Bayesian Calibration of Option Pricing Models Using Sequential Monte Carlo Samplers

R. Brignone[1] L. Gonzato[2] S. Knaust[1] E. Lütkebohmert[1]

ICCF24, Amsterdam, April 2-5, 2024

[1]Department of Quantitative Finance, University of Freiburg
[2]Department of Statistics and Operations Research, University of Vienna

Table of contents

Introduction

Motivation

- Standard calibration of option pricing models to the implied volatility (IV) surface is based on the following optimization:

$$\widehat{\Theta} = \arg\min_{\Theta} \frac{1}{n} \sum_{t=1}^{n_T} \sum_{j=1}^{n_K} \left(IV_{t,j}^{\text{mkt}} - IV_{t,j}^{\Theta} \right)^2 \tag{1}$$

  where $n = n_T \times n_K$, n_T is the number of maturities, n_K is the number of strikes. The calibration must be solved numerically

- Accurate model calibration is important for derivatives pricing, risk-management, portfolio selection and volatility forecasting

- Pitfalls of the standard calibration approach in (1):
  1. there is no established way of choosing the starting point
  2. multiple local minima
  3. measuring uncertainty is usually not possible

Contributions

- We turn the calibration problem into a Bayesian estimation task based on Sequential Monte Carlo (SMC) methods, avoiding the traditional issues of (1) thanks to density tempering
- We provide extensive results for the Stochastic Volatility Correlated Jumps (SVCJ) model of Duffie et al. (2000) both on simulated and real data
- Our approach largely outperforms the benchmark in terms of run-time accuracy and stability of estimated parameters
- We show how to speed up computations by leveraging delayed-acceptance MCMC methods and Deep Learning (DL)

Sequential Monte Carlo Optimization

SMC Beyond State-Space Models

- SMC can be used to sample from the posterior distribution of static parameters, $\pi(\Theta \mid \mathcal{D}) \propto \mathcal{L}(\Theta; \mathcal{D}) p(\Theta)$, where $\mathcal{L}(\Theta; \mathcal{D})$ is the likelihood function of data $\mathcal{D}$ and $p(\Theta)$ is a prior distribution
- The algorithm represents $\pi(\Theta \mid \mathcal{D})$ by simulating N weighted particles $\{w^{(i)}, \Theta_i\}_{i=1}^{N}$ using importance sampling and resampling

SMC Beyond State-Space Models

- SMC can be used to sample from the posterior distribution of static parameters, $\pi(\Theta \mid \mathcal{D}) \propto \mathcal{L}(\Theta; \mathcal{D}) p(\Theta)$, where $\mathcal{L}(\Theta; \mathcal{D})$ is the likelihood function of data $\mathcal{D}$ and $p(\Theta)$ is a prior distribution

- The algorithm represents $\pi(\Theta \mid \mathcal{D})$ by simulating N weighted particles $\{w^{(i)}, \Theta_i\}_{i=1}^N$ using importance sampling and resampling

- Main consequence:

  SMC also works with a generic objective function $h(\Theta; \mathcal{D})$. Let $\psi(\Theta; \mathcal{D}) \propto \exp[h(\Theta; \mathcal{D})]$, then $\psi(\Theta; \mathcal{D})$ can be seen as a density up to a normalizing constant such that $\pi(\Theta \mid \mathcal{D}) \propto \psi(\Theta; \mathcal{D}) p(\Theta)$

SMC Beyond State-Space Models

- SMC can be used to sample from the posterior distribution of static parameters, $\pi(\Theta \mid \mathcal{D}) \propto \mathcal{L}(\Theta; \mathcal{D}) p(\Theta)$, where $\mathcal{L}(\Theta; \mathcal{D})$ is the likelihood function of data $\mathcal{D}$ and $p(\Theta)$ is a prior distribution

- The algorithm represents $\pi(\Theta \mid \mathcal{D})$ by simulating N weighted particles $\{w^{(i)}, \Theta_i\}_{i=1}^N$ using importance sampling and resampling

- Main consequence:

  SMC also works with a generic objective function $h(\Theta; \mathcal{D})$. Let $\psi(\Theta; \mathcal{D}) \propto \exp[h(\Theta; \mathcal{D})]$, then $\psi(\Theta; \mathcal{D})$ can be seen as a density up to a normalizing constant such that $\pi(\Theta \mid \mathcal{D}) \propto \psi(\Theta; \mathcal{D}) p(\Theta)$

- The standard calibration can be easily cast into this framework!

