Particle Markov Chain Monte Carlo

with numerical simulations for the Heston model

Daniel Mitsutani

Department of Mathematics, University of Chicago

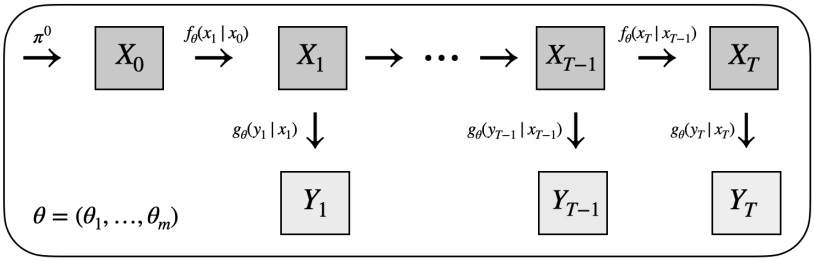
1 Introduction

A prominent problem in financial econometrics is modeling the volatility of an asset's price time series. The price series S_t is widely modeled by the stochastic differential equation $dS_t = \mu_t S_t\, dt + \sigma_t S_t\, dW_t$, where dW_t is a Brownian motion, μ_t is a time process known as drift and σ_t is known as the *volatility process*. When σ_t itself is a stochastic process, the model is referred to as a *stochastic volatility model*. A popular choice is to model σ_t as an Ornstein-Uhlenbeck process, which leads to the *Heston model* $d\sigma_t^2 = \kappa(\theta - \sigma_t^2)\, dt + \sqrt{\xi}\sigma_t dB_t$, where dB_t is another Brownian motion which is allowed to have correlation ρ with dW_t, and $\kappa, \theta, \xi > 0$ are further parameters in the model.

Prices are observable quantities, whereas volatility is modeled as a latent (or hidden) process, which is to be estimated based on observations of S_t. Since σ_t is a Markov process, such a model is referred to as a *hidden Markov model*. The goal of this project is to introduce particle Markov chain Monte Carlo (PMCMC) algorithms, which are a popular choice for Bayesian estimation of hidden Markov model parameters such as the ones in the Heston model. PMCMC, as we will see, combines two important Monte Carlo methods: Markov Chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC), by using SMC samples to construct MCMC kernels.

1.1 General Setting

The general setting for the project consists of a hidden Markov model with objects represented in the following diagram:



(Unknown Parameters):	$\theta = (\theta_1, \ldots, \theta_m) \in E \subseteq \mathbb{R}^m$
(Initial State):	$X_0 \sim \pi^0$
(Latent Markov Process):	$X_{t+1} \,\vert\, (X_t = x_t) \sim f_\theta(\cdot \,\vert\, x_t),\ \ 0 \le t \le T$
(Observable Process):	$Y_t \,\vert\, (X_t = x_t) \sim g_\theta(\cdot \,\vert\, x_t),\ \ 1 \le t \le T$

Then the joint distribution of the latent process and the data is given by:

$$p_\theta(x_{1:T}, y_{1:T}) = \pi^0(x_0) \prod_{t=1}^{T} g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1}).$$

Given fixed observations $y_{1:T} = y_1, \ldots, y_T$ and a prior $p(\theta)$ on the parameters, we will take it as our goal to sample from $p(\theta \,\vert\, y_{1:T})$, namely, the posterior of the parameters conditional on the observed process.

1.2 Motivation for PMCMC: Marginal Metropolis-Hastings

To understand the problem at hand, consider the idealized situation in which the model likelihood given the observations $\mathcal{L}(\theta) := p_\theta(y_{1:t})$ is tractable. In such a scenario, a Metropolis-Hastings mechanism can be used to sample from the posterior by choosing a proposal kernel for $\theta^* \sim q(\theta, \cdot)$ and accepting with probability:

$$1 \wedge \frac{q(\theta^*, \theta)}{q(\theta, \theta^*)} \cdot \frac{\mathcal{L}(\theta^*)p(\theta^*)}{\mathcal{L}(\theta)p(\theta)},$$

so that the samples of the chain have as invariant distribution the desired posterior distribution of θ, namely $p(\theta \,|\, y_{1:T}) \propto p_\theta(y_{1:T})p(\theta) = \mathcal{L}(\theta)p(\theta)$.

This idealized version of the classic Metropolis-Hastings algorithm applied to the sampling of the marginal $p(\theta \,|\, y_{1:T})$ is known as *marginal Metropolis-Hastings*. In practice, such an approach is not feasible since the evaluation of $\mathcal{L}(\theta)$ generally requires integration over all realizations of the latent hidden process in a high-dimensional space.

The core idea behind particle Markov chain Monte Carlo methods is to use a sequential Monte Carlo algorithm to instead generate such realizations of the latent process, and then use these realizations to obtain estimates $\widehat{\mathcal{L}}(\theta)$ to compute the acceptance probability. A surprising strength of this method is the fact that even though we use such an approximate acceptance probability, one can prove that the samples generated by the chain actually are distributed exactly according to $\mathcal{L}(\theta)p(\theta)$.

2 Sequential Monte Carlo

The *bootstrap particle filter* is a sequential Monte Carlo algorithm which produces weighted empirical measures $\widehat{\pi}^t$ which approximate the filtering distribution $\pi^t := p_\theta(x_t \,|\, y_{1:t})$. For the sake of establishing notation, we recall its pseudo-algorithm:

Algorithm 2.1 Bootstrap particle filter

Input: Hidden Markov model $(\theta,\, f_\theta,\, g_\theta,\, \pi^0)$ with fixed θ, sample size N, observations $y_{1:T}$, initial distribution $\widehat{\pi}^0 := \pi^0$.

