) for the eigenvalue θ_n , $\hat{J}\mathbf{v}^{(n)} = \theta_n\mathbf{v}^{(n)}$. Then, the weights ω_n are given by: $\omega_n = (\mathbf{v}_1^{(n)})^2$, $n = 1, \dots, N$.*

Proof. Proof can be found in [5] (p.179). ■

The pseudo algorithm to compute the pairs $\{\omega_n, \theta_n\}_{n=1}^N$ is provided in Algorithm 1 below.

Algorithm 1: Calculation of optimal quadrature weights and nodes $\{\omega_n, \theta_n\}_{n=1}^N$, The accompanying Python and MATLAB codes can be found at <https://github.com/LechGrzelak/Randomization>.

```
// Construction of the matrix M with size (N + 1 × N + 1) :
for i = 1 ... N + 1 do
    for j = 1 ... N + 1 do
        M[i, j] = E[ϑi+j-2]
    end
end
// Calculate upper diagonal Cholesky matrix from M:
Result: M = RTR
// Find α and β from matrix R, according to Equation (A.5)
α[1] = R[1, 2]/R[1, 1]
β[1] = (R[2, 2]/R[1, 1])2
for i = 2 : N - 1 do
    α[i] = R[i, i + 1]/R[i, i] - R[i - 1, i]/R[i - 1, i - 1];
    β[i] = (R[i + 1, i + 1]/R[i, i])2;
end
α[N] = R[N, N + 1]/R[N, N] - R[N - 1, N]/R[N - 1, N - 1];
// Construction of the matrix Ĵ, according to Equation (A.6):
J = zeros(N, N), J[1, 1] = α[1], J[1, 2] = √β[1]
for i = 2 ... N - 1 do
    J[i, i - 1] = √β[i - 1], J[i, i] = α[i], J[i, i + 1] = √β[i]
end
J[N, N] = α[N], J[N, N - 1] = √β[N - 1];
// Find the collocation points θn and the corresponding weights ωn
θ is computed from x = EigenValues(J)
ω is obtained by squaring the first row from eigenvalues vector
Result: {ωn, θn}n=1N
```

Appendix A.2. Impact of randomizers on implied volatilities

In this section, we illustrate the impact of randomizers on the implied volatilities under the Black-Scholes model. In particular we consider $\sigma \sim \Gamma(\hat{a}, \hat{b})$ and $\sigma \sim \hat{a}\chi^2(\hat{b}, \hat{c})$. The results are reported in Figure A.10.

Appendix A.3. ChF for the Bates model

For the Bates model, the complex valued functions $\bar{C}(u, \tau)$, $\bar{A}_H(u, \tau)$ are known explicitly and are given by:

$$\begin{aligned}\bar{C}(u, \tau) &= \frac{1 - e^{-D\tau}}{\gamma^2(1 - ge^{-D\tau})} (\kappa - \gamma\rho iu - D), \\ \bar{A}_H(u, \tau) &= r(iu - 1)\tau + \frac{\kappa\bar{v}\tau}{\gamma^2} (\kappa - \gamma\rho iu - D) - \frac{2\kappa\bar{v}}{\gamma^2} \log\left(\frac{1 - ge^{-D\tau}}{1 - g}\right), \\ \bar{A}_B(u, \tau) &= \bar{A}_H(u, \tau) - \lambda iu\tau \left(e^{\mu_J + \frac{1}{2}\sigma_J^2} - 1\right) + \lambda\tau \left(e^{iu\mu_J - \frac{1}{2}\sigma_J^2 u^2} - 1\right).\end{aligned}$$