😎

Likelihood Tempering

- Moving from $p(\Theta)$ to $\pi(\Theta \mid \mathcal{D})$ in one step can be very difficult. As in Del Moral et al. (2006), we can construct a sequence of intermediate target distributions $\pi_{\gamma_p}(\Theta \mid \mathcal{D})$, $p = 1, \ldots, P$

$$\pi_{\gamma_p}(\Theta \mid \mathcal{D}) \propto \mathcal{L}(\Theta; \mathcal{D})^{\gamma_p} p(\Theta) \tag{2}$$

  where $0 = \gamma_1 < \ldots < \gamma_P = 1$ is called tempering schedule

- The sequence $\gamma_{1:P}$ can be chosen adaptively to maintain a certain effective sample size, $\mathrm{ESS}_p = (\sum_{i=1}^N w_p^{(i)})^2 / \sum_{i=1}^N (w_p^{(i)})^2$, where

$$w_{p+1}^{(i)} = w_p^{(i)} \frac{\mathcal{L}(\Theta_i; \mathcal{D})^{\gamma_{p+1}} p(\Theta)}{\mathcal{L}(\Theta_i; \mathcal{D})^{\gamma_p} p(\Theta)} = w_p^{(i)} \mathcal{L}(\Theta_i; \mathcal{D})^{\gamma_{p+1} - \gamma_p} \tag{3}$$

  are the importance weights. We choose the next γ_{p+1} to ensure $N_{\mathrm{ESS}} = \eta \cdot N$, where $\eta \in (0, 1)$ and N is the number of particles

- Finally, we resample the particles using $w_p^{(i)}$ and then we move them by running N_{MH} iterations of a Metropolis-Hastings (MH) algorithm

Application to Option Pricing

SVCJ Model Specification

- Given $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{Q})$, the risk-neutral dynamics of the log-returns $X_t = \ln(S_t/S_0)$ is defined as follows:

$$dX_t = (r - 0.5V_t - \lambda\mu^\star)dt + \sqrt{V_t}dW_t^x + J_x dN_t \tag{4}$$

$$dV_t = k(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^v + J_v dN_t \tag{5}$$

  where $\mathbb{E}[dW_t^x dW_t^v] = \rho dt$, $dN_t \sim \text{Poi}(\lambda dt)$, $J_v \sim \text{Exp}(\mu_v)$ and $(J_x \mid J_v) \sim \mathcal{N}(\mu_J + \rho_J J_v, \sigma_J^2)$
- This model has been estimated by many researchers on time series of returns and option prices. We mention Eraker (2004), Broadie et al. (2007) and Dufays et al. (2022)
- There is common agreement that jump parameters are the most difficult to pin down

Standard and Bayesian Calibration

- The standard calibration consists in solving (1), where $\Theta = \{V_0, k, \theta, \sigma, \rho, \mu_V, \lambda, \mu_J, \sigma_J, \rho_J\}$ and where $IV^{\Theta}_{t,j}$ is computed numerically with the Fourier-cosine (COS) method. Then, We proceed in two steps:
    1. generate M parameters from $p(\Theta)$, compute the average MSE and store the $m = \lfloor\sqrt{M/100}\rfloor$ parameter sets with the smallest MSE
    2. run m different optimizations (we use the Nelder-Mead algorithm) using those points as initial guess

