Pricing Exotic Options using Monte Carlo Techniques

Oliver Bandosz, Tim Wolstenholme

Abstract

Something about pricing exotic options and a summary on what we cover in the paper...

Contents

1 Introduction

Quantitative finance and algorithmic trading has expanded dramatically over the last two decades. Automated, model-driven strategies now dominate many markets. For example, an estimated 70% of U.S. equity trading volume is initiated through algorithmic trading (Yadav, 2015). Global algorithmic trading activity continues to grow at double-digit rates annually, which is leading to a reliance on complex models. Options and other derivatives markets have similarly surged in activity. Cboe Global Markets reported a record 3.8 billion options contracts traded in 2024 (averaging 14.95 million contracts daily), marketing the fifth consecutive year of record-breaking volume.[1]

Quantitative hedge funds and trading firms that levrage these models have achieved remarkable success. Industry leaders like Renaissance Technologies and Marshall Wace delivered double-digit returns (over 20%) in 2024, outperforming many traditional asset managers. In fact, Marshall Wace's Eureka fund earned approximately 14.3%, and its Market Neutral Tops fund returned 22.6%, and its Alpha Plus fund gained 15.9% in 2024.[2] The scale of assets under management (AUM) at top quantitative firms has grown according, Citadel LLC manages about $397 billion as of August 2024,[3] and its market-making arm (Citadel Securities) reportedly generated a record $9.7 billion in trading revenue in 2024, a 55% increase from last year.[4] Such figures reflect not only impressive performance, but also the breadth of algorithmic option trading in today's markets. In particular, leading firms have built sophisticated option pricing engines (usually by stochastic models) that take advantage of small pricing inefficiencies at massive scale. This fusion of mathematical finance and computatiuon has been a key driver behind the success of quant hedge funds and the efficiency of derivative markets.

Despite these advances, traditional approaches to option pricing and model calibration face significant challenges when confronted with the complexity of real markets. Classical parameter estimation techniques such as *maximum likelihood estimation (MLE)* or least-squares calibration to market option prices encounter computational and statistical difficulties. Complex option pricing models that have stochastic volatility or jump processes, for example, have intractable likelihood functions that require integrating over unobserved state variables (like the entire path of volatility), which quickly becomes impractical in high dimensions (Brouillon et al., 2024). As the number of model parameters grows, conventional optimisation methods struggle. For example, the likelihood surfaces can be highly non-convex with multiple local optima, and it is difficult to establish convergence that guarantees for finding the true optimum. In practice, calibrating advanced models often demands brute-force searches or problem-specific approximations, which become computationally expensive as model complexity increases. Moreover, classical calibration typically creates point estimates that do not measure the uncertainty in the parameters. This is problematic in risk management and inference, where understanding the range of plausible parameter values and their joint confidence region is as important

[1] https://ir.cboe.com/news/news-details/2025/Cboe-Global-Markets-Reports-Trading-Volume-for-December-and-Full-Year-2024/

[2] https://www.reuters.com/markets/hedge-funds-deliver-double-digit-returns-2024-2025-01-02/#:∼:text=British%20hedge%20fund%20Marshall%20Wace,Eureka%20fund%2C%20a%20source%20said.

[3] https://www.investopedia.com/articles/personal-finance/011515/worlds-top-10-hedge-fund-firms.asp

[4] https://www.bloomberg.com/news/articles/2025-03-06/citadel-securities-9-7-billion-trading-revenue-passes-barclays

as finding a single best fit.

Similar issues arise on the pricing side with classical deterministic techniques. *Partial differential equation (PDE) solves* and lattice methods such as *binomial/trinomial tress* provide elegant solutions for simple option models, but they suffer from *th curse of dimensionality*/ The dimensionality of a PDE grows with the number of underlying risk factors of path dependencies, and standard grid-based methods become infeasible beyond a few dimensions (Han et al., 2018). This curse of dimensionality exponentially grows in computational cost with each additional underlying, making classical PDE approaches impractical for multi-asset or path-dependent options (such as American options). Although closed-form pricing formulas exist for a few idealised cases (such as *Black-Scholes* for European options), many exotic options have no analytic solution. Traders often resort to *Monte Carlo simulation* for such cases, however, straightforward Monte Carlo can converge slowly and embedding it inside an optimisation loop for calibration increases the computation load. Additionally, likelihood based inference for stochastic volatility models may require integrating out latent variables at each step, which closed-form solutions are rarely available for. These limitations of classical methods show that there is a need for more flexible, simulation-based approaches that can handle high-dimensional integrations and complex posterior spaces.

Over the past two decades, *Bayesian methods* powered by *Markov Chain Monte Carlo (MCMC)* have increased as an alternative for inference in quantitative finance. In a Bayesian framework, we treat model parameters and latent states as random variables with a prior distribution, and update this to a posterior distribution given observed data (such as historical prices). MCMC algorithms are then used to sample from these posterior distributions, which effectively solves inference problems by simulation rather than by analytical integration. MCMC evaluates the model's likelihood function (up to a constant) and the prior, it does not require closed-form solutions for the posterior or simplifying linear approximations. This approach avoids many of the problems of classical methods such as high-dimensional integrations, which is handled by random sampling, and there is no need to linearise nonlinear models to derive explicit formulae for complex likelihoods. In the context of financial models, researchers have found that Bayesian and MCMC methods are popular in option pricing models, especially those extending the Black-Scholes-Merton framework to include jumps and stochastic volatility.

The fundamental MCMC technique is the *Metropolis-Hsting (MH)* algorithm, which is often combined with *Gibbs sampling* for models that include latent variables. Each MH step proposes a change in some parameter or state and accepts or rejects it based on the *Metropolis-Hastings acceptance probability* to ensure the chain converges to the true posterior. Through these iterative simulations, the MCMC approach builds up a representative sample from the posterior distribution of all unknown quantities. The outcome is an entire distribution of parameter values consistent with the data, from which credibility intervals and other measures of uncertainty can be derived. This ability to measure parameter uncertainty is a major advantage of the Bayesian MCMC approach over classical point estimation.

Over time, researchers have developed a variety of enhanced MCMC techniques to improve efficiency and convergence in financial applications. One important class is *adaptive MCMC* algoirthms, which automatically tune their proposal distributions per iteration (Haario et al., 2001). Such adaptive schemes help in high-dimensional or strongly correlated parameter spaces, where a fixed proposal might either mix too slowly or be prone to rejection. Another powerful innovation is *Hamiltonian Monte Carlo (HMC)*, which leverages gradient informaiton from the target density to guide the sampling. HMC introduces auxiliary momentum variables and uses *Hamiltonian dynamics* to propose distant, yet high-probability, moves in the parameter space, which dramatically reduces the random walk behaviour of traditional MCMC (Hoffman et al., 2014). HMC can produce more effective samples per iteration by weakening the dependency between successive samples and imporving exploration efficiency. This feature is especially useful in complex posterior spaces such as those in high-dimensional stochastic volatility models. These samplers, along with other MCMC techniques like *particle MCMC* (which combines particle filters with MCMC for state-space models) (Andrieu et al., 2010) and *reversible jump MCMC*, have greatly expanded the toolkit for Bayesian computation in finance. Applications have ranged from Bayesian estimation of Heston's stochastic volatility model and Merton's jump diffusion model to filtering problems in interest rate and credit models.

In this paper, we build on these developments to study MCMC methods in exotic option pricing. We present a self-contained formulation and empirical evaluation covering both well-known option pricing models and modern MCMC techniques, applied to equity and commodity options. We begin with laying out the theoretical framework for a range of option pricing models. starting with the classic Black-Scholes model and then consider progressively richer models, including Merton's jump-diffusion model and Heston's stochastic volatility model. We additionally include the foundational knowledge of *Birth Death processes*, which will play an important role later in the paper. We then introduce the Bayesian inference approach and the specific MCMC algorithms employed. We formulate a Bayesian estimation problem for the parameters of the aforementioned models, given observed data. We specify prior distributions for model parameters and derive the likelihood functions for each model. The construction of the MCMC sampler is then described, including the choice of proposal kernels and any augmentation with latent variables. In particular, we discuss how latent volatility paths or latent jump indicators are introduced to facilitate inference in the Heston and jump-diffusion models, respectively. We employ a *Metropolis-Hastings within Gibbs sampler* for most parameters along with *adaptive proposal tuning* to achieve reasonable acceptance rates. For the Heston model, we also experiment with a Hamiltonian Monte Carlo step to efficiently sample the highly correlated volatility parameters. The goal of this section is to provide a clear blueprint of the inferential procedure, which we will integrate with the option pricing models of Section 2. Section 4 conducts analytical valuation of the exotic contracts considered in our study. We collect closed-form and transform-based results for geometric Asian options and Margrabe's exchange option under the models of Section 2. In particular, we derive the Black-Scholes lognormal formulas, the Merton mixture/conditional-normal representation, and the Heston formulas obtained via *Feynman-Kac arguments* (for Asians) (Shreve et al., 2004) and Fourier mthods based on the characteristic function of the integrated variance (for Margrabe) (Heston, 1993). Finally, we conclude this paper with tests on the MCMC algorithms on empirical data. First, we perform controlled experiments with synthetic exotics (Asian and Margrabe) to verify identifiably and coverage. Option values are generated from one model while inference is carried out under an alternative, and we examione posterior recovery, credible sets, and posterior predictive pricing errors, using corn (CORN) and soya bean (SOYB) data.

2 Stochastic Models

2.1 Birth Death Processes

2.2 Risk free Assets

In this section we introduce a brief foundation on financial mathematics and the models which we will be looking at. We begin with a market consisting of a *risk-free asset*, with price B_t given by

$$dB_t = rB_t dt,$$

where $r > 0$ a constant parameter to model interest.

2.3 Black-Scholes

2.4 Merton Jump Diffusion

2.5 Heston Model

3 Markov Chain Monte Carlo Techniques

3.1 Metropolis Hastings

Take a Black-Scholes market with one risky asset, modeled as:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t,$$

where $\{W_t\}_{t\geq 0}$ is the Wiener process. By Girsanov's theorem, under the risk-neutral measure $\mathbb{Q}$, we model the stock as

$$\frac{dS_t}{S_t} = r dt + \sigma dW_t.$$

By *Ito-Lemma*, we can obtain the log prices X_t:

$$dX_t = \left(r - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t.$$

We here apply *Euler's discretisation* to

$$X_{t+\Delta t} = X_t + (r - \frac{\sigma^2}{2})\Delta t + \sigma\sqrt{\Delta t} Z_t,$$

$$X_{t+\Delta t} | X_t \sim \mathcal{N}\left(\left(r - \frac{\sigma^2}{2}\right)\Delta t, \sigma^2 \Delta t\right).$$

Here we can take the Loglikelihood with respect to the parameters $\boldsymbol{\theta}$ as:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{X}|\boldsymbol{\theta}) = \log \ell(\boldsymbol{X}|\boldsymbol{\theta}) &= \sum_{t=0}^{T-1} \log \mathbb{P}(X_t|\boldsymbol{\theta}) \\
&= T \log\left(\frac{1}{\sqrt{2\pi\sigma^2\Delta t}}\right) - \sum_{t=0}^{T-1} \frac{\left(X_{t+\Delta t} - X_t - \left(r - \frac{\sigma^2}{2}\right)\Delta t\right)^2}{2\sigma^2 \Delta t}.
\end{aligned}$$

Since volatility is non-negative ($\sigma > 0$), we take the log prior as the log normal:

$$\begin{aligned}
\sigma \sim \text{LogNormal}(m, s^2), \quad \log \pi(\sigma) &= \log\left[\frac{1}{\sigma\sqrt{2\pi s^2}} \exp\left[-\frac{(\ln \sigma - m)^2}{2s^2}\right]\right] \\
&= -\ln \sigma - \frac{1}{2}\ln(2\pi s^2) - \frac{(\ln \sigma^2 - m)^2}{2s^2}.
\end{aligned}$$

Hence, by taking $\boldsymbol{\theta} = \sigma$, the log posterior, with $c \in \mathbb{R}$ constant, is:

$$\log \pi(\sigma|\boldsymbol{X}) = \mathcal{L}(\boldsymbol{X}|\sigma) + \log \pi(\sigma) + c.$$

Now suppose that we have K iterations in our MCMC algorithm and $\sigma^{(k)}$ at the k iteration with propose σ' from a proposal density $q(\sigma'|\sigma^{(k)})$. Then the Metropolis Hastings acceptance probability is

$$\alpha(\sigma^{(k)}, \sigma') = \min\left(1, \frac{\pi(\sigma'|\boldsymbol{X})q(\sigma^{(k)}|\sigma')}{\pi(\sigma^{(k)}|\boldsymbol{X})q(\sigma'|\sigma^{(k)})}\right),$$

which can be written in log form as

$$\log \alpha(\sigma^{(k)}, \sigma') = \min \left(0, \log \pi(\sigma'|\boldsymbol{X}) + \log q(\sigma^{(k)}|\sigma') - \log \pi(\sigma^{(k)}|\boldsymbol{X}) - \log q(\sigma'|\sigma^{(k)})\right).$$

The acceptance probability allows us to calculate the *empirical acceptance rate* for the algorithm. This allows us to determine how many proposed $\boldsymbol{\theta}$ terms are accepted through all iterations. At iteration k, we propose (for the Black-Scholes model), $\sigma' \sim q(\cdot|\sigma^{(k)})$ and accept it with probability $\alpha(\sigma^{(k)}, \sigma')$. We then draw a uniform $u \sim \text{Uniform}(0,1)$, such that if $u < \alpha(\sigma^{(k)}, \sigma')$ we accept the proposal and make $\sigma^{(k+1)} = \sigma'$. Otherwise, we reject the proposal and set $\sigma^{(k+1)} = \sigma^{(k)}$. This then leads us to the empirical acceptance rate:

$$AR = \frac{1}{K}\sum_{k=1}^{K} \mathbf{1}\{\text{move at } k \text{ was accepted}\}.$$

It is important to note that in practical cases, we aim to have the acceptance rate around 0.2-0.4.