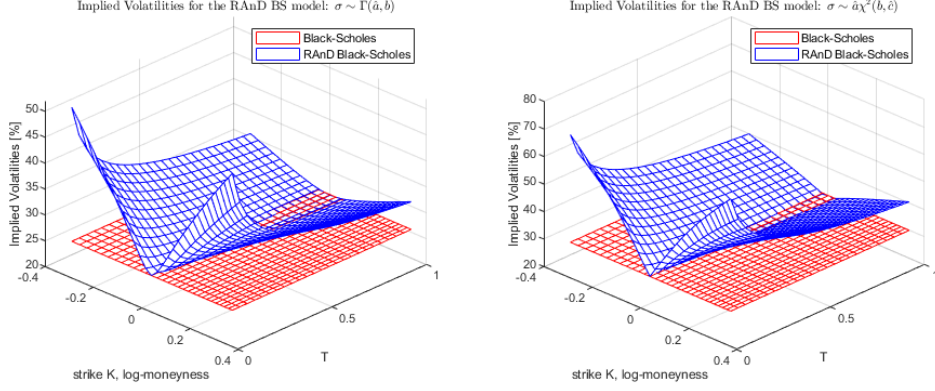


Figure A.10: Implied volatility surface for the RAnD BS model for different randomizers of σ .

for $D = \sqrt{(\kappa - \gamma\rho iu)^2 + (u^2 + iu)\gamma^2}$ and $g = \frac{\kappa - \gamma\rho iu - D}{\kappa - \gamma\rho iu + D}$ and where $\tau = T - t$.

Appendix A.4. Proof of Lemma 3.1

Using the dynamics of the Bates model in Equation (3.1) and the definition of the VIX (3.5) we have:

$$\log \frac{S(T)}{S(t)} = -\lambda \mathbb{E}[e^J - 1](T - t) - \frac{1}{2} \int_t^T v(u) du + \int_t^T \sqrt{v(u)} dW(u) + \int_t^T J dX_{\mathcal{P}}(u), \quad (\text{A.7})$$

with $\mathbb{E}_t[e^J] = e^{\mu_J + \frac{1}{2}\sigma_J^2}$ and all the model parameters as specified in (3.1). By taking the expectations, we find,

$$\mathbb{E}_t \left[\log \frac{S(T)}{S(t)} \right] = -\lambda \left(e^{\mu_J + \frac{1}{2}\sigma_J^2} - \mu_J - 1 \right) (T - t) - \frac{1}{2} \int_t^T \mathbb{E}_t[v(u)] du. \quad (\text{A.8})$$

The expectation, conditioned on time t , of the CIR process is known analytically and it is given by:

$$\mathbb{E}_t[v(T)] = \bar{c}(t, T)(\delta + \bar{\kappa}(t, T)),$$

$$\bar{c}(t, T) = \frac{1}{4\bar{\kappa}} \gamma^2 (1 - e^{-\kappa(T-t)}), \quad \delta = \frac{4\kappa\bar{v}}{\gamma^2}, \quad \bar{\kappa}(t, T) = \frac{4\kappa v(t) e^{-\kappa(T-t)}}{\gamma^2 (1 - e^{-\kappa(T-t)})}. \quad (\text{A.9})$$

By integration over the expectation in Equation (A.8) and substitution to definition of VIX in (3.5) we find:

$$\text{vix}^2(t, T) = 100^2 \times \frac{-2}{T-t} \mathbb{E}_t \left[\log \frac{S(T)}{S(t)} \right] = a(t, T)v(t) + b(t, T) + c, \quad (\text{A.10})$$

with

$$a(t, T) = \frac{1 - e^{-\kappa(T-t)}}{\kappa(T-t)}, \quad b(t, T) = \bar{v} (1 - a(t, T)), \quad c = 2\lambda \left(e^{\mu_J + \frac{1}{2}\sigma_J^2} - \mu_J - 1 \right).$$

Since the CIR process $v(t)|v(t_0)$, $t > 0$, follows a scaled, with $\bar{c}(t_0, t)$, non-central chi-square distribution, $\chi^2(\delta, \bar{\kappa}(t_0, t))$, where δ is the “degrees of freedom” parameter and the noncentrality parameter is $\bar{\kappa}(t_0, t)$, i.e.,

$$v(t)|v(t_0) \sim \bar{c}(t_0, t) \chi^2(\delta, \bar{\kappa}(t_0, t)), \quad t > t_0. \quad (\text{A.11})$$