Standard and Bayesian Calibration

- The standard calibration consists in solving (1), where $\Theta = \{V_0, k, \theta, \sigma, \rho, \mu_V, \lambda, \mu_J, \sigma_J, \rho_J\}$ and where $IV_{t,j}^{\Theta}$ is computed numerically with the Fourier-cosine (COS) method. Then, We proceed in two steps:
  1. generate M parameters from $p(\Theta)$, compute the average MSE and store the $m = \lfloor \sqrt{M/100} \rfloor$ parameter sets with the smallest MSE
  2. run m different optimizations (we use the Nelder-Mead algorithm) using those points as initial guess

- For the Bayesian calibration we exploit the link between (1) and a Gaussian likelihood function as follows:

$$\mathcal{L}(\Theta; \mathcal{D}) = \frac{1}{\sqrt{2\pi\tilde{\sigma}^2(\Theta)}} \exp\left\{ -\frac{\sum_{t=1}^{n_T} \sum_{j=1}^{n_K} (IV_{t,j}^{\text{mkt}} - IV_{t,j}^{\Theta})^2}{2\tilde{\sigma}^2(\Theta)} \right\} \tag{6}$$

  and sample from the resulting posterior by SMC

Application to Option Pricing

Simulated Data

Experiments Setup

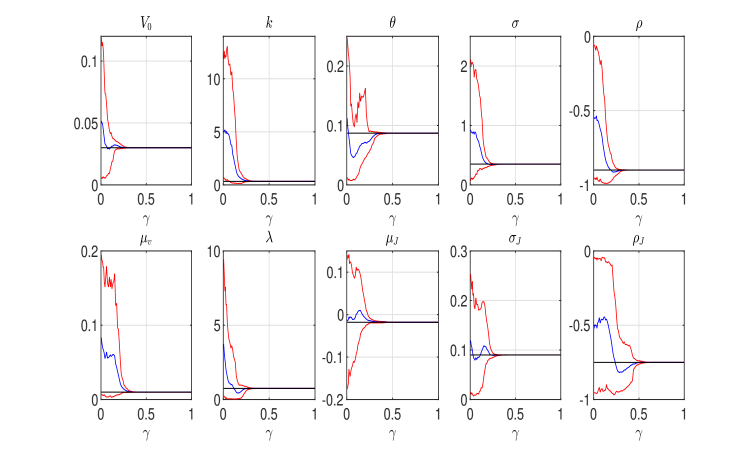
i) We assess the convergence of the Bayesian algorithm to the global optimum on simulated data (single day)

- We compute artificial IV surface for $\tau = \{1, 2, 3, 6, 9, 12\}$ months and $K = \{80, 85, 90, 95, 100, 105, 110, 115, 120\}$ using the parameters estimated by Dufays et al. (2022)
- Since the likelihood is a complicated non-linear function of Θ, we simply assign a Normal prior to all parameters
- We consider very conservative tuning parameters: $N = 512$ particles, $N_{\text{ESS}} = 0.9 \cdot N$ and $N_{\text{MH}} \approx 15$

ii) We compare our approach against the benchmark over 50 repetitions and different sets of parameters. We let N and M vary and fix $N_{\text{ESS}} = 0.7 \cdot N$ and $N_{\text{MH}} \approx 25$

i) Convergence to Global Optimum

Figure 1: Convergence of the parameters wrt γ on simulated data



Notes. Red line denotes (5%, 95%) percentiles, blue line denotes the posterior mean, black line denotes the true parameter value

ii) Standard vs Bayesian

Table 1: Run-time accuracy comparison. Parameters: DJLR (Dufays et al., 2022), BCJ (Broadie et al., 2007), E (Eraker, 2004)