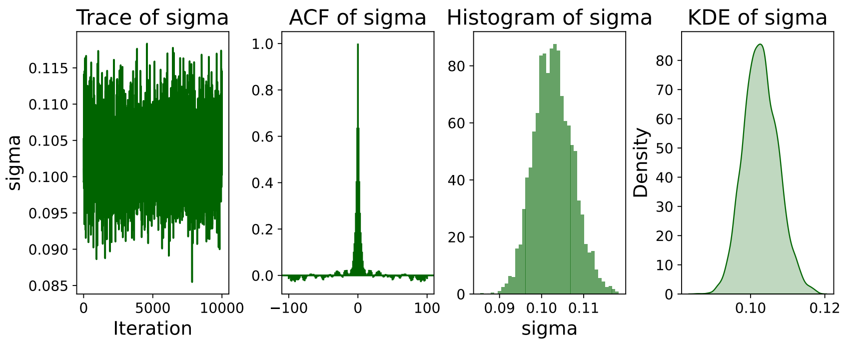

Figure 1: NEED TO WRITE CAPTION... Also Need to redo labels and axes - Too small

3.1.1 Effective Sample Size (ESS)

When working with MCMC algorithms, the successive MCMC draws $\{\boldsymbol{\theta}^{(k)}\}_{k=1}^{K}$ are serially correlated. The *Effective Sample Size (ESS)* allows us to quantify how many independent samples the chain is worth. Recall that for a sequence of independent identically distributed samples from π, $\boldsymbol{\theta}^{(\mathbf{1})}, \ldots, \boldsymbol{\theta}^{(K)}$, if we set $h(\boldsymbol{\theta})$ as a scalar summary, then we can define the *lag-t autocorrelation* as

$$\rho_t = \frac{\mathbb{C}\text{ov}(h(\boldsymbol{\theta}^{(i)}), h(\boldsymbol{\theta}^{(i+t)}))}{\mathbb{V}\text{ar}(h(\boldsymbol{\theta}))}.$$

We then define the ESS for $h(\boldsymbol{\theta})$ as

$$ESS = \frac{K}{1 + 2\sum_{t=1}^{L} \rho_t}.$$

Typically we examine the ESS for each component of $\boldsymbol{\theta}$ to ascertain whether sufficient mixing has occurred. A good ESS is one that allows us to calculate what you need to the required accuracy. Often an ESS of 500 is deemed sufficient, although an ESS below 100 means that inference from the MCMC sample should be treated with caution.

Table 1: Summary Statistics

Parameter	Mean	Standard Deviation	Effective Sample Size (ESS)
σ	0.1027	0.0045	2 134.5112

3.2 Metropolis-within-Gibbs

When applying MCMC to models with many parameters and latent variables, we apply *Gibbs sampling*. Gibbs sampling iteratively samples each parameter from its conditional distribution given all other variables. As such, we iteratively update latent states in blocks based on these distributions. However, when we do not have a closed form distribution, we apply Metropolis-Hastings for that block. This is known as *Metropolis-within-Gibbs*. The idea is we cycle through components of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$, and for each θ_i we perform a MH update to get the conditional density $\pi(\theta_i \mid \theta_{-i})$, for $i = 1, ..., p$, which will given the full joint posterior $\pi(\theta_1, ..., \theta_p)$ invariant. We use Gibbs for the rest of our pricing models due to the high dimensionality of both the options and complexity of the models.

Choosing a block is flexible, we can update one parameter at a time, or a group of highly correlated parameters together. For example, in the Heston model, we can use a Gibbs strategy that alternatives between sampling the volatility latent path $v_{0:T}$ given the parameters, and sampling the parameters $(\kappa, \theta, \xi, \rho)$ given the volatility path and price data. If we take the Heston model (without volatility jumps) below:

$$\begin{aligned}
dS_t &= \mu S_t dt + S_t \sqrt{v_t} dW^{\mathbb{P}}_{t,2}, \\
dv_t &= \kappa(\xi - v_t)dt + \sigma\sqrt{v_t} dW^{\mathbb{P}}_{t,2}, \\
d\langle W^{\mathbb{P}}_{t,1}, W^{\mathbb{P}}_{t,2}\rangle &= \rho dt,
\end{aligned}$$

where $2\kappa\xi > \sigma^2$, $\kappa, \sigma, \xi, \sqrt{v_t} > 0$, and $\rho \in (-1, 1)$. We can take $\boldsymbol{\theta} = (\kappa, \sigma, \xi, \rho)$By applying Girsanov's theorem[5], we convert into risk neutral measure:

$$\begin{aligned}
dS_t &= rS_t dt + \sqrt{v_t} S_t dW^{(1)}_t, \\
dv_t &= \kappa(\xi - v_t)dt + \sigma\sqrt{v_t} dW^{(2)}_t, \\
d\langle W^{(1)}_t, W^{(2)}_t\rangle &= \rho dt,
\end{aligned}$$

where $r > 0$ is the risk-free rate, and $W^{(1)}_t, W^{(2)}_t$ are the Wiener processes under the risk neutral measure $\mathbb{Q}$. Here the logarithmic prices $X_t = \ln S_t$ are:

$$dX_t = \left(r - \frac{v_t}{2}\right) dt + \sqrt{v_t} dW^{(1)}_t$$

Under the Euler discretisation, the discrete time approximation increments are:

$$\Delta X_t := X_{t+1} - X_t, \quad \Delta v_t := v_{t+1} - v_t,$$

are approximately jointly normal with mean

$$\mathbb{E}[\Delta X_t] = \left(r - \frac{1}{2}v_t\right)\Delta t = \mu_{X,t}\Delta t, \quad \mathbb{E}[\Delta v_t] = \kappa(\xi - v_t)\Delta t = \mu_{v,t}\Delta t,$$

[5] See appendix A

with the covariance matrix

$$\Sigma_t = \begin{pmatrix} v_t\Delta t & \rho\sigma v_t\Delta t \\ \rho\sigma v_t\Delta t & \sigma^2 v_t\Delta t \end{pmatrix} = \begin{pmatrix} \mathbb{V}\text{ar}(X_t) & \mathbb{C}\text{ov}(X_t, v_t) \\ \mathbb{C}\text{ov}(X_t, v_t) & \mathbb{V}\text{ar}(v_t) \end{pmatrix},$$

hence the joint PDF:

$$f(\Delta X_t, \Delta v_t, \boldsymbol{\theta}) = \frac{1}{2\pi\sqrt{\det \Sigma_t}} \exp\left(-\frac{1}{2}\left(\begin{pmatrix}\Delta X_t \\ \Delta v_t\end{pmatrix} - \begin{pmatrix}\mu_{X,t} \\ \mu_{v,t}\end{pmatrix}\right)^T \Sigma_t^{-1}\left(\begin{pmatrix}\Delta X_t \\ \Delta v_t\end{pmatrix} - \begin{pmatrix}\mu_{X,t} \\ \mu_{v,t}\end{pmatrix}\right)\right) \tag{1}$$

We can rewrite the joint density as:

$$f(\Delta X_t, \Delta v_t) = f(\Delta X_t|\Delta v_t)f(\Delta v_t),$$

and if we expand (1), we can rewrite it such that:

$$f($$

Hence,

$$\begin{aligned} \Delta X_t \mid \Delta v_t &\sim \mathcal{N}\left(\mu_X + \frac{\mathbb{C}\text{ov}(X_t, v_t)}{\mathbb{V}\text{ar}(X_t)},\quad \mathbb{V}\text{ar}(v_t) - \frac{(\mathbb{C}\text{ov}(X_t, v_t))^2}{\mathbb{V}\text{ar}(X_t)}\right), \\ \Delta v_t|\Delta X_t &\sim \mathcal{N}\left(\mu_v + \frac{\mathbb{C}\text{ov}(X_t, v_t)}{\mathbb{V}\text{ar}(v_t)},\quad \mathbb{V}\text{ar}(X_t) - \frac{(\mathbb{C}\text{ov}(X_t, v_t)^2}{\mathbb{V}\text{ar}(v_t)}\right), \end{aligned} \tag{2}$$

Thus our likelihood for both X and v respectively is:

When calculating the priors for $\boldsymbol{\theta} = (\sigma, \kappa, \xi, \rho)$ we have to take into account the constraints of the parameters. Each parameter in the Heston model follow the following constraints:

$$\sigma > 0, \quad \kappa > 0, \quad \xi > 0, \quad \rho \in (-1, 1)$$

Hence, we have to adopt the bijections:

$$\begin{aligned} \sigma &= \exp(\theta_1), \\ \kappa &= \exp(\theta_2), \\ \xi &= \exp(\theta_3), \\ \rho &= \tanh(\theta_4), \end{aligned}$$

where each $\theta_i \in \mathbb{R}$ is unconstrained. Hence our inverse mapping is

$$\theta_1 = \ln(\sigma), \quad \theta_2 = \ln(\kappa), \quad \theta_3 = \ln(\xi), \quad \theta_4 = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right).$$

If we want to transform the densities from $\boldsymbol{\theta}_{unc}$ back to $\boldsymbol{\theta}$, then we require the Jacobian determinant:

$$\left|\det\left(\frac{d\boldsymbol{\theta}}{d\boldsymbol{\theta}_{unc}}\right)\right| = \left|\det\begin{pmatrix} \frac{\partial\sigma}{\partial\theta_1} & 0 & 0 & 0 \\ 0 & \frac{\partial\kappa}{\partial\theta_2} & 0 & 0 \\ 0 & 0 & \frac{\partial\xi}{\partial\theta_3} & 0 \\ 0 & 0 & 0 & \frac{\partial\rho}{\partial\theta_4} \end{pmatrix}\right|.$$

This gives us the following

$$\left|\det\left(\frac{d\boldsymbol{\theta}}{d\boldsymbol{\theta}_{unc}}\right)\right| = \sigma\kappa\xi(1-\rho^2).$$

Interestingly, the Jacobian matrix cancels out if we place our priors on $\boldsymbol{\theta}_{unc}$ directly, and hence doesn't appear in the acceptance ratio. This two-block Gibbs can be broken down further by breaking down the volatility path into smaller segments, such that we update (in each iteration) each individual volatility path given the parameters and data. This is known *Single-site Metropolis*.