Therefore the CDF for VIX is given by:

$$\begin{aligned} \mathbb{P}[\text{vix}(t, T) \leq x] &= \mathbb{P} \left[\sqrt{a(t, T)v(t) + b(t, T) + c} \leq x \right] \\ &= \mathbb{P} \left[v(t) \leq \frac{1}{a(t, T)} (x^2 - b(t, T) - c) \right] = F_{v(t)} \left(\frac{1}{a(t, T)} (x^2 - b(t, T) - c) \right). \end{aligned}$$

Since the unconditional CDF for $v(t)$ is known explicitly, we have

$$F_{v(t)}(x) = \mathbb{Q}[v(t) \leq x] = \mathbb{Q}\left[\chi^2(\delta, \bar{\kappa}(t_0, t)) \leq \frac{x}{\bar{c}(t_0, t)}\right] = F_{\chi^2(\delta, \bar{\kappa}(t_0, t))}\left(\frac{x}{\bar{c}(t_0, t)}\right), \quad (\text{A.12})$$

thus finally we have:

$$F_{\text{vix}}(x) = F_{\chi^2(\delta, \bar{\kappa}(t_0, t))}\left(\frac{x^2 - b(t, T) - c}{a(t, T)\bar{c}(t_0, t)}\right).$$

By differentiation, the probability density function reads:

$$\begin{aligned} f_{\text{vix}}(x) &:= \frac{d}{dx} F_{\chi^2(\delta, \bar{\kappa}(t_0, t))}\left(\frac{x^2 - b(t, T) - c}{a(t, T)\bar{c}(t_0, t)}\right) \\ &= \frac{2x}{a(t, T)\bar{c}(t_0, t)} f_{\chi^2(\delta, \bar{\kappa}(t_0, t))}\left(\frac{x^2 - b(t, T) - c}{a(t, T)\bar{c}(t_0, t)}\right) =: 2\alpha_1 x f_{\chi^2(\delta, \bar{\kappa}(t_0, t))}(\alpha_1(x^2 - \alpha_2)), \end{aligned}$$

with $\alpha_1 = \frac{1}{a(t, T)\bar{c}(t_0, t)}$ and $\alpha_2 = b(t, T) + c$.

Having all the ingredients at hand we derive the ChF for $\text{vix}^2(t, T)$. By definition we find:

$$\begin{aligned} \phi_{\text{vix}^2(t, T)}(u) &= \mathbb{E}_{t_0} \left[e^{iu(a(t, T)v(t) + b(t, T) + c)} \right] \\ &= e^{iu(b(t, T) + c)} \mathbb{E}_{t_0} \left[e^{iua(t, T)v(t)} \right] = e^{iu(b(t, T) + c)} \mathcal{M}_{v(t)}(iua(t, T)), \end{aligned}$$

where $\mathcal{M}_{v(t)}(iua(t, T))$ is the moment-generating function of $v(t)$, known explicitly (see [26] p.314):

$$\mathcal{M}_{v(t)}(u) = \mathbb{E}_{t_0} \left[e^{uv(t)} \right] = \left(\frac{1}{1 - 2u\bar{c}(t_0, t)} \right)^{\frac{1}{2}\delta} \exp\left(\frac{u\bar{c}(t_0, t)\bar{\kappa}(t_0, t)}{1 - 2u\bar{c}(t_0, t)}\right).$$

Appendix A.5. Implied volatilities for RAnD BS model with uniform randomizer

Table A.5: Maximum implied volatility, $\text{IV}(\cdot)$, error computed against Monte Carlo results with 10 million samples, with $K = S_0 e^{0.1\sqrt{T}\delta}$, $\delta = [-3, -2, -1, -0.5, 0, 0.5, 1, 2, 3]$.