Standard	DJLR		BCJ		E	
M	IV RMSE	time	IV RMSE	time	IV RMSE	time
400	2.52E-03	3.10	1.77E-03	4.20	1.38E-03	3.4
1600	1.05E-03	6.40	6.39E-04	8.90	4.44E-04	6.90
6400	5.14E-04	13.10	2.65E-04	19.00	1.74E-04	14.0
25600	**9.91E-05**	**27.80**	**1.20E-04**	**38.6**	**8.15E-05**	**28.10**
Bayesian	DJLR		BCJ		E	
N	IV RMSE	time	IV RMSE	time	IV RMSE	time
80	**1.10E-05**	**3.10**	1.86E-04	2.90	**4.16E-05**	**2.80**
160	8.52E-06	6.50	1.44E-04	6.00	3.61E-05	5.50
320	6.90E-06	13.20	**1.05E-04**	**12.50**	2.54E-05	12.60
640	3.36E-06	26.50	6.78E-05	26.70	2.12E-05	23.10

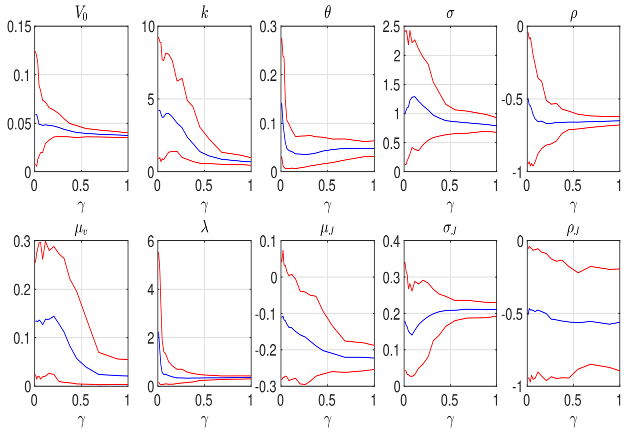
Notes. IV RMSE and computing time (in minutes) averaged across 50 repetitions

Application to Option Pricing

S&P500 Data (March 3, 2021)

Bayesian Calibration (I)

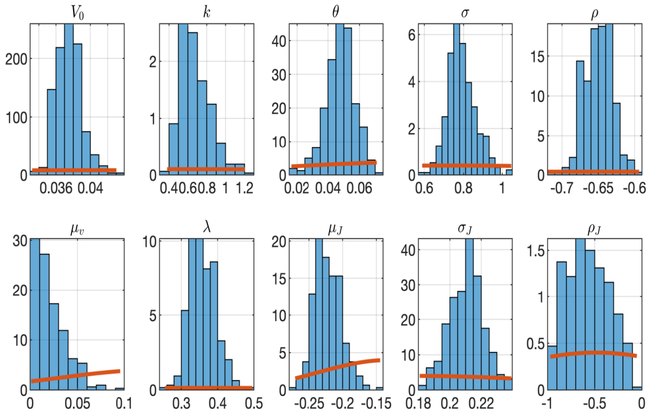
Figure 2: Parameters convergence



Notes. Red line denotes (5%, 95%) percentiles, blue line denotes the posterior mean.
Settings: $N = 320$, $N_{\mathrm{ESS}} = 0.7 \cdot N$, $N_{\mathrm{MH}} \approx 25$

Bayesian Calibration (II)

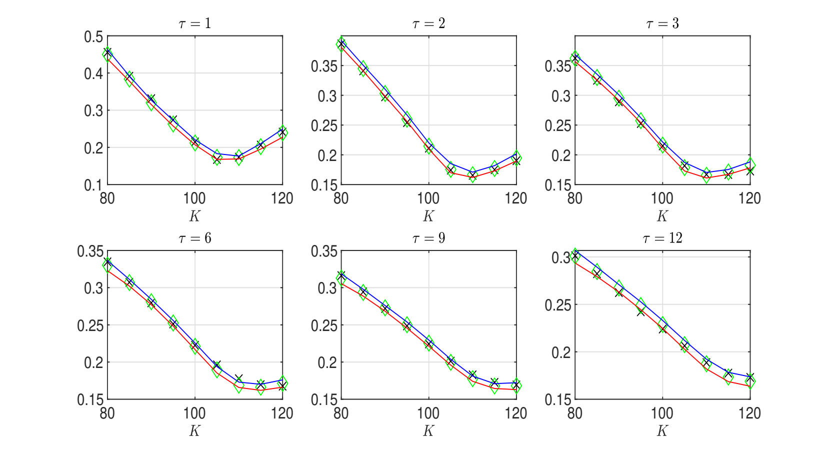
Figure 3: Prior vs Posterior



Notes. The red line represents the prior, while the blue histogram represents the posterior