Table 2: Summary Statistics

Parameter	Mean	Standard Deviation	Effective Sample Size (ESS)
σ	0.0024	0.0252	111.6038
κ	8 297 784.4025	9 960 223.6531	2.8301
ξ	7 909 900 299 532.2783	11 693 122 264 912.1875	3.2745
ρ	-0.8547	0.3487	7.8790

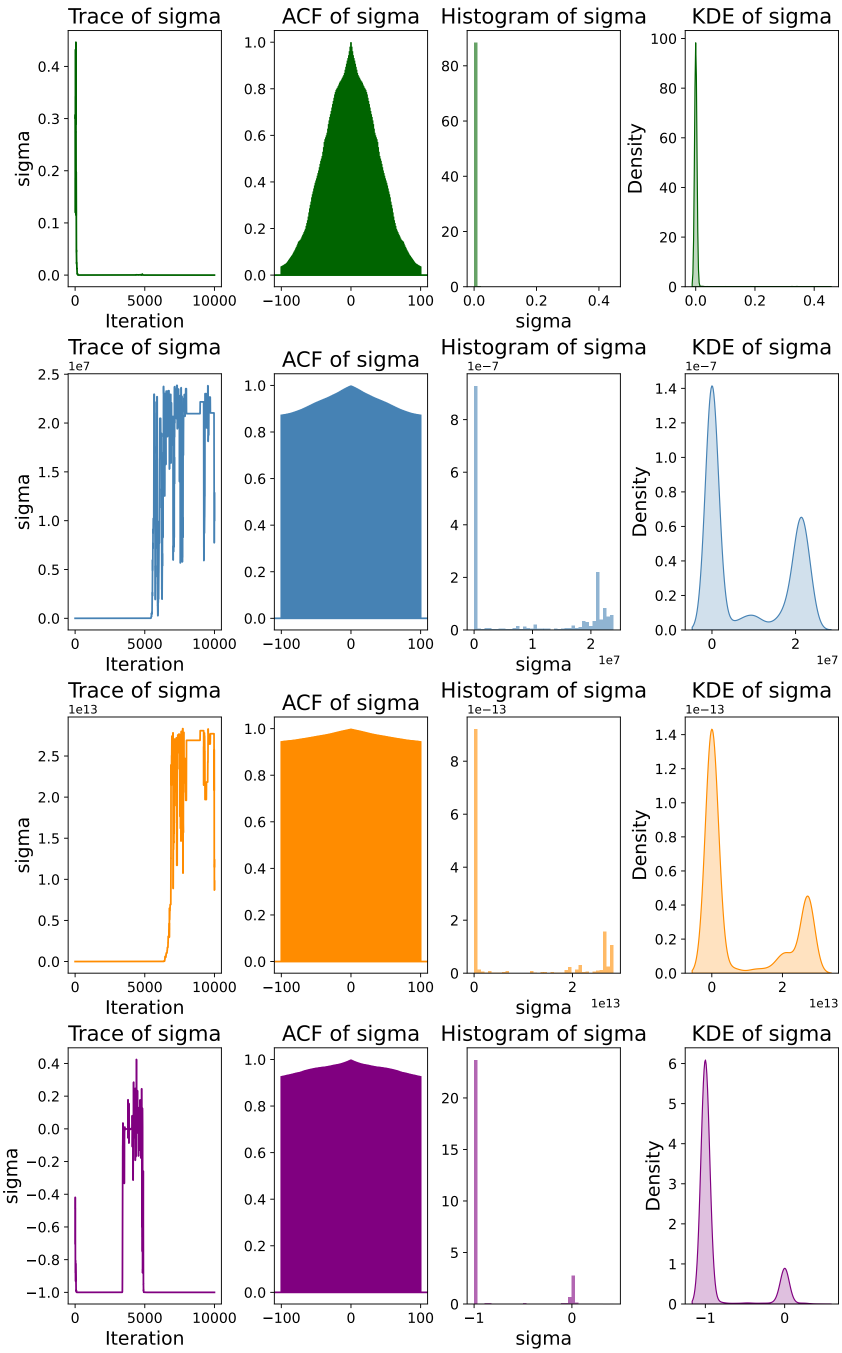

Figure 2: NEED TO WRITE CAPTION... Also Need to redo labels and axes - Too small

3.3 Adaptive Metropolis Hastings

A standard random walk metropolis algorithm uses a fixed proposal covariance, C. If C is too small, the chain mixes slowly (high autocorrelation). On the other hand, if too large, proposals are often rejected. In high or moderate dimensions, optimal choices of C depend on the correlation structure of the target posterior.

We want to let the chain learn an approximately optimal proposal shape on each iteration by estimating the posterior's local covariance structure from past samples. The ideal proposal for a Gaussian target $\mathcal{N}(\mu, \Sigma)$ is $\mathcal{N}(\theta, \Sigma)$. *Adaptive-Metropolis Hastings AMH* seeks to mimic this by continually updating an empirical covariance, allowing the proposal distribution to evolve in response to information gathered from past samples, thereby automatically tuning itself to the shape of the target posterior.

Say we had a two risky assets S_1, S_2 under the Black-Scholes, with correlation $\rho \in (-1, 1)$, such that:

$$\begin{aligned} dS_{1,t} &= rS_{1,t}dt + \sigma S_{1,t}dW_{1,t}^{\mathbb{Q}}, \\ dS_{2,t} &= rS_{2,t}dt + \nu S_{2,t}dW_{2,t}^{\mathbb{Q}}, \\ d\langle W_{1,t}^{\mathbb{Q}}, W_{2,t}^{\mathbb{Q}}\rangle &= \rho dt, \end{aligned}$$

where $\sigma, \nu > 0$, and $W_{i,t}^{\mathbb{Q}}$ are the Wiener processes in the risk neutral measure for each risky asset $i = 1, 2$. If we let $\boldsymbol{\theta} \in \mathbb{R}^3$ denote the current state of the Markov chain, and let $\pi(\boldsymbol{\theta}|\boldsymbol{X})$ be the three dimensional posterior over (σ, ν, ρ) for the two-asset Black-Scholes model, reparameterised via $\theta_1 = \ln \sigma, \theta_2 = \ln \nu, \theta_3 = \operatorname{arctanh}(\rho)$.

At iteration n, given the current state $\boldsymbol{\theta}^{(n)}$ and a symmetric positive-definite proposal covariance $C^{(n)}$, the AMH algorithm generates a candidate $\boldsymbol{\theta}'$ by drawing from the Gaussian proposal distribution $\mathcal{N}(\boldsymbol{\theta}^{(n)}, C^{(n)})$. The acceptance probability for the Metropolis algorithm is then computed as

$$\alpha(\boldsymbol{\theta}^{(n)}, \boldsymbol{\theta}') = \min\left(1, \frac{\pi(\boldsymbol{\theta}'|\boldsymbol{X})q(\boldsymbol{\theta}^{(n)}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(n)}|\boldsymbol{X})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(n)})}\right) = \min\left(1, \frac{\pi(\boldsymbol{\theta}'|\boldsymbol{X})}{\pi(\boldsymbol{\theta}^{(n)}|\boldsymbol{X})}\right),$$

and $\boldsymbol{\theta}^{(n+1)}$ is set equal to $\boldsymbol{\theta}'$ with probability α, or remains at $\boldsymbol{\theta}^{(n)}$ otherwise. Upon accepting or rejecting this proposal, we update the empirical mean $M^{(n)}$ and sum of squares matrix $S^{(n)}$ using the *one-pass algorithms* (Welford, 1962):

$$\begin{aligned} M^{(n+1)} &= M^{(n)} + \frac{\boldsymbol{\theta}^{(n+1)} - M^{(n)}}{n+1} \\ S^{(n+1)} &= S^{(n)} + (\boldsymbol{\theta}^{(n+1)} - M^{(n)})(\boldsymbol{\theta}^{(n+1)} - M^{(n+1)})^\top, \end{aligned}$$

so that the empirical covariance is $\hat{\Sigma}^{(n+1)} = \frac{S^{(n+1)}}{n}\ \forall n \geq 1$. The idea of adaptation consists in blending the previous proposal covariance $C^{(n)}$ with a scaled version of the empirical covariance. Specifically, after a predetermined burn-in period of n_0 iterations, we define the scaled empirical covariance as

$$\Sigma_{\text{emp}}^{(n+1)} = \frac{(2.38)^2}{d}\hat{\Sigma}^{(n+1)} + \epsilon I_d,$$

where the factor $\frac{(2.38)^2}{d}$ is known from the optimal-scaling theory of Gaussian targets to make an asymptotic acceptance rate of approximately 23.8%, and $\epsilon > 0$ is a small ridge term to ensure eigenvalues remain bounded away from zero (Gelman et al., 1996). Hence, the new proposal covariance is then set to

$$C^{(n+1)} = \gamma C^{(n)} + (1-\gamma)\Sigma_{\text{emp}}^{(n+1)},$$

with $\gamma \in (0, 1)$ chosen so that changes to the proposal kernel become smaller as n grows. For $n < n_0$, we let $C^{(n+1)} = C^{(n)}$. The two theoretical properties that guarantee the AMH chain remains ergodic with the correct stationary distribution are the *diminishing adaptation condition* and the *containment condition*. The former condition requires that the total-variation distance between successive proposal kernels tends to zero. In other words, if $K_n(\boldsymbol{\theta}, \cdot)$ is the proposal kernel at step n, then we require

$$\sup_{\boldsymbol{\theta}} ||K_{n+1}(\boldsymbol{\theta}, \cdot) - K_n(\boldsymbol{\theta}, \cdot)||_{\mathrm{TV}} \longrightarrow 0 \quad \text{as } n \to 0.$$

This is ensured by the convex combination with weight γ and by the fact that empirical covariances converge (Haario et al., 2001). The containment condition requires that the family of proposal kernels does not wander into explosive patterns. This is enforced by delaying adaptation until after n_0 iterations and by the ridge term ϵI_d, which keeps all eigenvalues within a controlled range.

Table 3: Summary Statistics

Parameter	Mean	Standard Deviation	Effective Sample Size (ESS)
σ	-1.1344	0.0468	1 134.3722
ν	-1.3671	0.0448	755.0555
ρ	0.6548	0.0630	706.1350

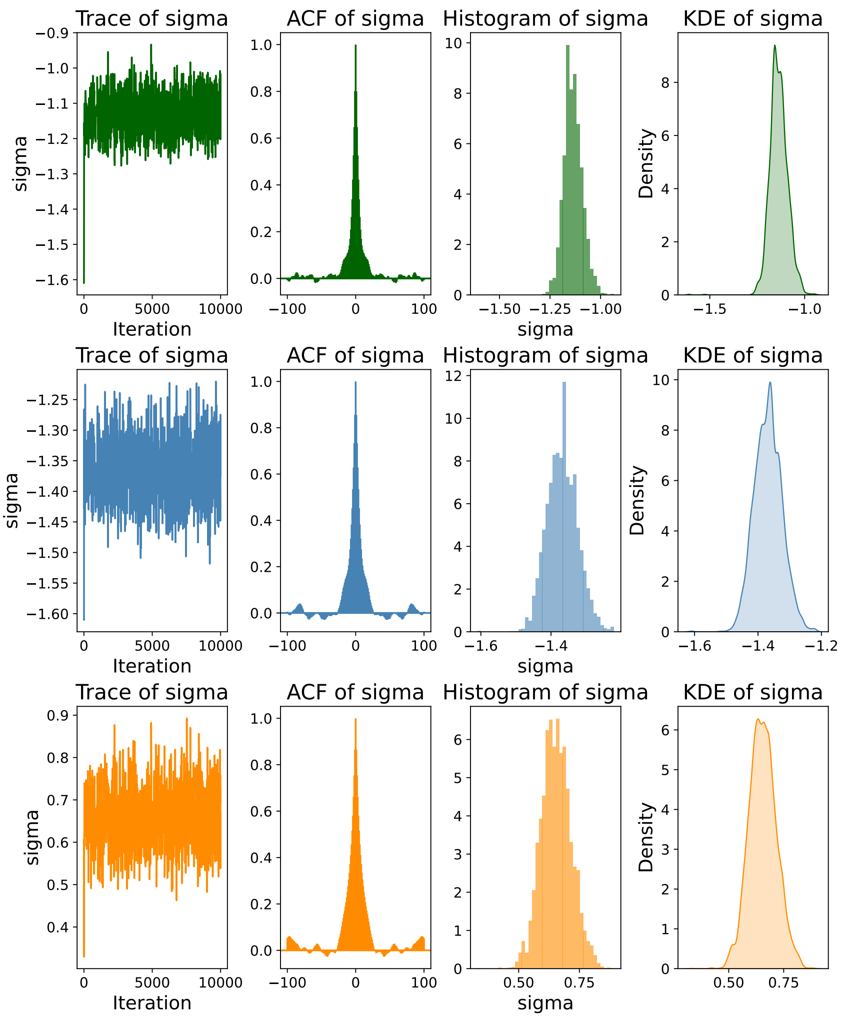

Figure 3: NEED TO WRITE CAPTION... Also Need to redo labels and axes - Too small

Note: acceptance rate 0.2173

3.4 Birth-Death MCMC

In option pricing with jump diffusions, the jump component of asset returns is naturally modeled as a count process such as a Poisson process to take into account the occurrence of discontinuities. A Metropolis-within-Gibbs would treat the integer-valued latent counts with a propose Gaussian random-walk which moves on the count parameter or discrete jumps of arbitrary size. The former violates the integer support and must be ad-hoc rounded, while the latter often leads to vanishing acceptance rates or poor mixing as the chain struggles to traverse the count space in small or overly large increments.