$\sigma \sim \mathcal{U}([a, b])$	error = $\max_i \text{IV}(t_0, K_i) - \text{IV}_{\text{ref}}(t_0, K_i) $							
	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$
$T = 1d$	0.16 %	0.04 %	0.02 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
$T = 1w$	0.16 %	0.04 %	0.01 %	0.01 %	0.00 %	0.00 %	0.00 %	0.00 %
$T = 2w$	0.16 %	0.03 %	0.02 %	0.02 %	0.02 %	0.02 %	0.02 %	0.02 %
$T = 1m$	0.15 %	0.03 %	0.01 %	0.01 %	0.00 %	0.01 %	0.01 %	0.01 %
$T = 3m$	0.15 %	0.03 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
$T = 6m$	0.15 %	0.03 %	0.01 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
$T = 12m$	0.15 %	0.03 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %

Appendix A.6. Impact of randomization on Bates model parameters

The impact of different randomizers on implied volatilities under the Bates model is presented in Figure A.11 and Figure A.12.

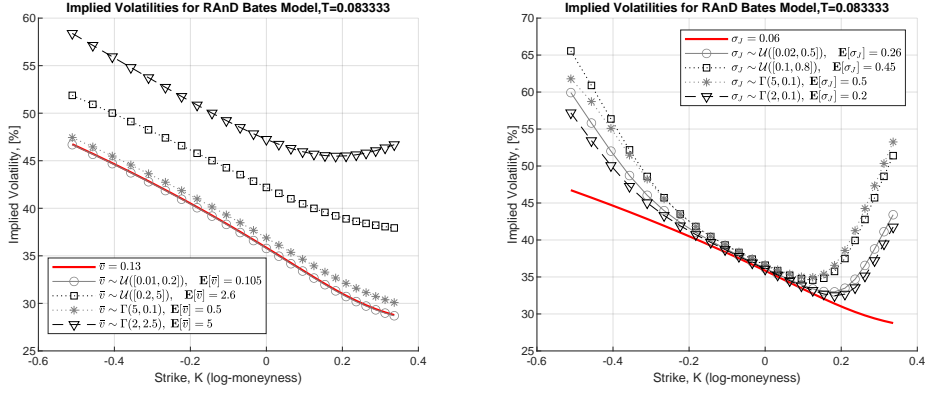


Figure A.11: Impact of randomized parameters on implied volatilities. Left: randomized long-term vol, \bar{v} . Right: randomized jump's size vol, σ_J .

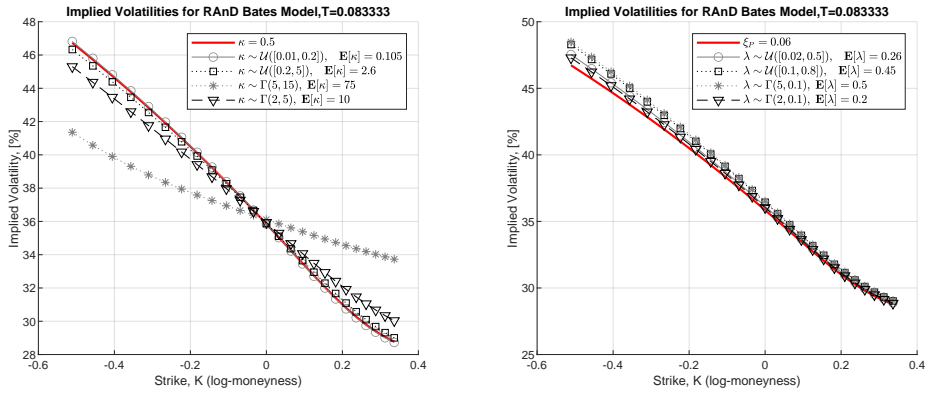


Figure A.12: Impact of randomized parameters on implied volatilities. Left: randomized speed of mean reversion, κ . Right: randomized jump's intensity, λ .