Bayesian Calibration (III)

Figure 4: Reproducing the IV surface on March 3, 2021

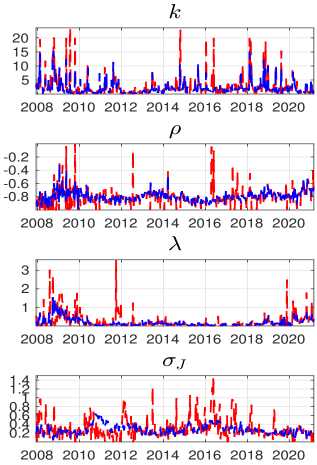


Notes. IV^{mkt} (black crosses) and IV^{Θ} (green diamonds) is the posterior mean

Application to Option Pricing

S&P500 Data (05.12.2007 - 03.03.2021)

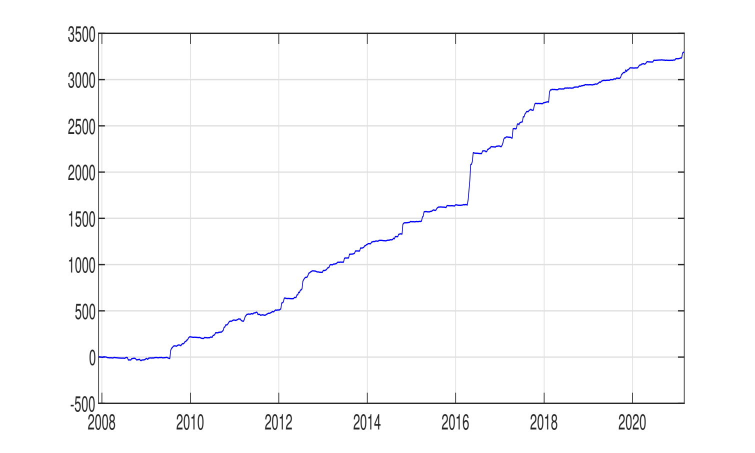
Sequential Calibration: Standard vs Bayesian



Notes. Blue line denotes the Bayesian calibration, red line denotes the standard one

Cumulative Log-Likelihood Ratio

Figure 5: Cumulative log-likelihood ratio between the Bayesian and the standard calibration



Speeding Up:
Delayed-Acceptance MCMC +
Deep Learning

Pricing via Deep Learning

- With more complex models, the likelihood evaluation (involving the calculation of option prices) will require either a numerical solution of a ODE system or a large number of MC simulations for each particle
- We suggest replacing the COS option pricing function by means of Neural Networks (NN)
- Issue: how to perform the NN training effectively?
  1. instead of randomly generating model parameters, we consider as input the posteriors obtained with $N = 800$ particles over the entire data set (i.e., 553 600 set of parameters)
  2. we exploit closed-form risk-neutral cumulants as additional inputs
- We improve the MSE in the validation set when compared to the traditional approach

appendix cumulants + validation

Delayed-Acceptance MCMC

- Bayesian calibration via DL is extremely fast, but its performance highly depends on the NN training. On the other hand, the COS method is slower but more accurate
- By using delayed-acceptance MCMC (Golightly et al., 2015) we combine the speed of DL with the accuracy of COS
- At every MCMC step, we compute a first acceptance probability with DL, and then we compute a second acceptance probability with COS only for those particles accepted in the first stage
- In this way, we avoid expensive calculations for proposals that are likely to be rejected. The algorithm still targets the correct stationary distribution

Results

Table 2: Time Normalized Implied Volatility Root Mean Square Error (TNIVRMSE = time × IVRMSE) for different methods and different experiments