In the Merton jump-diffusion model, where the likelihood couples each latent jump count to both continuous diffusion residuals and a Poisson prior, these issues cause even more of an issue. For example, the improper proposals distort the posterior shape, and mixing stalls in regions of high latent-count probability . To address these problems, we use a *birth-death Metropolis Hastings* step within the Gibbs sampler. By proposing only local birth or death moves, the *Birth-Death Markov Chain Monte Carlo (BDMCMC)* maintains detailed balance with the exact target posterior and achieves higher acceptance rates and more efficient exploration of the latent count space than ad-hoc discrete updates in a standard Metropolis within Gibbs framework (Green, 1995).

3.4.1 Birth-Death Process

A birth-death process is a continuous-time Markov chain $\{N(t)\}_{t\geq 0}$ taking values in the nonnegative integers $\mathbb{N}_0$, in which transitions may only increase the state by one (birth) or decrease by one (death). When the process is in state n, births occur at rate λ_n and deaths at rate μ_n, with $\mu_0 = 0$ by convention so that the chain cannot exit $\mathbb{N}_0$ downwards (think of the process as if it is a population, the population can never be negative only non-negative).Let $P_{mn}(t) = \Pr\{N(t) = n|N(0) = m\}$, where $N(t)$ is the birth-death process, then the forward-Kolmogorov equations are:

$$\frac{d}{dt}P_{mn}(t) = \lambda_{n-1}P_{m,n-1}(t) + \mu_{n+1}P_{m,n+1}(t) - (\lambda_n + \mu_n)P_{mn}(t),$$

with $\lambda_{-1} = 0$ and $\mu_0 = 0$. Equivalently, the infinitesimal generator $\mathcal{Q} = (q_{ij}$ has nonzero entries

$$q_{n,n+1} = \lambda_n, \quad q_{n,n-1} = \mu_n, \quad q_{nn} = -(\lambda_n + \mu_n),$$

and $q_{ij} = 0$ for $|i - j| > 1$. Since the total exit rate from state n is $\alpha_n = \lambda_n + \mu_n$, the waiting time in n is $\mathrm{Exp}(\alpha_n)$. Hence, the chain jumps to $n+1$ with probability $\frac{\lambda_n}{\alpha_n}$ and to $n-1$ with probability $\frac{\mu_n}{\alpha_n}$. If the chain is positive recurrent, it admits a unique stationary distribution $\{\pi\}$ satisfying the detailed balance conditions:

$$\pi_n\lambda_n = \pi_{n+1}\mu_{n+1}, \quad \sum_{n\geq 0}\pi_n = 1.$$

Which can be solved recursively giving a solution for π_n:

$$\pi_n = \pi_0\prod_{k=0}^{n-1}\frac{\lambda_k}{\mu_{k+1}}, \pi_0 = \left(1 + \sum_{n\geq 1}\prod_{k=0}^{n-1}\frac{\lambda_k}{\mu_{k+1}}\right)^{-1}.$$

3.4.2 Birth-Death MCMC

Similarly to section 3.2, we want to sample an unobserved count $N \in \mathbb{N}_0$ from its posterior distribution $\pi(n) \propto \ell(n)p(n)$, where $\ell(n)$ is the likelihood of the data given $N = n$, and $p(n)$ is the prior probability of $N = n$. To incorporate a birth-death Metropolis Hastings step within a larger Gibbs sampler, we suppose that the current latent state is $N^{(i^-)} = n$ at iteration i. Then, we define the proposal distribution for birth and death moves:

$$b_n = \Pr\{n \to n+1\}, \quad d_n = \Pr\{n \to n - `1\},$$

such that $b_n + d_n = 1$, $d_0 = 0$, and $b_n, d_n > 0$ for $n \geq 1$.[6] We then draw $U \sim \text{Uniform}(0,1)$. If $U < b_n$, we set the candidate $n^* = n+1$; otherwise $n^* = \max(n-1, 0)$. To enforce detailed balance with respect to $\pi(\cdot)$, we compute the following acceptance probability:

$$\alpha = \frac{\pi(n^*)q(n^* \to n)}{\pi(n)q(n \to n^*)},$$

where $q(n \to m)$ is the proposal probability of moving from state n to state m (Robert et al., 1999). It is important to note that the birth-death Metropolis-Hastings chain is irreducible whenever $b_n > 0\ \forall n$ and $d_n > 0\ \forall n$, and aperiodic since there is always a positive probability of rejecting a move and remaining at n. Hence, the chain converges in distribution to the target posterior $\pi(n)$.

3.4.3 Applying Birth-Death MCMC to Merton Jump Diffusion Model

Let S_t follow the Merton risk-neutral model:

$$\frac{dS_t}{S_{t^-}} = (r - \lambda k)dt + \sigma dW_t + (e^Y - 1)dN_t,$$

where N_t is a Poisson process with rate λ and jump sizes $Y \sim \mathcal{N}(\nu, \delta^2)$ meaning that $k = \mathbb{E}[e^Y - 1]$. Over a discretisation into intervals of length Δ, let $N = N_T$ be the total count.

We want to sample from the joint posterior of parameters and latent count N, but we will focus here on updating N conditional on all else. The full conditional is $\pi(n) \propto \ell(n)p(n)$, where $\ell(N)$ is the likelihood of the observed price increments $\{S_{t_{i+1}} - S_{t_i}\}|N$, and $p(N) = \mathbb{P}(N = n) = e^{-\lambda T}\frac{(\lambda T)^n}{n!}$. At iteration i, suppose the current latent count is $N^{(i)} = n$. We draw $U \sim \text{Uniform}(0,1)$ and set the proposed count

$$n^* = \begin{cases} n+1, & U < b_n, \\ n-1, & U \geq b_n, \end{cases}$$

where $b_n \in (0,1)$ is our chosen birth and $d_n = 1 - b_n$ the death probability, with $d_0 = 0$ so we never propose negative counts. Then to preserve detailed balance with respect to $\pi(n)$, we accept the move $n \to n^*$ with probability

$$\alpha = \min\left(1, \frac{\pi(n^*)q(n^* \to n)}{\pi(n)q(n \to n^*)}\right),$$

[6] We will be defining $b_n = d_n = \frac{1}{2}$ for $n \geq 1$, $b_0 = 1$.

such that the proposal kernel satisfies

$$q(n \to n+1) = b_n, \quad q(n \to n-1) = d_n,$$

and likewise for $q(n^* \to n)$. Since

$$\frac{p(n+1)}{p(n)} = \frac{e^{-\lambda T}(\lambda T)^{n+1}/(n+1)!}{e^{-\lambda T}(\lambda T)^n/n!} = \frac{\lambda T}{n+1}, \quad \frac{p(n-1)}{p(n)} = \frac{n}{\lambda T},$$

the prior ratio is explicit. Furthermore, under the data model, conditional on $N = n$ the log-price increment over $[0, T]$ is

$$\ln S_T - \ln S_0 = \left(r - \lambda k - \frac{1}{2}\sigma^2\right) T + \sigma W_T + \sum_{i=1}^{n} Y_i,$$

so that the continuous part and the sum of n independently identically distributed normals combine into a single normal. Formally, for two neighbouring counts, $\ell(n) \sim \mathcal{N}(\mu_0 + n\nu, \sigma^2 T + n\delta^2)$, where $\mu_0 := (r - \lambda k - \frac{1}{2}\sigma^2)T$. Thus

$$\frac{\ell(n+1)}{\ell(n)} = \frac{\varphi(x; \mu_0 + (n+1)\nu, \sigma^2 T + (n+1)\delta^2)}{\varphi(x; \mu_0 + n\nu, \sigma^2 T + n\delta^2)},$$

where $\varphi(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{(x-\mu)^2}{2\sigma})$. Hence the birth move $n \to n+1$ is

$$\alpha_b = \min\left(1, \frac{\ell(n+1)}{\ell(n)} \frac{\lambda T}{n+1} \frac{d_{n+1}}{b_n}\right),$$

and the death move $n \to n-1$ is

$$\alpha_d = \min\left(1, \frac{\ell(n-1)}{\ell(n)} \frac{n}{\lambda T} \frac{b_{n-1}}{d_n}\right).$$

Table 4: Summary Statistics

Parameter	Mean	Standard Deviation	Effective Sample Size (ESS)
σ	0.0017	0.0214	5.408911
λ	0.2582	0.2949	4.111688
ν	0.3422	0.1463	2.679296
δ	0.6329	0.2877	3.398248

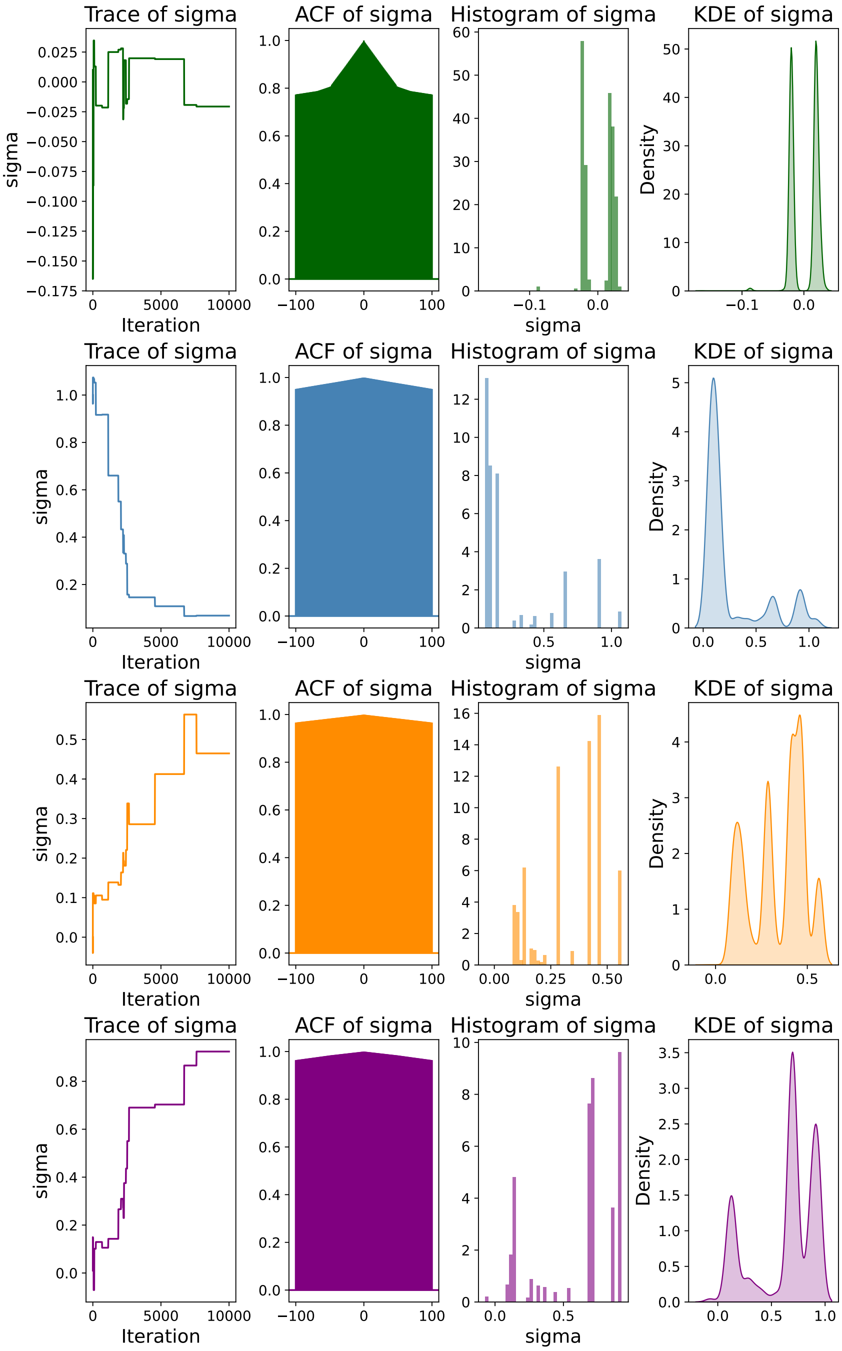

Figure 4: NEED TO WRITE CAPTION... Also Need to redo labels and axes - Too small

3.5 Hamiltonian Monte Carlo

Up to this point, we have struggled to get good results for complex pricing models such as Heston and Merton jump diffusion models. In many Bayesian inference problems, standard random-walk Metropolis or Gibbs samplers suffer from poor scaling in high-dimensional parameter spaces or when strong posterior correlation exist. *Hamiltonian Monte Carlo (HMC)* fixes these limitations by introducing *auxiliary momentum* variables and simulating Hamiltonian dynamics to propose moves that preserve detail balance. The main idea is to view sampling from a target density $\pi(\boldsymbol{\theta})$ as exploring the level sets of an energy function, and to exploit gradient information to traverse parameter space efficiently (Duane et al., 1987).