For $0 \leq t \leq T-1$, do for $1 \leq n \leq N$:

1: Sample $\chi_t^n \sim \widehat{\pi}^t(\cdot)$.

2: Propagate with Markov kernel: $X_{t+1}^n \sim f_\theta(\cdot \,|\, \chi_t^n)$.

3: Compute weights and normalized weights:

$$w_{t+1}^n = g_\theta(y_{t+1} \,|\, X_{t+1}^n), \quad \widetilde{w}_{t+1}^n = w_{t+1}^n / \left(\sum_{n=1}^N w_{t+1}^n \right).$$

4: Set $\widehat{\pi}^{t+1}(x) = \sum_{n=1}^N \widetilde{w}_{t+1}^n \delta(x - X_{t+1}^n)$.

Output: Emprical weight measures $\widehat{\pi}^t$ approximating $p_\theta(x_t \,|\, y_{1:t})$.

We introduce two objects obtained from an SMC run needed for PMCMC.

2.1 Likelihood estimation

The likelihood $\mathcal{L}(\theta) = p_\theta(y_{1:T})$ can be approximated from the bootstrap filter inductively using the factorization

$$p_\theta(y_{1:T}) = p_\theta(y_1) \prod_{t=2}^{T} p_\theta(y_t \,|\, y_{1:t-1}).$$

Recalling that $\widehat{\pi}^{t-1}(dx_{t-1})$ approximates $p_\theta(x_{t-1} \,|\, y_{1:t-1})\, dx_{t-1}$, the individual terms can be rewritten as:

$$\begin{aligned} p_\theta(y_t \,|\, y_{1:t-1}) &= \int p_\theta(y_t \,|\, x_t, y_{1:t-1})\, p_\theta(x_t \,|\, y_{1:t-1})\, dx_t \\ &= \int g_\theta(y_t \,|\, x_t)\, p_\theta(x_t \,|\, x_{t-1})\, p_\theta(x_{t-1} \,|\, y_{1:t-1})\, dx_{t-1}\, dx_t \\ &= \int g_\theta(y_t \,|\, x_t)\, \mathcal{P}\pi^{t-1}(x_t)\, dx_t \approx \mathbb{E}_{\mathcal{S}^N \mathcal{P} \widehat{\pi}^{t-1}}[\, g_\theta(y_t \,|\, \cdot)\,], \end{aligned}$$

where $\mathcal{P}$ and $\mathcal{S}^N$ are the prediction and sampling operators. Note that we compute this expectation by propagating, sampling, and evaluating g_θ, which precisely the t-th step of the algorithm, which therefore yields $p_\theta(y_t \,|\, y_{1:t-1}) \approx (\sum_{n=1}^N w_t^n)/N$. This motivates defining the likelihood estimator by:

$$\widehat{\mathcal{L}}(\theta) := \prod_{t=1}^{T} \left(\frac{1}{N} \sum_{n=1}^{N} w_t^n \right) = \prod_{t=1}^{T} \left(\frac{1}{N} \sum_{n=1}^{N} g_\theta(y_t | X_t^n) \right),$$

which, as previously mentioned, is the basis for the PMCMC algorithms we will describe.

2.2 Ancestor map and full path sampling

Recall that X_t^n is obtained by sampling some χ_{t-1}^n from $\widehat{\pi}^{t-1}$ and propagating via the Markov kernel. Since $\widehat{\pi}^{t-1}$ is supported on the previous X_{t-1}^*'s, we have $\chi_{t-1}^n = X_{t-1}^m$ for some $1 \le m \le N$. We then write $a_t^n := m$ and call X_{t-1}^m the *ancestor* of X_t^n. Note that with the notation above X_0^n is technically not defined, so that for the sake of consistency we set $X_0^n := \chi_0^n$.

The above definition can be used to define full paths inductively backwards. Namely, given $1 \le n \le N$, we start from X_T^n and trace back its ancestors inductively, which defines the *full paths*

$$X_{0:T}^n := (X_0^{B_0^n}, X_1^{B_1^n}, \dots X_T^{B_T^n}),$$

where B_t^n are the indices obtained from inductively applying the ancestor map, i.e. $B_t^n = a_t^{B_{t+1}^n}$ starting with $B_T^n = n$. To then sample full paths $X_{0:T}$ from the SMC run, one can sample from $\{X_{0:T}^n\}_{n=1}^N$ with probability vector $(\tilde{w}_T^n)_{n=1}^N$, and it can be shown the samples obtained have distribution approximately $p_\theta(x_{0:T} \,|\, y_{1:T})$.

3 Particle Marginal Metropolis-Hastings Sampler

We introduce the particle marginal Metropolis-Hastings (PMMH) algorithm, a PMCMC algorithm with a simple core idea: we run the marginal Metropolis-Hastings algorithm, while using an SMC of choice to approximate $\mathcal{L}(\theta)$ as in Section 2.1.

Algorithm 3.1 Particle marginal Metropolis-Hastings sampler

Input: Hidden Markov model $(f_\theta,\ g_\theta,\ \pi^0)$, sample size N, number of iterations M, observations $y_{1:T}$, initial distribution $\widehat{\pi}^0 := \pi^0$, Markov kernel q, initialization θ_0 and prior $p(\theta)$ for θ, SMC algorithm of choice.

Initialization: Run an SMC algorithm with parameter θ to obtain $\widehat{\mathcal{L}}(\theta_0)$.