	Simulated Data	March 3, 2021	10.03.21–13.12.23
COS	**5.85E-05**	3.84E-02	6.59E-02
DL	1.96E-04	**1.50E-03**	**1.50E-03**
DL-COS	1.61E-04	1.90E-02	4.70E-02

Key takeaways:

1. On simulated data, it is not possible to achieve the same accuracy as the COS-based approach
2. The DL-based method, equipped with an efficient NN training, outperforms the alternatives on real data
3. The DL-COS approach can be effective if efficient NN training is not available and when the computational cost increases

Conclusions

Final Remarks

- We revisit the standard calibration approach by turning the optimization problem into a Bayesian estimation task
- We show the superiority of our approach wrt the benchmark both on simulated and real option data in terms of run-time accuracy and stability of estimated parameters over time
- We can drastically reduce the computational time by exploiting a NN trained on the sequential posterior estimated on real data
- To avoid a tight dependence on the NN training, it is possible to use delayed-acceptance MCMC to ensure accuracy similar to the COS while reducing computing time (from 20 to 50%)

Final Remarks

- We revisit the standard calibration approach by turning the optimization problem into a Bayesian estimation task
- We show the superiority of our approach wrt the benchmark both on simulated and real option data in terms of run-time accuracy and stability of estimated parameters over time
- We can drastically reduce the computational time by exploiting a NN trained on the sequential posterior estimated on real data
- To avoid a tight dependence on the NN training, it is possible to use delayed-acceptance MCMC to ensure accuracy similar to the COS while reducing computing time (from 20 to 50%)

Thank you!

Main References

Y. Cui, S. del Bano Rollin, and G. Germano.
Full and fast calibration of the Heston stochastic volatility model.
European Journal of Operational Research, 263(2):625–638, 2017.

P. Del Moral, A. Doucet, and A. Jasra.
Sequential Monte Carlo samplers.
Journal of the Royal Statistical Society: Series B, 68(3):411–436, 2006.

J. C. Duan, S. Li, and Y. Xu.
Sequential Monte Carlo optimization and statistical inference.
Wiley Interdisciplinary Reviews: Computational Statistics, 15(3):e1598, 2023.

A. Dufays, K. Jacobs, Y. Liu, and J. Rombouts.
Fast filtering with large option panels: Implications for asset pricing.
Journal of Financial and Quantitative Analysis, pages 1–56, 2023.

A. Golightly, D. Henderson, and C. Sherlock.
Delayed acceptance particle MCMC for exact inference in stochastic kinetic models.
Statistics and Computing, 25:1039–1055, 2015.

Appendix

Appendix

A1. Risk-neutral Cumulants

SVCJ Model

- Under the SVCJ model the MGF of X is the following

$$m_X(u) = \mathbb{E}^{\mathbb{Q}}\left[e^{uX_\tau}\right] = e^{A(u,\tau)+B(u,\tau)V_0}$$

where

$$\frac{\partial A(u,\tau)}{\partial \tau} = ru + k\theta B(u,\tau) - \lambda\mu^* u + \lambda\left(\frac{e^{u\mu_J + u^2\sigma_J^2/2}}{1 - B(u,\tau)\mu_V - \rho_J\mu_V u} - 1\right)$$

$$\frac{\partial B(u,\tau)}{\partial \tau} = -\frac{1}{2}(u - u^2) - (k - \rho\sigma u)B(u,\tau) + \frac{1}{2}\sigma^2 B(u,\tau)^2$$