Formally, let $\boldsymbol{\theta} \in \mathbb{R}^D$ denote the vector of parameters whose posterior density is $\pi(\boldsymbol{\theta}) \propto e^{-U(\boldsymbol{\theta})}$, where

$$U(\boldsymbol{\theta}) = -\log[\pi(\boldsymbol{\theta})]$$

is called the potential energy. We then introduce an auxiliary momentum variable $p \in \mathbb{R}^D$with a conditional Gaussian density $p \sim \mathcal{N}(0, M)$, where M is a positive-define *mass matrix*. The joint density over (θ, p) is then

$$\Pi(\boldsymbol{\theta}, p) = \pi(\boldsymbol{\theta})\mathcal{N}(p|0, M) \propto \exp(-H(\boldsymbol{\theta}, p)),$$

with Hamiltonian

$$H(\boldsymbol{\theta}, p) = U(\boldsymbol{\theta}) + K(p), \quad K(p) = \frac{1}{2}p^\top M^{-1}p,$$

the kinetic energy.

We can describe how $(\boldsymbol{\theta}, p)$ evolve in continuous time t through the *Hamilton's equations of motion*:

$$\frac{d\boldsymbol{\theta}}{dt} = \frac{\partial H}{\partial p} = M^{-1}p, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = -\nabla U(\boldsymbol{\theta}). \tag{3}$$

These deterministic dynamics preserve the joint density $\Pi(\boldsymbol{\theta}, p)$, i.e, they conserve H, and are volume-preserving in phase space. Since the exact integration of (3) is generally impossible, we apply a symplectic integrator, usually *leapfrog scheme*, to approximate the flow with small step size ϵ and repeating the following three steps L times to yield a proposal $(\boldsymbol{\theta}^*, p^*)$:

$$\begin{aligned}
p\left(t + \frac{\epsilon}{2}\right) &= p(t) - \frac{\epsilon}{2}\nabla U(\boldsymbol{\theta}(t)) \qquad \text{(Half-momentum update)},\\
\boldsymbol{\theta}(t+\epsilon) &= \boldsymbol{\theta}(t) + \epsilon M^{-1} p\left(t + \frac{\epsilon}{2}\right) \qquad \text{(Position update)},\\
p(t+\epsilon) &= p\left(t + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2}\nabla U(\boldsymbol{\theta}(t+\epsilon)) \qquad \text{Half-momentum update)}.
\end{aligned}$$

As numerical integration incurs discretisation error, the Hamiltonian is no longer exactly conversed, and to correct for this and to ensure detailed balance, we perform a Metropolis acceptance step to accept $(\boldsymbol{\theta}^*, p^*)$ with probability

$$\min\{1, \exp[-H(\boldsymbol{\theta}^*, p^*) + H(\boldsymbol{\theta}, p)]\}$$

Finally, new momentum is resampled from $\mathcal{N}(0, M)$ before commencing the next trajectory. HMC can propose moves that travel in parameter space with high probability by using gradient information, reducing the random-walk behaviour that simpler MCMC methods use.

3.5.1 Applying HMC to the Heston Model

Recall that the Heston model under the risk-neutral measure is

$$\begin{aligned}
dS_t &= \mu S_t dt + S_t\sqrt{v_t}dW_{t,1}^{\mathbb{Q}},\\
dv_t &= \kappa(\xi - v_t)dt + \sigma\sqrt{v_t}dW_{t,2}^{\mathbb{Q}},\\
d\langle W_{t,1}^{\mathbb{Q}}, W_{t,2}^{\mathbb{Q}}\rangle &= \rho dt.
\end{aligned}$$

For simplicity, we will write $W_{t,i}^{\mathbb{Q}} := W_{t,i}$ for $i = 1, 2$. To apply HMC to Heston, we require defining a suitable potential energy function and computing its gradient with respect to the model parameters $\boldsymbol{\theta} = (\kappa, \xi, \sigma, \rho)$ and the latent volatility path $\{v_t\}$. Given discrete observations $S_{t_0}, S_{t_1}, ..., S_{t_N}$, we can write the joint posterior as

$$\pi(\boldsymbol{\theta}, \{v_i\}|\{S_i\}) \propto \pi(\boldsymbol{\theta})\pi(v_0) \prod_{i=1}^{N} \ell(S_i|S_{i-1}, v_{i-1}, \boldsymbol{\theta})\ell(v_i|v_{i-1}, \boldsymbol{\theta}).$$

Similarly to section 3.2, we can rewrite the problem under Euler discretisation as:

$$\mathcal{L}(S_i|S_{i-1}, v_{i-1}, \boldsymbol{\theta}) := \log \ell(S_i|S_{i-1}, v_{i-1}, \boldsymbol{\theta}) = -\frac{1}{2}\left[\log(2\pi v_{i-1} S_{i-1}^2 \Delta t) + \frac{(\Delta \log S_i - (r - \frac{1}{2} v_{i-1} \Delta t)^2}{v_{i-1}\Delta t}\right],$$
$$\mathcal{L}(v_i|v_{i-1}, \boldsymbol{\theta}) := \log \ell(v_i|v_{i-1}, \boldsymbol{\theta}) = -\frac{1}{2}\left[\log(2\pi\sigma^2 v_{i-1}\Delta t) + \frac{(\Delta v_i - \kappa(\xi - v_{i-1})\Delta t)^2}{\sigma^2 v_{i-1}\Delta t}\right],$$

where $\Delta \log S_i = \log(\frac{S_i}{S_{i-1}})$ and $\Delta v_i = v_i - v_{i-1}$. Thus the potential energy is

$$U(\boldsymbol{\theta}.\{v_i\}) = -\log \pi(\boldsymbol{\theta}) - \log \pi(v_0) - \sum_{i=1}^{N} \left[\mathcal{L}(S_i|S_{i-1}, v_{i-1}, \boldsymbol{\theta}) + \mathcal{L}(v_i|v_{i-1}, \boldsymbol{\theta})\right].$$

Gradient components are then obtained by differentiating U with respect to each parameter and each latent v_i. For example,

$$\frac{\partial U}{\partial v_{i-1}} = -\frac{\partial}{\partial v_{i-1}}\mathcal{L}(S_i|S_{i-1}, v_{i-1}, \boldsymbol{\theta}) - \frac{\partial}{\partial v_{i-1}}\mathcal{L}(v_i|v_{i-1}, \boldsymbol{\theta}) - \frac{\partial}{\partial v_{i-1}}\mathcal{L}(v_{i-1}|v_{i-2}, \boldsymbol{\theta}),$$

each term makes closed form expressions involving polynomials and inverse functions of v_{i-1}. We then assemble these gradients into the vector $\nabla_{\boldsymbol{\theta},v} U$ and implement the leapfrog steps in the joint space of $\boldsymbol{\theta}$ and $\{v_i\}$, tuning ϵ, M, and L.

4 Exotic Options

4.1 Asian options

Asian options are a type of single asset exotic option where the payoff function is defined in terms of the average price of the underlying, instead of the final price. We will be considering Geometric Asian options in this paper. The payoff function of this option is:

$$\Phi = \max(A_T - K, 0)$$

$$where\ A_T = \exp\Big\{\frac{1}{T}\int_0^T \log(S_t)\, dt\Big\}$$

4.2 Margrabe's Option & Rainbow Options

4.3 Pricing Under Black-Scholes

4.3.1 Asian

Recall from section TODO the dynamics of the log prices in a Black-Scholes pricing model under risk neutral measure $\mathbb{Q}$:

$$dX_t = (r - \frac{1}{2}\sigma^2)dt\ + \sigma dW_t$$

Integrating both sides between 0 and t, we obtain

$$\int_0^t dX_t = \int_0^t (r - \frac{1}{2}\sigma^2)dt + \sigma \int_0^t dW_t$$

$$X_t - X_0 = (r - \frac{1}{2}\sigma^2)t + \sigma dW_t$$

$$\log(S_t) = \log(S_0) + (r - \frac{1}{2}\sigma^2)t + \sigma dW_t$$

Now, substituting this expression for $\log(S_t)$ into our expression for A_T:

$$A_T = \exp\Big\{\frac{1}{T}\int_0^T \log(S_t)\, dt\Big\} = \exp\Big\{\frac{1}{T}\int_0^T \log(S_0) + (r - \frac{1}{2}\sigma^2)t + \sigma W_t\, dt\Big\}$$

Considering only the exponentiated term:

$$let\ Y = \frac{1}{T}\int_0^T \log(S_0) + (r - \frac{1}{2}\sigma^2)t + \sigma W_t\, dt$$

$$Y = \log(S_0) + \frac{T}{2}(r - \frac{1}{2}\sigma^2) + \frac{\sigma}{T}\int_0^T W_t dt$$

As W_t is a stochastic process, this integral will not have an exact solution. However, due to properties of Brownian motion, we know this integral is normally distributed with mean 0. So consider the variance of this distribution:

$$\mathbb{V}ar\Big[\int_0^T W_t dt\Big] = \mathbb{E}\Big[\Big(\int_0^T W_t dt\Big)^2\Big]$$

$$\mathbb{V}ar\Big[\int_0^T W_t dt\Big] = \mathbb{E}\Big[\Big(\int_0^T W_t dt\Big)\Big(\int_0^T W_t dt\Big)\Big] = \mathbb{E}\Big[\Big(\int_0^T W_t dt\Big)\Big(\int_0^T W_s ds\Big)\Big]$$

Now rewriting this integral as a double integral:

$$\mathbb{V}ar\Big[\int_0^T W_t dt\Big] = \mathbb{E}\Big[\int_0^T \int_0^T W_s W_t ds dt\Big]$$

$$= \int_0^T \int_0^T \mathbb{E}[W_s W_t] ds dt = \int_0^T \int_0^T \min(t, s) ds dt$$

Now, dividing this $[0, T] \times [0, T]$ region into 2 parts, one with $s \leq t$, and the other with $s > t$

$$\mathbb{V}ar\Big[\int_0^T W_t dt\Big] = \int\int_{s \leq t} s \, ds dt \, + \int\int_{s>t} t \, ds dt$$

Now substituting in bounds for these regions:

$$\mathbb{V}ar\Big[\int_0^T W_t dt\Big] = \int_0^T \int_0^t s \, ds dt \, + \int_0^T \int_0^s t \, ds dt = 2\int_0^T \int_0^t s \, ds dt = 2 \cdot \frac{T^3}{6} = \frac{T^3}{3}$$

Thus $\mathbb{V}ar\Big[\frac{\sigma}{T}\int_0^T W_t dt\Big] = \frac{\sigma^2}{T^2} \cdot \frac{T^3}{3} = \frac{\sigma^2 T}{3}$.

Therefore, we can see that $\int_0^T W_t \sim N(0, \frac{T^3}{3})$, so using properties of the normal distribution:

$$Y \sim N(\hat{\mu}, \nu^2),$$

Where:

- $\hat{\mu} = \log(S_0) + \frac{T}{2}(r - \frac{1}{2}\sigma^2)$
- $\nu = \sigma\sqrt{\frac{T}{3}}$

Now we can rewrite the expression for A_T as:

$$A_T = e^Y$$

From this we can see that A_T is log-normally distributed, so the price of the option will take a similar form to that of Black-Scholes for a European call option.
Defining the standard normal variable as Z:

$$Z = \frac{Y - \hat{\mu}}{\nu} \implies Y = \nu Z + \hat{\mu}$$

$$\mathbb{Q}(A_T > K) = \mathbb{Q}(\log(A_T) > \log(K))$$

$$= \mathbb{Q}(\nu Z + \hat{\mu} > \log(K)) = \mathbb{Q}\Big(Z > \frac{\log(K) - \hat{\mu}}{\nu}\Big)$$

$$\mathbb{Q}(A_t > K) = 1 - \mathbb{Q}\Big(Z < \frac{\log(K) - \hat{\mu}}{\nu}\Big) = 1 - \Phi\Big(-\frac{\log(K) - \hat{\mu}}{\nu}\Big)$$

As d_2 is the risk neutral probability that the option finishes in the money:

$$\mathbb{Q}(A_T > K) = \mathbb{Q}(Z > -d_2) = 1 - \mathbb{Q}(Z < d_2)$$

From this it can be seen that:

$$d_2 = \frac{\hat{\mu} - \log(K)}{\nu}$$

Just as with European call options, $d_2 = d_1 - \nu \implies d_1 = d_2 + \nu$ So:

$$d_1 = \nu + d_2 = \frac{\hat{\mu} + \nu^2 - \log(K)}{\nu}$$

Thus we have all of the expressions required for the final option price.

$$C = e^{-rT}(S_0 e^{\hat{\mu}+\frac{1}{2}\nu^2}\Phi(d_1) - K\Phi(d_2))$$

Where:

- $\hat{\mu} = \log(S_0) + \frac{T}{2}(r - \frac{1}{2}\sigma^2)$
- $\nu = \frac{\sigma^2 T}{3}$
- $d_1 = \frac{\hat{\mu}+\nu^2-\log(K)}{\nu}$
- $d_2 = \frac{\hat{\mu}-\log(K)}{\nu}$

4.3.2 Margrabe's

Margrabe's two-asset exchange option is an option which one holds the right to exchange one unit of asset 2 for one unit of asset 1 at time T. Formally speaking, the payoff function is

$$\Phi(S_{1,T}, S_{2,T}, T) = \max(S_{1,T} - S_{2,T}), 0). \tag{4}$$

To price this claim we introduce the log-ratio process

$$X_t = \ln S_{1,t} - \ln S_{2,t}. \tag{5}$$

Recall that under risk-neutral measure $\mathbb{Q}$, each asset price $S_{i,t}$ in the Black-Scholes framework satisfies

$$d \ln S_{i,t} = \left(r - \frac{1}{2}\sigma_i^2\right) dt + \sigma_i dW_{i,t},$$

for $i = 1, 2$, where r is the risk-free rate, σ_i, the volatility of asset i, and the Brownian motions satisfy $d\langle W_{t,1}, W_{t,2}\rangle = \rho dt$. By applying Itô's lemma to (5) and linearity of stochastic integrals,

$$dX_t = d\ln S_{1,t} - d\ln S_{2.t} = \left[\left(r - \frac{1}{2}\sigma_1^2\right) - \left(r - \frac{1}{2}\sigma_2^2\right)\right] dt + \sigma_1 dW_{1,t} - \sigma_2 dW_{2,t}$$

. Since $d\langle W_{t,1}, W_{t,2}\rangle = \rho dt$, we cam define a single Brownian motion W_t and an effective volatility

$$\sigma_{\text{eff}} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2},$$

so that

$$\sigma_1 dW_{1,t} - \sigma_2 dW_{2,t} = \sigma_{\text{eff}} dW_t.$$

Hence, we can write dX_t as

$$dX_t = -\frac{1}{2}(\sigma_1^2 - \sigma_2^2)dt + \sigma_{\textbf{eff}} dW_t.$$

Integrating both sides from 0 to T gives

$$X_T - X_0 = -\frac{1}{2}(\sigma_1^2 - \sigma_2^2)T + \sigma_{\text{eff}} \int_0^T dW_t$$

$$\Rightarrow X_T = X_0 - \frac{1}{2}(\sigma_1^2 - \sigma_2^2)T + \sigma_{\text{eff}} W_T,$$

and therefore X_T is normally distributed with mean

$$\mu_X = X_0 - \frac{1}{2}(\sigma_1^2 - \sigma_2^2)T,$$

and variance

$$\mathbb{V}\text{ar}[X_T] = \sigma_{\text{eff}}^2 T.$$

The exchange-option payoff can be written as

$$(S_{1,T} - S_{2,T})^+ = S_{2,T}(\exp(X_T) - 1)^+.$$

Its time-zero fair price is the discounted risk-neutral expectation:

$$C = e^{-rT}\, \mathbb{E}_{\mathbb{Q}}\left[S_{2,T}(\exp(X_T) - 1)^+\right].$$

Since under $\mathbb{Q}^{(2)}$ the drift of X_t becomes $-\frac{1}{2}\sigma_{\text{eff}}^2$, the law od X_T under this measure is

$$X_T \sim \mathcal{N}\left(X_0 - \frac{1}{2}\sigma_{\text{eff}}^2 T, \sigma_{\text{eff}}^2 T\right).$$

Hence, if we let $F_t = \frac{S_{1,t}}{S_{2,t}} = e^{X_t}$, then F_t is log-normal, and we can define

$$d_1 = \frac{X_0 + \frac{1}{2}\sigma_{\text{eff}}^2 T}{\sigma_{\text{eff}}\sqrt{T}}, \quad d_2 = d_1 - \sigma_{\text{eff}}\sqrt{T},$$

where $X_0 = \ln \frac{S_{1,0}}{S_{2,0}}$. The expectaion of $(F_T - 1)^+$ under the log-normal law is then

$$\mathbb{E}_{\mathbb{Q}^{(2)}}\left[(F_T - 1)^+\right] = N(d_1) - N(d_2),$$

where N is the standard normal cumulative distribution. Hence the Margrabe formula in closed form is

$$C = S_{1,0}N(d_1) - S_{2,0}N(d_2). \tag{6}$$

4.4 Pricing Under Merton Jump Diffusion

4.4.1 Asian

To begin, we denote N_t as a Poisson process with rate λ, and by $\{Y_i\}_{i\geq 1}$ as an independent identically distributed sequence of jump sizes with

$$Y_i \sim \mathcal{N}(\nu, \delta^2),$$

and define the compensator $k = \mathbb{E}[e^Y - 1] = e^{\nu + \frac{1}{2}\delta^2} - 1$. Then the risk-neutral SDE is

$$\frac{dS_t}{S_{t^-}} = (r - \lambda k)dt + \sigma dW_t + (e^Y - 1)dN_t.$$

By using Itô's lemma, the log-price $X_t = \ln S_t$ satisfies

$$dX_t = (r - \lambda k - \frac{1}{2}\sigma^2)dt + \sigma dW_t + Y dN_t.$$

We now integrating both sides of the log price SDE from 0 to t. The jump term gives,

$$\int_0^t Y dN_s = \sum_{i=1}^{N_t} Y_i,$$

since each jump at time $\tau_i \leq t$ contributes Y_i. Hence the integration of the log price becomes

$$X_t = X_0 + (r - \lambda k - \frac{1}{2}\sigma^2)t + \sigma W_t + \sum_{i=1}^{N_t} Y_i. \tag{7}$$

Since we are interested in the geometric Asain payoff

$$A_T = \exp\left\{\frac{1}{T}\int_0^T X_s ds\right\},$$

we will set

$$Y_T := \frac{1}{T}\int_0^T X_s ds, \tag{8}$$

so that $A_T = e^{Y_T}$. By substituting (7) into Y_t, we get

$$\begin{aligned}
Y_T &= \frac{1}{T}\int_0^T \left[X_0 + (r - \lambda k - \frac{1}{2}\sigma^2)s + \sigma W_s + \sum_{i=1}^{N_s} Y_i\right] ds \\
&= \frac{1}{T}\int_0^T X_0 ds + \frac{1}{T}\int_0^T (r - \lambda k - \frac{1}{2}\sigma^2)s ds + \frac{1}{T}\int_0^T \sigma W_s ds + \frac{1}{T}\int_0^T \sum_{i=1}^{N_s} Y_i ds \\
&= X_0 + \left(r - \lambda k - \frac{1}{2}\sigma^2\right)\frac{T}{2} + \frac{1}{T}\int_0^T \sigma W_s ds + \frac{1}{T}\int_0^T \sum_{i=1}^{N_s} Y_i ds.
\end{aligned}$$

To handle $\frac{\sigma}{T}\int_0^T W_s ds$, we must find the mean and variance. However, we know $\int_0^T W_s ds$ is Gaussian with mean zero and variance

$$\mathbb{V}\mathrm{ar}\left[\int_0^T W_s ds\right] = \int_0^T \int_0^T \mathbb{E}[W_u W_v] du dv = \int_0^T \int_0^T \min(u, v) du dv = \frac{T^3}{3}.$$

Hence

$$\frac{\sigma}{T}\int_0^T W_s ds \sim \mathcal{N}\left(0, \frac{\sigma^2 T}{3}\right).$$

The jump process term can be solved using Fubini:

$$\frac{1}{T}\int_0^T \sum_{i=1}^{N_s} Y_i ds = \frac{1}{T}\sum_{i=1}^{N_s}\int_0^T Y_i ds = \frac{1}{T}\sum_{i=1}^{N_T} Y_i (T - \tau_i),$$

where $\tau_i \leq T$. Therefore the jump term is

$$J_T := \frac{1}{T}\sum_{i=1}^{N_T} Y_i (T - \tau_i).$$

Given $N_T = n$. the jump times $\{\tau_i\}$ are independent identically distributed uniform on $[0, T]$. Hence

$$\mathbb{E}[T - \tau_i | N_T = n] = \frac{T}{2},$$

and $\mathbb{E}[Y_i] = \nu$. Thus

$$\mathbb{E}[J_t | N_T = n] = \frac{1}{T} n\nu \frac{T}{2} = \frac{n\nu}{2}.$$

By averaging over $N_T \sim \text{Pois}(\lambda T)$, we get

$$\mathbb{E}[J_T] = \sum_{n=0}^{\infty} \frac{n\nu}{2} e^{-\lambda T} \frac{(\lambda T)^n}{n!} = \frac{\nu}{2}\,\mathbb{E}[N_T] = \frac{\nu}{2}\lambda T.$$

To find the variance, we first note that conditional on $N_T = n$ and jump times, the variance of the sum is

$$\mathbb{V}\text{ar}\left[\frac{1}{T}\sum_{i=1}^{N_T} Y_i(T - \tau_i)\right] = \frac{1}{T^2}\sum_{i=1}^{n} \mathbb{V}\text{ar}[Y_i(T - \tau_i)],$$

since the $Y_i(T - \tau_i)$ are independent given the τ_i. Now,

$$\mathbb{V}\text{ar}[Y_i(T - \tau_i)] = \mathbb{E}[Y_i^2]\,\mathbb{E}[(T - \tau_i)^2] - (\mathbb{E}[Y_i])^2(\mathbb{E}[T - \tau_i])^2.$$

From generating functions, we have $\mathbb{E}[Y_i^2] = \nu^2 + \delta^2$, $\mathbb{E}[T - \tau_i] = \frac{T}{2}$, and

$$\mathbb{E}[(T - \tau_i)^2] = \int_0^T \frac{(T - u)^2}{T} du = \frac{T^2}{3}.$$

Hence

$$\mathbb{V}\text{ar}[Y_i(T - \tau_i)] = (\nu^2 + \delta^2)\frac{T^2}{3} - \nu^2\frac{T^2}{4} = \frac{T^2}{12}(4\nu^2 + 4\delta^2 - 3\nu^2) = \frac{T^2}{12}(\nu^2 + 4\delta^2).$$

Therefore, conditional on $N_T = n$,

$$\mathbb{V}\text{ar}[J_T | N_t = n] = \frac{n}{T^2}\frac{T^2}{12}(\nu^2 + 4\delta^2) = \frac{n}{12}(\nu^2 + 4\delta^2)$$

By using the property $\mathbb{V}\text{ar}[J_t] = \mathbb{E}[\mathbb{V}\text{ar}(J_t|N_T)] + \mathbb{V}\text{ar}(\mathbb{E}[J_T|N_T])$, and $\mathbb{E}[N_T] = \lambda T$ and $\mathbb{V}\text{ar}[N_T] = \lambda T$,

$$\begin{aligned}
\mathbb{V}\text{ar}[J_T] &= \frac{\lambda T}{12}(\nu^2 + 4\delta^2) + \mathbb{V}\text{ar}\left(\frac{N_t \nu}{2}\right) = \frac{\lambda T}{12}(\nu^2 + 4\delta^2) + \left(\frac{\nu}{2}\right)^2 \lambda T \\
&= \lambda T\left[\frac{\nu^2 + 4\delta^2}{12} + \frac{\nu^2}{4}\right] = \frac{\lambda T}{3}(\nu^2 + \delta^2).
\end{aligned}$$

By plugging back into (8), we can represent Y_T in its Gaussian form with mean

$$\mu_Y = X_0 + \left(r - \lambda k - \frac{1}{2}\sigma^2\right)\frac{T}{2} + \frac{\lambda \nu T}{2},$$

and variance

$$\sigma_Y^2 = \frac{\sigma^2 T}{3} + \frac{\lambda T}{3}(\nu^2 + \delta^2).$$

Since A_T is log normal with these parameters, the time sero price of the geometric Asain call, with strike K, is

$$C = e^{-rT}\,\mathbb{E}[(A_T - K)^+] = e^{-rT}[e^{\mu_Y + \frac{1}{2}\sigma_Y^2} N(d_1) - KN(d_2)],$$

where N is the standard normal cumulative distribution and

$$d_1 = \frac{\mu_Y + \sigma_Y^2 - \ln K}{\sigma_Y}, \quad d_2 = d_1 - \sigma_Y.$$