For $1 \le i \le M$:

1: Draw $\theta^* \sim q(\theta_{i-1},\ \cdot)$.
2: Run an SMC algorithm with parameter θ^* to obtain $\widehat{\mathcal{L}}(\theta^*)$.
3: Set

$$a = \frac{q(\theta^*, \theta_{i-1})}{q(\theta_{i-1}, \theta^*)} \cdot \frac{\widehat{\mathcal{L}}(\theta^*)\, p(\theta^*)}{\widehat{\mathcal{L}}(\theta_{i-1})\, p(\theta_{i-1})}$$

4: With probability $1 \wedge a$, set $\theta_i = \theta^*$ otherwise set $\theta_i = \theta_{i-1}$.

Output: Samples $\{\theta_i\}_{i=1}^M$ from a Markov chain with stationary distribution $p(\theta \,|\, y_{1:T})$.

3.1 Exactness of Stationary Distribution

Unlike MALA or HMC, it is not evident for the PMMH sampler that the transition kernel should satisfy detailed balance with respect to the target $p(\theta \,|\, y_{1:T}) = \mathcal{L}(\theta)p(\theta)$, since the acceptance probability is calculated using $\widehat{\mathcal{L}}$ as opposed to $\mathcal{L}$. However, we can show that the invariant distribution is exactly $\mathcal{L}(\theta)p(\theta)$ by proving that $\widehat{\mathcal{L}}(\theta)$ is unbiased with respect the random draws of the SMC algorithm.

For an SMC run, we write $\mathbb{X}_{0:p} := (\{X_t^n\}_{n=1}^N)_{t=0}^p, \mathbb{A}_{1:p} := (\{a_t^n\}_{n=1}^N)_{t=1}^p$ for the draws up to time p, $\mathbb{X}_p := \{X_p^n\}_{n=1}^N$, $\mathbb{A}_p := \{a_p^n\}_{n=1}^N$ the draws at time p and by q_θ^{SMC} the joint probability density of the full draws: $\mathbb{X}_{0:T}, \mathbb{A}_{1:T} \sim q_\theta^{\mathrm{SMC}}(\cdot\,,\,\cdot)$. By definition, taking the expectation $\mathbb{E}_{\mathbb{X}_{0:T}, \mathbb{A}_{1:T}}$ with respect to the SMC randomness is the same as integrating against the density $q_\theta^{\mathrm{SMC}}(\mathbb{X}_{0:T}, \mathbb{A}_{1:T})\, d\mathbb{X}_{0:T}\, d\mathbb{A}_{1:T}$. We emphasize that although we omit the arguments for readability, $\widehat{\mathcal{L}}(\theta)$ is actually a (deterministic) function of $\theta, \mathbb{X}_{0:T}$ and $\mathbb{A}_{1:T}$. The key proposition, which we prove at the end of the section, is:

Proposition 3.1. *The approximate likelihood $\widehat{\mathcal{L}}(\theta)$ is an unbiased estimator of $\mathcal{L}(\theta)$:*

$$\mathbb{E}_{\mathbb{X}_{0:T}, \mathbb{A}_{1:T}}[\widehat{\mathcal{L}}(\theta)] = \mathcal{L}(\theta).$$

Now to show that the chain $\{\theta_i\}$ from the PMMH sampler admits $p(\theta \,|\, y_{1:T})$ as a stationary distribution, we consider its embedding into a larger Markov chain with state variables $(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})$, where $\mathbb{X}_{0:T}, \mathbb{A}_{1:T}$ are the SMC draws of the latest accepted proposal. In this extended space, the PMMH proposal kernel is given by:

$$q^{\mathrm{ext}}(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T}; \theta^*, \mathbb{X}_{0:T}^*, \mathbb{A}_{1:T}^*) := q(\theta, \theta^*)\, q_{\theta^*}^{\mathrm{SMC}}(\mathbb{X}_{0:T}^*, \mathbb{A}_{1:T}^*).$$

Letting $\Pi(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})$ be a density $\propto \widehat{\mathcal{L}}(\theta)\, p(\theta)\, q_\theta^{\mathrm{SMC}}(\mathbb{X}_{0:T}, \mathbb{A}_{1:T})$, we see that in the extended space the PMMH acceptance probability can be rewritten as:

$$a = \frac{q(\theta^*, \theta)}{q(\theta, \theta^*)} \cdot \frac{\widehat{\mathcal{L}}(\theta^*)\, p(\theta^*)}{\widehat{\mathcal{L}}(\theta)\, p(\theta)} = \frac{q^{\mathrm{ext}}(\theta^*, \mathbb{X}_{0:T}^*, \mathbb{A}_{1:T}^*; \theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})}{q^{\mathrm{ext}}(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T}; \theta^*, \mathbb{X}_{0:T}^*, \mathbb{A}_{1:T}^*)} \cdot \frac{\Pi(\theta^*, \mathbb{X}_{0:T}^*, \mathbb{A}_{1:T}^*)}{\Pi(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})},$$

since the q^{SMC} terms cancel out. Hence, the MH-adjusted PMMH kernel satisfies detailed balance with respect to $\Pi(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})$, which is therefore the stationary distribution of the PMMH extended chain in the $(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})$ variables.