with $\tau = T - t$, $T > t$, and initial conditions $A(u,0) = B(u,0) = 0$

- Taking the derivative of $\log m_X(u)$ we get the cumulants of X

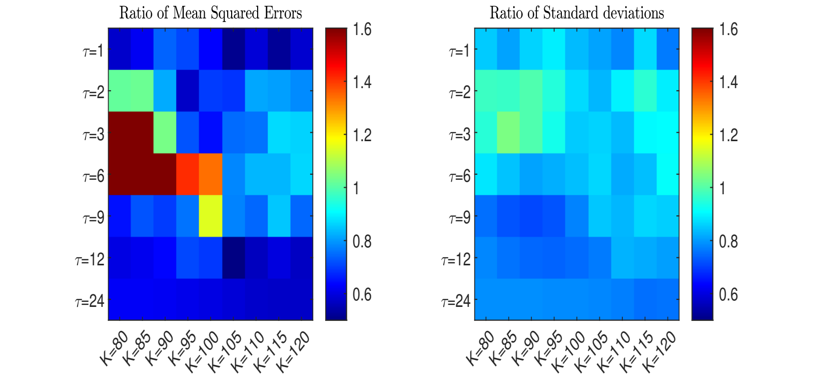
$$c_{n,\tau} = \left.\frac{\partial^n \log m_X(u)}{\partial u^n}\right|_{u=0} = \left.\frac{\partial^n A(u,\tau)}{\partial u^n}\right|_{u=0} + \left.\frac{\partial^n B(u,\tau)}{\partial u^n}\right|_{u=0} V_0$$

where $\left.\frac{\partial^n A(u,\tau)}{\partial u^n}\right|_{u=0}$ and $\left.\frac{\partial^n B(u,\tau)}{\partial u^n}\right|_{u=0}$ are known in closed-form

Back

Performance of the NN (Validation Set)

Figure 6: Ratio of MSE and standard deviation of the pricing functions computed using NNs without and with the first four risk-neutral cumulants



Notes. A value smaller than 1 indicates better performances for the case with four cumulants

Appendix

A2. Feller Condition

Feller Condition 🤔

- In ALL numerical experiments (including the NN training), we do not enforce the Feller condition
- Practical motivation: to obtain a good fit, a violation of this condition is typically required, as shown for instance in Broadie and Kaya (2007) and Cui et al. (2017)
- General warnings:
  1. MC simulation should exploit suitable algorithms to handle this violation (see Begin et al., 2015)
  2. the characteristic function can be discontinuous and give wrong prices with Fourier-based methods (see Cui et al., 2017)
  3. if under $\mathbb{P}$ the Feller condition is violated, an equivalent risk-neutral measure $\mathbb{Q}$ may not exists (see Desmettre et al., 2021)
- Since we are already working under $\mathbb{Q}$, this does not pose any theoretical problem (see Desmettre et al., 2021)

Appendix

A.3 Prior and Bounds

Prior and Bounds

Table 3: Prior specification. The mean of the prior, μ_0, depends on the experiment at hand. For instance, on real data μ_0 is the average between the parameters estimated in Broadie et al. (2007) and Eraker (2004)

Θ	Distrib.	σ_0	Support	Θ	Distrib.	σ_0	Support
V_0	Tr. Normal	0.05	$(0, \infty)$	k	Tr. Normal	4.00	$(0, \infty)$
θ	Tr. Normal	0.10	$(0, \infty)$	σ	Tr. Normal	1.00	$(0, \infty)$
ρ	Tr. Normal	1.00	$(-1, 0)$	μ_V	Tr. Normal	0.10	$(0, \infty)$
λ	Tr. Normal	3.00	$(0, \infty)$	μ_J	Normal	0.10	$(-\infty, \infty)$
σ_J	Tr. Normal	0.10	$(0, \infty)$	ρ_J	Tr. Normal	1.00	$(-1, 0)$

To prevent the possibility of obtaining totally unrealistic parameter estimates, we impose additional upper and lower bounds for the standard calibration

$$V_0 \in (0, 1.5), \quad k \in (0, 20), \quad \theta \in (0, 0.6), \quad \sigma \in (0, 5), \quad \rho \in (-1, 0),$$
$$\mu_V \in (0, 2), \quad \lambda \in (0, 10), \quad \mu_J \in (-1, 1), \quad \sigma_J \in (0, 1.5), \quad \rho_J \in (-1, 0)$$