4.4.2 Margrabe's

Recall from subsection 4.3.2 that $X_t = \ln S_{1,t} - \ln S_{2,t}$, where

$$\frac{dS_{i,t}}{S_{i,t^-}} = (r - \lambda_i k_i)dt + \sigma_i dW_{i,t} + (e^{Y_i} - 1)dN_{i,t},$$

$$d\langle W_1, W_2\rangle = \rho dt, \quad i = 1, 2.$$

We can apply Itô's lemma to get

$$dX_t = [r - \lambda_1 k_1 - \frac{1}{2}\sigma_1^2]dt - [r - \lambda_2 k_2 - \frac{1}{2}\sigma_2^2]dt + \sigma_1 dW_{1,t} - \sigma_2 dW_{2,t} + Y_1 dN_{1,t} - Y_2 dN_{2,t}.$$

We can rewrite drift as

$$\mu_0 := [r - \lambda_1 k_1 - \frac{1}{2}\sigma_1^2] - [r - \lambda_2 k_2 - \frac{1}{2}\sigma_2^2] = -\frac{1}{2}[\sigma_1^2 - \sigma_2^2] - [\lambda_1 k_1 - \lambda_2 k_2],$$

and similarly in section 4.3.2, we can combine the Brownian terms into a single Brownian motion W_t with effective volatility:

$$\sigma_{\text{eff}} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2},$$

hence giving

$$dX_t = \mu_0 dt + \sigma_{\text{eff}} dW_t + Y_1 dN_{1,t} - Y_2 dN_{2,t}.$$

Integrating both sides then gives:

$$\begin{aligned} X_T - X_0 &= \int_0^T \mu_0 dt + \int_0^T \sigma_{\text{eff}} dW_t + \sum_{i=1}^{N_{1,T}} Y_{1,i} - \sum_{j=1}^{N_{2,T}} Y_{2,j}, \\ &= \mu_0 + \sigma_{\text{eff}} W_T + \sum_{i=1}^{N_{1,T}} Y_{1,i} - \sum_{j=1}^{N_{2,T}} Y_{2,j}. \end{aligned}$$

Recall that the Margrabe payoff is $\max(S_{1,T} - S_{2,T}, 0) = S_{2,T}(e^{X_T} - 1)^+$. Under the risk neutral measure $\mathbb{Q}_2$ with S_2 as numéraire, the discount ratio $F_t = \frac{S_{1,t}}{S_{2,t}} = e^{X_t}$ is a martingale.[7] Thus the time-zero price is

$$\Pi = S_{2,0}\,\mathbb{E}_{\mathbb{Q}^{(2)}}\left[\left(e^{X_T} - 1\right)^+\right]. \tag{9}$$

[7] The proof is straight forward since $S_{1,t}$ and $S_{2,t}$ are both martingales.

We can now condition on the realised jump counts $N_{1,T} = n_1$ and $N_{2,T} = n_2$. Since $N_{1,T}, N_{2,T}$ are independent Poisson random variables with parameters $\lambda_1 T$ and $\lambda_2 T$ respectively, we can write

$$\mathbb{E}_{\mathbb{Q}^{(2)}}\left[\left(e^{X_T}-1\right)^+\right] = \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} E_{\mathbb{Q}^{(2)}}\left[\left(e^{X_T}-1\right)^+ | N_{1,T}=n_1, N_{2,T}=n_2\right] \mathbb{P}(N_{1,T}=n_1)\mathbb{P}(N_{2,T}=n_2).$$

For fixed (n_1, n_2), the conditional distribution of

$$X_T = X_0 + \mu_0 T + \sigma_{\text{eff}} W_t + \sum_{i=1}^{n_1} Y_{1,i} - \sum_{j=1}^{n_2} Y_{2,j},$$

will have a normal distribution of:

$$X_T \sim \mathcal{N}(X_0 + \mu_0 T + n_1\nu_1 - n_2\nu_2, \sigma_{\text{eff}}^2 T + n_1\delta_1^2 + n_2\delta_2^2).$$

This is due to the two sums of independent normals are themselves normal with means $n_1\nu_1$ and $n_2\nu_2$, and variance $n_1\delta_1^2$ and $n_2\delta_2^2$ respectively. Under this conditional normal law, the expectation of $(e^{X_T}-1)^+$ is exactly the Black-Scholes-style one-asset call formula applied to the log-normal variable e^{X_T}.Define

$$d_1(n_1,n_2) = \frac{X_0 + \mu_0 T + n_1\nu_1 - n_2\nu_2 + \sigma_{\text{eff}}^2 T + n_1\delta_1^2 + n_2\delta_2^2}{\sqrt{\sigma_{\text{eff}}^2 T + n_1\delta_1^2 + n_2\delta_2^2}},$$

$$d_2(n_1,n_2) = d_1(n_1,n_2) - \sqrt{\sigma_{\text{eff}}^2 T + n_1\delta_1^2 + n_2\delta_2^2}.$$

For simplicity we will define the conditional distribution of X_T as $X_T \sim \mathcal{N}(m, v)$. We can express the expectation of (9) with the density of X_T:

$$\begin{aligned}
\mathbb{E}[(e^{X_T}-1)^+] &= \int_0^\infty (e^x - 1) f(x)dx = \int_0^\infty e^x f(x)dx - \int_0^\infty f(x)dx \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi v}} \exp\left(x - \frac{(x-m)^2}{2v}\right) dx - \mathbb{P}[X_T > 0] \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi v}} \exp\left(x - \frac{(x-m)^2}{2v}\right) dx - \Phi\left(\frac{m}{\sqrt{v}}\right),
\end{aligned}$$

where Φ is the standard normal cumulative density function. By completing the square we can rearrange the exponential term in the first integral as

$$e^{m+\frac{1}{2}v} \int_0^\infty \frac{1}{\sqrt{2\pi v}} \exp\left(\frac{(x-(m+v))^2}{2v}\right) dx,$$

hence, by a change in variable of $z = \frac{x-(m+v)}{\sqrt{v}}$, we get the following:

$$e^{m+\frac{1}{2}v} \int_{-(m+v)/\sqrt{v}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = e^{m+\frac{1}{2}v}\Phi\left(\frac{m+v}{\sqrt{v}}\right) = e^{m+\frac{1}{2}v}\Phi(d_1),$$

where $d_1 = \frac{m+v}{\sqrt{v}}$. If we take $d_2 = \frac{m}{\sqrt{v}}$, then we can write the expectation from (9) as

$$\mathbb{E}[(e^{X_T}-1)^+] = \int_0^\infty (e^x - 1) f(x)dx = e^{m+\frac{1}{2}v}\Phi(d_1) - \Phi(d_2).$$

Hence,

$$\mathbb{E}_{\mathbb{Q}_2}[(e^{X_T}-1)^+] = \Phi(d_1) - \Phi(d_2).$$

and

$$\mathbb{E}_{\mathbb{Q}_2}[(e^{X_T}-1)^+ | n_1, n_2] = \Phi(d_1(n_1,n_2)) - \Phi(d_2(n_1,n_2)).$$

Putting everything together we get:

$$\Pi = S_{2,0} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} [\Phi(d_1(n_1,n_2)) - \Phi(d_2(n_1,n_2))] \frac{e^{-\lambda_1 T}(\lambda_1 T)^{n_1}}{n_1!} \frac{e^{-\lambda_2 T}(\lambda_2 T)^{n_2}}{n_2!}.$$

4.5 Pricing Under Heston Model

4.5.1 Asian

Recall under risk-neutral measure $\mathbb{Q}$, the Heston model is

$$\begin{aligned} dS_t &= \mu S_t dt + S_t \sqrt{v_t} dW_{t,1}^{\mathbb{Q}}, \\ dv_t &= \kappa(\xi - v_t)dt + \sigma\sqrt{v_t} dW_{t,2}^{\mathbb{Q}}, \\ d\langle W_{t,1}^{\mathbb{Q}}, W_{t,2}^{\mathbb{Q}} \rangle &= \rho dt. \end{aligned}$$

Similarly as before using Itô's lemma,

$$dX_t = \left(r - \frac{1}{2}v_t\right) dt + \sqrt{v_t} dW_{t,1}.$$

If the geometric Asian payoff is

$$A_T = \exp\left(\frac{1}{T}\int_0^T X_s ds\right)$$

with the fair call price

$$\Pi = e^{-rT}\,\mathbb{E}_{\mathbb{Q}}[(A_T - K)^+],$$

then to get a closed-form expression, we must first calculate the moment generating function of $\frac{1}{T}\int_0^T X_s ds$, i.e., $M(u) = \mathbb{E}[\exp(\frac{1}{T}\int_0^T X_s ds)]$. If we consider the conditional expectation

$$F(t,x,v) = \mathbb{E}\left[\exp\left(u\frac{1}{T}\int_t^T X_s ds\right) | X_t = x, v_t = v\right],$$

then over the next small interval $[t, t+\Delta t]$, we earn the factor $\exp(\frac{u}{T}\int_t^{t+\Delta t} X_s ds) \approx \exp(\frac{u}{T}x\Delta t)$. Hence we are in the new state $(X_{t+\Delta t}, v_{t+\Delta t})$ from which the remaining expected payoff is $F(t+\Delta t, X_{t+\Delta t}, v_{t+\Delta t})$:

$$F(t,x,v) = \mathbb{E}\left[e^{\frac{u}{T}x\Delta t} F(t+\Delta t, X_{t+\Delta t}, v_{t+\Delta t}) | X_t = x, v_t = v\right]. \tag{10}$$

Additionally, if we let $\mathcal{L}$ be the infinitesimal generator of the 2D diffusion (X_t, v_t), then the generator acting on a twice-differentiable test function $f(x,v)$ is

$$\mathcal{L}f = \left(r - \frac{1}{2}v\right)\partial_x f + \kappa(\xi - v)\partial_v f + \frac{1}{2}v\partial_{xx} f + \rho\sigma v \partial_{xv} f + \frac{1}{2}\sigma^2 v \partial_{vv} f.$$

If we then expand $F(t+\Delta t, X_{t+\Delta t}, v_{t+\Delta t})$ to first order in Δt using Itô's Lemma and the generator $\mathcal{L}$ of the (X_t, v_t) process, we get a time expansion of $F(t+\Delta t, X_{t+\Delta t}, v_{t+\Delta t}) = F(t,x,v) + \partial_t F \Delta t + o(\Delta t)$ and a space expansion via the continuous diffusion part:

$$\mathcal{L}F = \frac{1}{2} v \partial_{xx} F + \rho\sigma v \partial_{xv} F + \frac{1}{2}\sigma^2 v \partial_{vv} F + \left(r - \frac{1}{2}v\right)\partial_x F + \kappa(\xi - v),$$

hence by Itô,

$$\mathbb{E}[F(t+\Delta t, X_{t+\Delta t}, v_{t+\Delta t})] = F(t,x,v) + (\partial_t F + \mathcal{L}F)\Delta t + o(\Delta t). \tag{11}$$

We can expand $e^{(u/T)x\Delta t}$ to get $e^{(u/T)x\Delta t} = 1 + \frac{u}{T}x\Delta t + \wr(\Delta t)$. By substituting this and (11) into (10) we get:

$$\begin{aligned} F(t,x,v) &= \mathbb{E}\left[(1 + \frac{u}{T}x\Delta t)(F + (\partial_t F + \mathcal{L}F)\Delta t)\right] + o(\Delta t) \\ &= F + \left[\partial_t F + \mathcal{L}F + \frac{u}{T}xF\right]\Delta t + o(\Delta t). \end{aligned}$$

$$\implies \partial_t F + \mathcal{L}F + \frac{u}{T}xF = 0.$$

By multiplying by -1 we get the backward form equation:

$$\partial_t F = \mathcal{L}F - \frac{u}{T}xF.$$

Substituting in the generator terms we get

$$-\partial_t F = \frac{1}{2}v\partial_{xx}F + \rho\sigma v \partial_{xv}F + \frac{1}{2}\sigma^2 v \partial_{vv}F + (r - \frac{1}{2}v)\partial_x F + \kappa(\xi - v)\partial_v F - \frac{u}{T}xF,$$

with terminal condition $F(T,x,v) = 1$.[8] We now take the exponential-affine form in the variables x and v as

$$F(t,x,v) = \exp\left(A(\tau) + B(\tau)x + C(\tau)v\right),$$

where $\tau = T - t$, and A, B, C are scalar functions of τ only (Heston, 1993). Computing the derivatives we get:

$$\begin{aligned} &\partial_t F = -\frac{dA}{d\tau}F - \frac{dB}{d\tau}xF - \frac{dC}{d\tau}vF, \\ &\partial_x F = BF, \quad \partial_{xx}F = B^2 F, \\ &\partial_v F = CF, \quad \partial_{vv}F = C^2 F, \quad \partial_{xv}F = BCF. \end{aligned}$$