What Proposition 3.1 shows in this context is that the marginal distribution of θ in the distribution Π is precisely our target:

$$\begin{aligned}\Pi(\theta) &\propto \int \widehat{\mathcal{L}}(\theta)\, p(\theta)\, q_\theta^{\mathrm{SMC}}(\mathbb{X}_{0:T}, \mathbb{A}_{1:T})\, d\mathbb{X}_{0:T}\, d\mathbb{A}_{1:T} \\ &= \mathbb{E}_{\mathbb{X}_{0:T}, \mathbb{A}_{1:T}}[\widehat{\mathcal{L}}(\theta)] p(\theta) = \mathcal{L}(\theta) p(\theta).\end{aligned}$$

Under suitable conditions on the support of the SMC extended proposal to ensure ergodicity of the chain, what we have thus shown is that:

Theorem 3.2. *The chain $\{\theta_i\}_{i=1}^M$ from the particle marginal Metropolis-Hastings sampler has stationary distribution $p(\theta \mid y_{1:T}) = \mathcal{L}(\theta)p(\theta)$.*

3.2 Proof of Proposition 3.1

To keep the notation cleaner, we omit the dependence on the fixed θ in this section. Define a family of *unnormalized* measures γ^t for $1 \le t \le T$ using their integral against a test function $\psi(x_t)$ by:

$$\gamma^t(\psi) := \int \pi^0(x_0)\psi(x_t) \prod_{p=1}^{t} f(x_p \mid x_{p-1}) g(y_p \mid x_p)\, dx_{0:t} = \int \psi(x_t) p(x_{0:t} \mid y_{1:t})\, dx_{0:t}.$$

This family satisfies a recursive relation similar to the recursive relation $\pi^t = \mathcal{A}\mathcal{P}\pi^{t-1}$ for the filtering distribution, where $\mathcal{P}$ and $\mathcal{A}$ are the prediction and analysis operators. In fact, the only difference is that the analysis operator $\mathcal{A}$ requires a normalization step which is not applied in the recursive relation for γ^t. We briefly denote this unnormalized version of $\mathcal{A}$ by $\mathcal{A}^{\mathrm{un}}$ so that $\gamma^t = \mathcal{A}^{\mathrm{un}}\mathcal{P}\gamma^{t-1}$.

Heuristically, the idea of the proof is as follows. One can show, essentially by Fubini's theorem, that the expectation $\mathbb{E}$ with respect to the randomness of SMC commutes with $\mathcal{P}$, and that it commutes with $\mathcal{A}$ *except for the normalizing step*. In other words, the expectation commutes with $\mathcal{A}^{\mathrm{un}}$ and $\mathcal{P}$ and so $\mathbb{E}\mathcal{A}^{\mathrm{un}}\mathcal{P} = \mathcal{A}^{\mathrm{un}}\mathcal{P}\mathbb{E}$. Further, using the fact that our resampling scheme is unbiased, one can show that $\mathbb{E}\mathcal{S}^n = \mathbb{E}$ as well. Thus, starting with an unbiased $\widehat{\gamma}^0 := \widehat{\pi}^0$, we can inductively commute the expectation through the recursive relations to show that a $\widehat{\gamma}^t$ satisfying $\widehat{\gamma}^t = \mathcal{A}^{\mathrm{un}}\mathcal{S}^n\mathcal{P}\widehat{\gamma}^{t-1}$ is unbiased.

We make this intuition rigorous below.

Proposition 3.3. *Define for $1 \le t \le T$ the unnormalized empirical measures:*

$$\widehat{\gamma}^t := \Gamma^{t-1}\left(\frac{1}{N}\sum_{n=1}^{N} w_p^n \delta_{X_t^n}\right), \text{ where } \Gamma^{t-1} := \prod_{p=1}^{t-1}\left(\frac{1}{N}\sum_{n=1}^{N} w_t^n\right).$$

Then the $\widehat{\gamma}^t$ are unbiased estimators of γ^t. That is, for any function ψ:

$$\mathbb{E}_{\mathbb{X}_{0:t}, \mathbb{A}_{1:t}}[\widehat{\gamma}^t(\psi)] = \gamma^t(\psi).$$

Proof. Given $\psi(x_t)$, let

$$\Psi(x_{t-1}) := \int g(y_t \,|\, x_t) f(x_t \,|\, x_{t-1})\, \psi(x_t) dx_t,$$

and observe that $\psi \mapsto \Psi$ is dual to $\mathcal{A}^{\mathrm{un}}\mathcal{P}$, i.e., $\gamma^{t-1}(\Psi) = \gamma^t(\psi)$.

The proof is by induction on t. Suppose the statement holds for $t-1$. We begin by computing for $1 \le n \le N$:

$$\begin{aligned}
\mathbb{E}_{\mathbb{X}_t, \mathbb{A}_t}\left[w_t^n \psi(X_t^n) \,|\, \mathbb{X}_{1:t-1} \right] &= \mathbb{E}_{\mathbb{X}_t, \mathbb{A}_t}\left[g(y_t \,|\, X_t^n)\, \psi(X_t^n) \,|\, \mathbb{X}_{1:t-1} \right] \\
&= \mathbb{E}_{\mathbb{A}_t}\left[\int g(y_t \,|\, x_t) f(x_t \,|\, X_{t-1}^{a_t^n})\, \psi(x_t)\, dx_t \,|\, \mathbb{X}_{1:t-1} \right] \\
&= \mathbb{E}_{\mathbb{A}_t}\left[\Psi(X_{t-1}^{a_t^n}) \,|\, \mathbb{X}_{1:t-1} \right] = \sum_{n=1}^N \tilde{w}_{t-1}^n \Psi(X_{t-1}^n),
\end{aligned}$$

where the last equality uses that the *resampling is unbiased*. That is, note that $\mathbb{A}_t$ draws a_t^n from $\{1, \dots N\}$ with probability vector $\{\tilde{w}_{t-1}^n\}_{n=1}^N$ so that:

$$\mathbb{E}_{\mathbb{A}_t}\left[\Psi\left(X_{t-1}^{a_t^n}\right) \,\middle|\, \mathbb{X}_{1:t-1} \right] = \sum_{n=1}^N \tilde{w}_{t-1}^n \Psi(X_{t-1}^n).$$

Using the definition of $\widehat{\gamma}^t$, the identity just proved, and the tower property of expectation:

$$\begin{aligned}
\mathbb{E}_{\mathbb{X}_{1:t}, \mathbb{A}_{1:t}}[\widehat{\gamma}^t(\psi)] &= \mathbb{E}_{\mathbb{X}_{1:t}, \mathbb{A}_{1:t}}\left[\Gamma^{t-1}\left(\frac{1}{N} \sum_{n=1}^N w_p^n\, \psi(X_t^n) \right) \right] \\
&= \mathbb{E}_{\mathbb{X}_{1:t-1}, \mathbb{A}_{1:t-1}}\left[\Gamma^{t-1} \sum_{n=1}^N \tilde{w}_{t-1}^n \Psi(X_{t-1}^n) \right] \\
&= \mathbb{E}_{\mathbb{X}_{1:t-1}, \mathbb{A}_{1:t-1}}\left[\Gamma^{t-2} \sum_{n=1}^N w_{t-1}^n \Psi(X_{t-1}^n) \right] \\
&= \mathbb{E}_{\mathbb{X}_{1:t-1}, \mathbb{A}_{1:t-1}}\left[\widehat{\gamma}^{t-1}(\Psi) \right] = \gamma^{t-1}(\Psi),
\end{aligned}$$

where the third equality follows from $\Gamma^{t-1}/\Gamma^{t-2} = \sum_{n=1}^N w_{t-1}^n$ (with the convention $\Gamma^0 = 1$), the fourth equality is by definition, and the last equality is the inductive hypothesis. The proof of the inductive step is now complete since $\gamma^{t-1}(\Psi) = \gamma^t(\psi)$.

The induction base case $t = 1$ follows along the same lines noting that $\widehat{\pi}^0 = \frac{1}{N}\sum_{n=1}^N \delta_{X_0^n}$ is unbiased for π^0 since $X_0^n \sim^{iid} \pi^0$. □

Now we can prove Proposition 3.1:

Proof of Proposition 3.1. Observe that taking $\psi \equiv 1$ gives $\gamma^t(1) = p(y_{1:t})$. Since by definition $\widehat{\mathcal{L}} = \widehat{\gamma}^T(1)$ and $\mathcal{L} = \gamma^T(1)$, the proof now follows from Proposition 3.3. □

4 Particle Gibbs sampler

An alternative approach to sample from $p_\theta(y_{1:T})$ is available when the *full conditionals* $p(\theta \,|\, x_{0:T}, y_{1:T})$ can be sampled; namely, to combine SMC with a Markov chain obtained via a Gibbs sampler as opposed to a Metropolis-Hastings sampler. To accomplish this, the chain now has to take place in the state space of $(\theta, X_{0:T})$. The θ update can be easily performed since $p(\theta \,|\, x_{0:T}, y_{1:T})$ can be sampled by hypothesis, but the $X_{0:T}$ update has to be done via a so-called *conditional SMC update*.

This conditional SMC update requires a pre-specified fixed *reference trajectory* $X^\star_{0:T}$. We illustrate this by giving the conditional version of the bootstrap filter, and refer to Section 2.2 to remind the reader how one can generate full trajectories from the samples and ancestor map of an SMC run.

Algorithm 4.1 Conditional bootstrap filter update

Input: Same as bootstrap particle filter, plus reference trajectory $X^\star_{0:T}$.

Initialization: Set $X^N_{0:T} := X^\star_{0:T}$ and for $1 \le t \le T$ set weights $w^N_t := g_\theta(y_t \,|\, X^\star_t)$.

For $0 \le t \le T-1$, do for $1 \le n \le N-1$:

1: Sample $\chi^n_t \sim \widehat{\pi}^t(\cdot)$.

2: Propagate with Markov kernel: $X^n_{t+1} \sim f_\theta(\cdot \,|\, \chi^n_t)$.

3: Compute weights and normalized weights:

$$w^n_{t+1} = g_\theta(y_{t+1} \,|\, X^n_{t+1}), \quad \widetilde{w}^n_{t+1} = w^n_{t+1} \Big/ \left(\sum_{n=1}^N w^n_{t+1} \right).$$

4: Set $\widehat{\pi}^{t+1}(x) = \sum_{n=1}^N \widetilde{w}^n_{t+1} \delta(x - X^n_{t+1})$.

End for.

1: Sample k from $\{1, \dots N\}$ with probability vector $(\widetilde{w}^n_T)_{n=1}^N$.

Output: Conditionally updated $X_{0:T} := X^k_{0:T}$, where $\{X^n_{0:T}\}_{n=1}^N$ are as in Section 2.2.

Simply put, the conditional SMC update is a standard SMC update which requires that the full reference trajectory $X^\star_{0:T}$ and its ancestry be preserved through the algorithm. Since for every t the resampling can generate descendants of the reference trajectory, the conditional SMC implicitly guides the newly generated particles towards the reference trajectory $X^\star_{0:T}$.