Substitute into the PDE and cancelling out F:

$$\frac{dA}{d\tau} + \frac{dB}{d\tau}x + \frac{dC}{d\tau}v = \frac{1}{2}vB^2 + \rho\sigma v BC + \frac{1}{2}\sigma^2 v C^2 + \left(r - \frac{1}{2}v\right)B + \kappa(\xi - v)C - \frac{u}{T}x.$$

[8] This is formally known as the *Feynman-Kac PDEs* (Oksendal, 2013), where F satisfies the backward PDE on $[0,T]$

Collecting terms:

$$\begin{aligned}
\frac{dB}{d\tau} &= -\frac{u}{T}, \quad B(0) = 0 \quad (\text{since } F(T,\cdot) = 1), \\
\Longrightarrow \quad B(\tau) &= -\frac{u}{T}\tau, \\
\frac{dC}{d\tau}(\tau) &= \frac{1}{2}(B(\tau))^2 + \rho\sigma B(\tau)C(\tau) + \frac{1}{2}\sigma^2 C(\tau)^2 - \frac{1}{2}B(\tau) - \kappa C(\tau), \quad C(0) = 0, \\
\frac{dA}{d\tau}(\tau) &= rB(\tau) + \kappa\xi C(\tau), \quad A(0) = 0.
\end{aligned}$$

[9]Using the solutions to A, B, C up to $\tau = T$, we have at $t = 0$:

$$M(u) = F(0, X_0, v_0) = \exp(A(T) + B(T)X_0 + C(T)v_0).$$

In particular $M(1) = \mathbb{E}[e^{\frac{1}{T}\int_0^T X_s ds}]$ and $M(0) = 1$. The geometric-Asian call price is then

$$\Pi = e^{-rT}\left[M(1) - KM(0)\right] = e^{-rT}\left[e^{A(T)+B(T)X_0+C(T)v_0} - K\right].$$

4.5.2 Margrabe's

Under the original risk neutral measure $\mathbb{Q}$, we define each asset price $S_{i,t}$ and its variance $v_{i,t}$ as

$$\begin{aligned}
dS_{i,t} &= rS_{i,t}dt + \sqrt{v_{i,t}}S_{i,t}dW_{i,t}, \\
dv_{i,t} &= \kappa_i(\xi_i - v_{i,t})dt + \sigma_i\sqrt{v_{i,t}}dZ_{i,t}, \\
d\langle W_1, W_2\rangle_t &= \rho_S dt, \quad d\langle W_i, Z_i\rangle_t = \rho_i dt,
\end{aligned}$$

and all other covariations zero, and $i = 1, 2$. By using Itô's lemma on $X_t = \ln S_{1,t} - \ln S_{2,t}$,

$$dX_t = \left(r - \frac{1}{2}v_{1,t}\right)dt - \left(r - \frac{1}{2}v_{2,t}\right)dt + \sqrt{v_{1,t}}dW_{1,t} - \sqrt{v_{2,t}}dW_{2,t}.$$

Grouping the variance and combining the two Brownian increments into a single dW_t,

$$\Sigma_t^2 = v_{1,t} + v_{2,t} - 2\rho_s\sqrt{v_{1,t}v_{2,t}},$$

$$\Longrightarrow dX_t = -\frac{1}{2}(v_{1,t} - v_{2,t})dt + \Sigma_t dW_t.$$

We want to switch to the S_2 numéraire measure $\mathbb{Q}_2$, under which the discounted ratio $F_t = e^{X_t}$ must be a martingale. Applying Itô to e^{X_t} and setting its continuous dt term to zero forces $-\frac{1}{2}(v_{1,t} - v_{2,t}) = -\frac{1}{2}\Sigma_t^2$. Thus under $\mathbb{Q}^{(2)}$,

$$dX_t = -\frac{1}{2}\Sigma_t^2 dt + \Sigma_t dW_t.$$

Integrating from 0 to T:

$$X_T - X_0 = -\frac{1}{2}\int_0^T \Sigma_t^2 dt + \int_0^T \Sigma_t dW_t.$$

Let $I := \int_0^T \Sigma_t^2 dt$, then given the entire path $\{\Sigma\}_t$, the stochastic integral $\int_0^T \Sigma_t dW_t$ is normal with mean 0 and variance I, and hence, conditionally $X_T|\{\Sigma\}_t \sim \mathcal{N}(X_0 - \frac{1}{2}I, I)$.

[9]The solution to these PDEs is in the Appendix.

Like previously, if we write the margrabe payoff as $(E^{X_T}-1)^+$, its conditional expectation is exactly the Black-Scholes exchange formula with log-mean $m = X_0 - \frac{1}{2}I$ and varianec $\nu = I$. Hence, similarly to previous subsections:

$$\mathbb{E}\left[\left(e^{X_T}-1\right)\big|\,\Sigma\right] = \Phi\left(\frac{m+\nu}{\sqrt{\nu}}\right) - \Phi\left(\frac{m}{\sqrt{\nu}}\right),$$

where $d_1(I) = \frac{X_0+\frac{1}{2}I}{\sqrt{I}}$ and $d_2(I) = \frac{X_0-\frac{1}{2}I}{\sqrt{I}}$. If we let $g(I) = \Phi\left(\frac{X_0+\frac{1}{2}I}{\sqrt{I}}\right) - \Phi\left(\frac{X_0-\frac{1}{2}I}{\sqrt{I}}\right)$, and I is non-negative random variable, then by the law of the unconscious statistician,

$$\mathbb{E}[g(I)] = \int_0^\infty g(j) f_I(j) dj,$$

where $f_I(i)$ is the unknown density of I. Rather than work with f_I directly, we will use its characteristic function

$$\varphi(z) = \mathbb{E}[e^{izI}] = \int_0^\infty e^{izj} f_I(j) dj$$

which can be written in closed form due to the affine structure of the Heston variance (Duffie et al., 2000). The identity that links expectations of $g(I)$ to $\varphi(z)$ is the *Fourier-inversion formula.* Formally, if we define the Fourier transform of g by

$$\tilde{g}(z) = \int_0^\infty e^{izj} g(j) dj,$$

the Fubini's theorem gives

$$\int_0^\infty g(j) f_I(j) dj = \int_0^\infty g(j)\left[\frac{1}{2\pi}\int_{-\infty}^\infty e^{-izj}\varphi(z)dz\right] dj = \frac{1}{2\pi}\int_{-\infty}^\infty \varphi(z)\tilde{g}(-z)dz.$$

Since $g(j)$ vanishes for $j < 0$ and is real-valued, we can reduce this to an integral over $z \in [0, \infty]$ of the real part:

$$\mathbb{E}[g(I)] = \frac{1}{2\pi}\int_{-\infty}^\infty \varphi(z)\tilde{g}(-z)dz.$$

We can evaluate this integral using the fact that g is real valued, so $\tilde{g}(-z) = \overline{\tilde{g}(-z)}$, and $\varphi(-z) = \overline{\varphi(z)}$, giving us

$$\int_{-\infty}^\infty \varphi(z)\tilde{g}(-z)dz = \int_{-\infty}^0 \varphi(z)\tilde{g}(-z)dz + \int_0^\infty \varphi(z)\tilde{g}(-z)dz.$$

Let $u = -z$, which means that z goes from $-\infty$ to 0, u goes from $+\infty$ to 0 and $dz = -du$. Thus

$$\int_{-\infty}^0 \varphi(z)\tilde{g}(-z)dz = \int_\infty^0 \varphi(-u)\tilde{g}(u)(-du) = \int_0^\infty \varphi(-u)\tilde{g}(u)du.$$

Since g is real-valued, its transform satisfies $\tilde{g}(u) = \overline{\tilde{g}(-u)}$, and the characteristic function of a real-valued random variable satisfies $\varphi(-u) = \overline{\varphi(u)}$, hence $\varphi(-u)\tilde{g}(u) = \overline{\varphi(u)}\;\overline{\tilde{g}(-u)} = \overline{\varphi(u)\tilde{g}(-u)}$. Thus giving:

$$\int_{-\infty}^\infty \varphi(z)\tilde{g}(-z)dz = \int_0^\infty \varphi(z)\tilde{g}(-z)dz + \int_0^\infty \overline{\varphi(z)\tilde{g}(-z)}dz = 2\int_0^\infty \Re[\varphi(z)\tilde{g}(-z)]dz.$$

Hence

$$\mathbb{E}[g(I)] = \frac{1}{2\pi} \cdot 2\Re\left\{\int_0^\infty \varphi(z)\tilde{g}(-z)dz\right\} = \frac{1}{\pi}\int_0^\infty \Re[\varphi(z)\tilde{g}(-z)]dz$$

(Lewis, 2001) showed that this converts into

$$\mathbb{E}[g(I)] = g(\infty) - \frac{1}{\pi}\int_0^\infty \Re[\varphi(z)\hat{g}(-z)]dz,$$

where $g(\infty) = \lim_{i\to\infty} g(i) = 1$, since as $i \to \infty$ the option is almost surely in the money. Additionally, $\hat{g}(z)$ is a slightly modified transform of g, explicitly computable in terms of $e^{\frac{1}{2}iz} - 1$ and the factor $\frac{1}{iz}$ that arises from integrating by parts to handle the payoff kink at $i = 0$.

Finally, this gives us a fair price for the Margrabe option of

$$\Pi = S_{2,0}\left[1 - \frac{1}{\pi}\int_0^\infty \Re\left[\varphi(z)\frac{e^{\frac{1}{2}iz} - 1}{iz}e^{-izX_0}\right]dz\right].$$

5 Results

6 Conclusion

7 Appendix

7.1 Appendix A: Girsanov's Theorem for the Heston Model

7.2 Appendix B: Solution to the Geometric Asian Heston PDEs

Recall we have

$$\begin{aligned}
&\frac{dB}{d\tau} = -\frac{u}{T}, \quad B(0) = 0 \quad (\text{since } F(T, \cdot) = 1), \\
&\implies \quad B(\tau) = -\frac{u}{T}\tau, \\
&\frac{dC}{d\tau}(\tau) = \frac{1}{2}(B(\tau))^2 + \rho\sigma B(\tau)C(\tau) + \frac{1}{2}\sigma^2 C(\tau)^2 - \frac{1}{2}B(\tau) - \kappa C(\tau), \quad C(0) = 0, \\
&\frac{dA}{d\tau}(\tau) = rB(\tau) + \kappa\xi C(\tau), \quad A(0) = 0.
\end{aligned}$$

Since $\frac{dB}{d\tau} = -\frac{u}{T}$ is constant, we have $B(\tau) = B(0) - \frac{u}{T}\tau = -\frac{u}{T}\tau$. Substituting $B(\tau)$,

$$\frac{dC}{d\tau}(\tau) = \frac{1}{2}\frac{u^2}{T^2}\tau^2 + \frac{u}{2T}\tau + \rho\sigma\left(-\frac{u}{T}\tau\right)C(\tau) + \frac{1}{2}\sigma^2 C(\tau)^2 - \kappa C(\tau), \quad C(0) = 0.$$

Let $\gamma(\tau) := \frac{1}{2}\frac{u^2}{T^2}\tau^2 + \frac{u}{2T}\tau$ and $p(\tau) := \rho\sigma\left(-\frac{u}{T}\tau\right) - \kappa$, then we get the Riccati-type ODE

$$\frac{dC}{d\tau} = \frac{1}{2}\sigma^2 C(\tau)^2 + p(\tau)C(\tau) + \gamma(\tau), \quad C(0) = 0.$$

Take the integrating factor

$$\begin{aligned}
K(\tau) &= \exp\left(-\int_0^\tau p(s)ds\right) = \exp\left(\int_0^\tau (\rho\sigma\frac{u}{T}s + \kappa)ds\right) \\
&= \exp\left(\rho\sigma\frac{u}{T}\frac{\tau^2}{2} + \kappa\tau\right).
\end{aligned}$$

We can note that

$$\frac{d}{d\tau}[K(\tau)C(\tau)] = K(\tau)\gamma(\tau).$$

Therefore,

$$C(\tau) = \frac{1}{K(\tau)}\int_0^\tau K(s)\gamma(s)ds.$$

Substituting back into $C(\tau)$, we get

$$C(\tau) = \exp\left(-\frac{\rho\sigma u}{2T}\tau^2 - \kappa\tau\right)\int_0^\tau \exp\left(\frac{\rho\sigma u}{2T}s^2 + \kappa s\right)\left(\frac{u^2}{2T^2}s^2 + \frac{u}{2T}s\right)ds.$$

Thus, we can write $A(\tau)$ as

$$A(\tau) = \int_0^\tau \left[r\left(-\frac{u}{T}s\right) + \kappa\xi C(s)\right]ds = -r\frac{u}{2T}\tau^2 + \kappa\xi\int_0^\tau C(s)ds.$$