The key observation, which we will not prove, is that this conditional SMC update can be interpreted as exact variable updates via their full conditionals on the extended state space of the variables $(\theta, k, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})$, where $k \in \{1, \dots N\}$ is a random variable which specifies the reference trajectory amongst the full trajectories generated, and as before $\mathbb{X}_{0:T}, \mathbb{A}_{1:T}$ are the full draws from the last conditional SMC run. Informally speaking, in this state space the conditional SMC performs two updates: first it samples $(\mathbb{X}_{0:T}, \mathbb{A}_{1:T})$ conditional on (θ, k) and then it samples k conditional on $(\theta, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})$.

All that is left for the full Gibbs sampler sweep is to update θ conditional on the remaining variables. This leads to the particle Gibbs sampler (PG).

Algorithm 4.2 Particle Gibbs sampler

Input: Same as PMMH, minus kernel q, plus arbitrary initial trajectory $X_{0:T}(0)$.

For $1 \leq i \leq M$:

1: Sample $X_{0:T}(i)$ from the conditional SMC update, using $X_{0:T}(i-1)$ as a reference trajectory and parameter θ_{i-1}.
2: Sample θ_i from the full conditional $p(\theta \,|\, X_{0:T}(i), y_{1:T})$.

Output: Samples $\{(\theta_i, X_{0:T}(i))\}_{i=1}^M$ from a Markov chain with stationary distribution $p(\theta, x_{0:T} \,|\, y_{1:T})$.

In the same spirit of Proposition 3.1, one can then show that the joint marginal distribution of θ and the reference trajectory $X_{0:T}$ in the full state space of $(\theta, k, \mathbb{X}_{0:T}, \mathbb{A}_{1:T})$ is in fact exactly $p(\theta, x_{0:T} \,|\, y_{1:T})$ so that again, even though we use approximations provided by the SMC algorithm, the generated chain has the exact target distribution.

Remark 4.1. Compared to PMMH, PG has the significant advantage of not requiring the construction of a proposal kernel for θ, which can be hard to tune for a model with many parameters such as the Heston model. On the other hand, it requires sampling from $p(\theta \,|\, x_{0:T}, y_{1:T})$.

A more significant issue with PG is *path degeneracy*, that is, the tendency due to importance resampling to generate full paths with very similar pasts. This, in turn, can cause the update of $X_{0:T}$ to have bad exploration properties, since $X_{0:T}$ can be very similar to the reference trajectory $X^\star_{0:T}$. A significant improvement over the PG can be obtained by resampling the ancestors of the reference trajectory in an unbiased way, which leads to the particle Gibbs with ancestor sampling (PGAS).

5 Numerical simulations for the Heston model

The observed process Y_t we will use is given by the log returns $Y_t = \log(S_t/S_{t-1})$. Rewriting the stochastic differential equations for Y_t instead of S_t and letting $X_t = \sigma_t^2$, Euler discretization of the price and Heston volatility SDEs gives:

$$Y_t = (\mu_t - X_{t-1}/2)\Delta t + (X_{t-1})^{1/2}[\rho\, \varepsilon_t^X + (1-\rho^2)^{1/2}\varepsilon_t^Y],$$

$$X_t = X_{t-1} + \kappa(\theta - X_{t-1})\Delta t + (\xi X_{t-1})^{1/2}\, \varepsilon_t^X,$$

where $\varepsilon_t^Y, \varepsilon_t^X$ are independent $\sim N(0, \Delta t)$. We set $\Delta t = 1/252$ ($= 1/$number of trading days in a year), and make the ad-hoc assumption of no drift $\mu_t \equiv 0$.

The parameters $\Theta = (\rho, \kappa, \theta, \xi)$ in the Heston model have the following interpretation: volatility X_t is a mean reverting process, with mean given by θ and mean-reversion rate κ. The *leverage effect* ρ governs the correlation between stochastic movements of volatility and prices. For example, in equity markets high volatility often correlates to negative returns and vice versa, and ρ is quite negative. Finally, ξ governs the stochasticity of X_t itself, and is known as *vol of vol*.

Observe that Y_t depends directly on X_t and X_{t-1}, yet our theory can be adjusted without trouble to handle this case by considering the state space of the latent chain to be the pair (X_t, X_{t-1}). The densities in the model are given by:

$$f_\Theta(X_t \,|\, X_{t-1}) \sim N(\kappa(\theta - X_{t-1})\Delta t + X_{t-1},\, \xi X_{t-1}\Delta t),$$
$$g_\Theta(Y_t \,|\, X_t, X_{t-1}) \sim N((\mu - X_{t-1}/2)\Delta t + (X_{t-1})^{1/2}\rho\,\varepsilon_t^X,\, (1-\rho^2)X_{t-1}\Delta t),$$

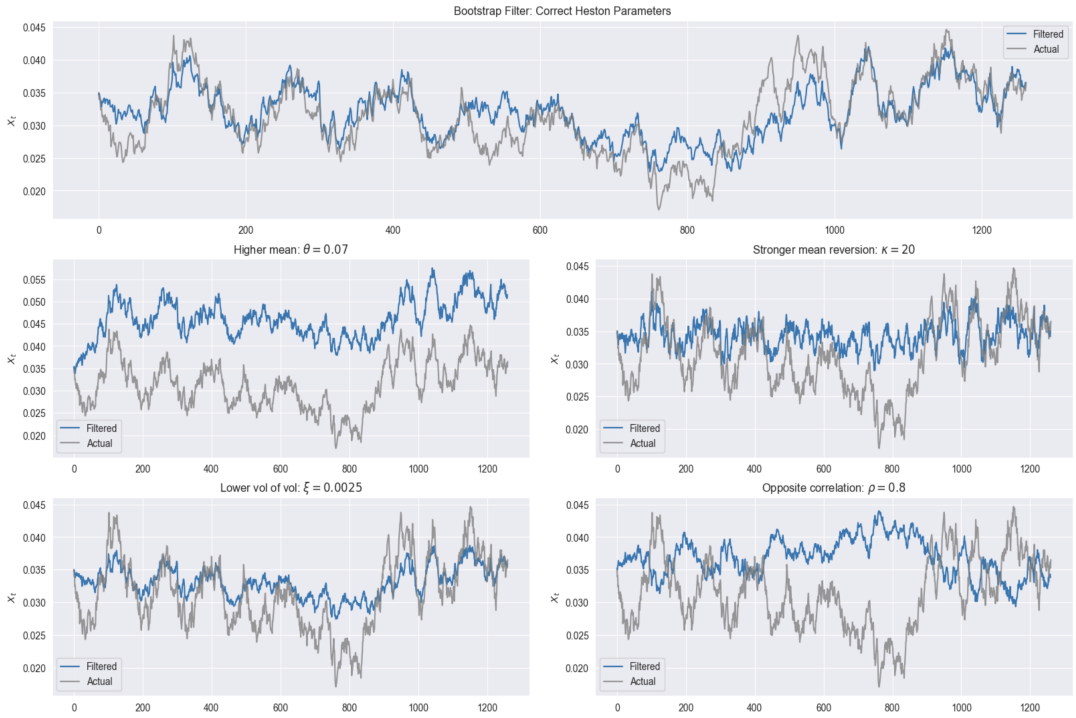
where ε_t^X can be computed from X_t and X_{t-1} by solving for ε_t^X in the discretized equation for X_t.

5.1 Filtering results

We run a Heston simulation for 5 years ($T = 1260$ iterations) with simulation parameters

$$\rho = -0.8,\ \kappa = 4,\ X_0 = \theta = 0.035,\ \xi = 0.008,$$

which are within the range satisfying $2\theta\kappa > \xi$ needed to ensure that vol is positive. Then we apply the bootstrap filter with $N = 2000$ particles and Heston parameters used in the simulation to validate its implementation; for illustration, we also include some results obtained by mis-specifying the parameters in the figure below (simulated returns omitted for the sake of space).



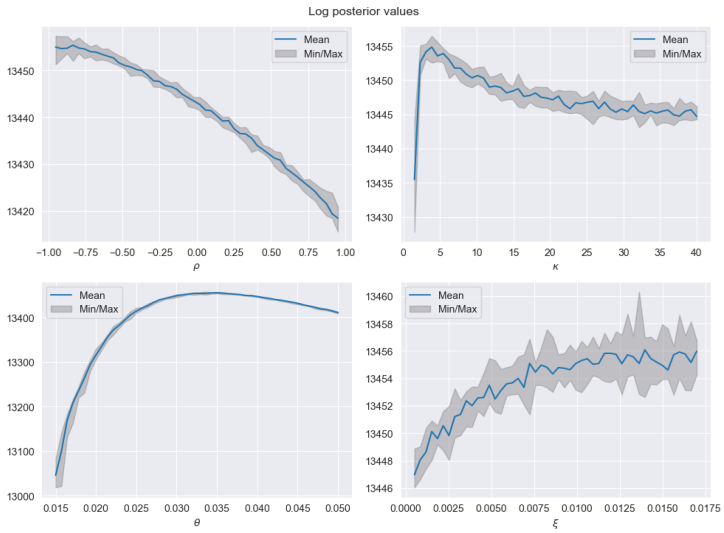
5.2 PMMH Results

To run PMMH, we specify priors centered at the correct parameters but which are essentially flat, and thus allow the likelihood $\widehat{\mathcal{L}}(\Theta)$ to dominate the posterior:

$$p(\rho) \sim \text{Unif}(-1, 1),\ p(\kappa) \sim N(4, 100),\ p(\theta) \sim N(0.035, 10),\ p(\xi) \sim N(0.008, 10).$$

We then obtain estimates for the log posterior as a function of the each of the parameters, the others being held at the simulation values, using 20 runs of the bootstrap filter per value per parameter.

Observing the scale of the y-axis in the plots, we see that the posterior is comparatively much flatter as a function of ξ (and to some degree of κ as well), and its simulation parameter is far from the mode for its posterior.

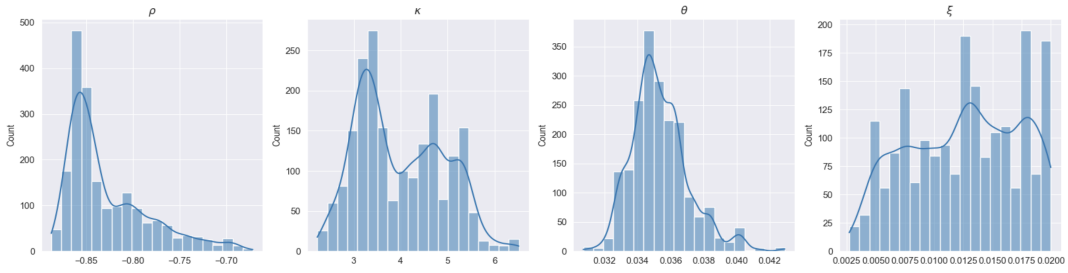


Since the mode for ξ is significantly higher than the simulation parameter, its prior's support was artificially capped at 0.02 to ensure that $2\kappa\theta > \xi$. We initialize PMMH with somewhat far off starting values $\rho_0 = 0$, $\theta_0 = 0.08$, $\kappa_0 = 2$, $\xi_0 = 0.005$ and use Gaussian proposals (in appropriate coordinates changed to ensure the parameters remained in valid ranges) with variance tuned to an acceptance rate ≈ 0.25.

In roughly 3000 iterations the chain found the mode of the posteriors, and discarding these as a burn-in period, we obtain 2000 samples from the posterior and compute statistics for the samples obtained, displayed in the table below along with the samples' histogram. As expected from the strength of concavity of the posteriors, the confidence interval is much tighter for θ and ρ than for κ and ξ. All means were fairly close to simulated values, except for ξ, although it still fell in the 90% confidence interval.

One further interesting observation is that the collected posterior ξ and κ samples have somewhat high correlation ≈ 0.45, indicating that the effects on the model likelihood of higher vol of vol and mean reversion rate cancel each other out some degree.

	Mean	5%	95%	Simul.
ρ	-0.83	-0.87	-0.73	-0.80
κ	3.99	2.71	5.44	4.00
θ	0.035	0.033	0.038	0.035
ξ	0.012	0.004	0.019	0.008



Remark 5.1. An attempt at sampling with a naive RWMH in the full 1264-dimensional state space $(\Theta, X_{0:T})$ failed to find the modes of the parameters and to fit to the simulated volatility even with 5×10^7 iterations, illustrating the curse of dimensionality.

6 Discussion and Bibliography

The material in this project is largely based on the seminal work of Andrieu, Doucet and Holenstein [2], which systematically introduced Particle Markov chain Monte Carlo methods as presented here, and which also contains a numerical study of PMMH applied to a different stochastic volatility model. Our focus was to present the subject concisely, while providing a convincing and readable proof of Theorem 3.2 accessible to a reader familiar with the lecture notes [12, Chapter 9] for the course on Monte Carlo simulation taught at the University of Chicago.

The proof of Theorem 3.2 contains two core parts. The first is the idea to use an unbiased estimator of the likelihood within the marginal Metropolis-Hastings sampler. Such algorithms predate PMCMC and are known in the literature as *pseudo-marginal Metropolis Hastings samplers.* PMCMC algorithms are a specific instance of these algorithms which use SMC to obtain the unbiased estimator. The introduction of pseudo-marginal Metropolis Hastings traces back to [1], a work which was itself motivated by the method introduced [3] in the context of genealogical inference in genetics. The second part of Theorem 3.2 is the proof of unbiasedness of the likelihood estimator derived from sequential Monte Carlo (Proposition 3.1). This proof is often presented within the framework of the Feynman-Kac formalism, such as in [6] and [11]. Due to the theoretical importance of the proposition for all of PMCMC methods, and since I could not find a reference that is both self-contained and thorough, I dedicated a significant part of this project to its proof.

The last algorithm presented, the particle Gibbs sampler, was also introduced in [2]. Its version with ancestor resampling (PGAS), briefly mentioned in Remark 4.1, was first proposed in [10]. PGAS is a significant improvement over particle Gibbs which computationally requires a very minor modification of the original idea. Unfortunately, we did not have the space to include it in our numerical simulations, nor to give a proof of its exactness analogous to Theorem 3.2.

A comprehensive analysis of the performance of PMCMC algorithms applied to stochastic volatility models can be found in the thesis [13], where SMC algorithms beyond the bootstrap filter, volatility models beyond the Heston model, and other PMCMC methods such as the aforementioned PGAS are analyzed. The book [8] also contains a discussion of PMMH applied to the Heston model. Still within the context of stochastic volatility models, we remark that if one is only interested in maximum likelihood estimation, [4] presents methods combining sequential Monte Carlo ideas with the EM-algorithm, based on the general method first proposed in [9].

Finally, we mention that a natural extension of PMCMC methods is to apply a particle filter (SMC) not only to sample from $X_{1:T}$, as discussed in this project, but also to use SMC to improve the random-walk MCMC sampler for θ. This nested SMC algorithm leads to the class of samplers known as SMC^2, introduced in [5] and [7].

References

[1] C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient monte carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.

[2] C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(3):269–342, 2010.

[3] M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160, 2003.

[4] K. Chen and S.-H. Poon. Variance swap premium under stochastic volatility and self-exciting jumps. *Available at SSRN 2200172*, 2013.

[5] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. Smc2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(3):397–426, 2013.

[6] A. Doucet, N. De Freitas, and N. Gordon. An introduction to sequential monte carlo methods. *Sequential Monte Carlo methods in practice*, pages 3–14, 2001.

[7] A. Fulop and J. Li. Efficient learning via simulation: A marginalized resample-move approach. *Journal of Econometrics*, 176(2):146–161, 2013.

[8] D. Hainaut. Continuous time processes for finance. *Switching, self-exciting, fractional and other recent dynamics. Bocconi & Springer Series*, 2022.

[9] M. Hürzeler and H. R. Künsch. Monte carlo approximations for general state-space models. *Journal of Computational and graphical Statistics*, 7(2):175–193, 1998.

[10] F. Lindsten, M. I. Jordan, and T. B. Schön. Particle gibbs with ancestor sampling. *The Journal of Machine Learning Research*, 15(1):2145–2184, 2014.

[11] P. Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Springer, 2004.

[12] D. Sanz-Alonso and O. Al-Ghattas. A first course in monte carlo methods. *arXiv preprint arXiv:2405.16359*, 2024.

[13] H. Yang. Markov chain monte carlo estimation of stochastic volatility models with finite and infinite activity lévy jumps: Evidence for efficient models and algorithms. 2015.