MTH6113 Mathematical Tools for Asset Management

Stochastic Models for Stock Prices

Dr. Melania Nica

Plan

- Geometric Brownian Motion for representing stock prices;
- Learn how to simulate stock prices from a GBM/lognormal model;
- How good is the log-normal model?
- Compare log-normal model to market data
- Stylized facts
- Better models
  - Stochastic volatility models
  - An autoregressive model

EMH

- Competition among the participants ensures that new information regarding securities is rapidly absorbed and reflected in prices.
- If security prices reflect all available information, the market is said to be efficient.
- Testing Weak form Efficiency: Random Model of Stock Prices

Brownian Motion

- **Brownian motion** is a random walk occurring in continuous time
  - *with movements that are continuous rather than discrete*
- Brownian motion is traditionally regarded as discovered by the botanist Robert Brown in 1827.
  - study of pollen particles floating in water under the microscope: pollen grains executing a random motion.
- Louis Bachelier in 1900 in his PhD thesis "The theory of speculation" used Brownian Motion to analyse the movements of the Paris stock exchange index.

Standard Brownian Motion

Definition

Standard Brownian Motion, SBM, is a stochastic process $\{B_t : t \geq 0\}$, with state space $S = \mathbb{R}$ (set of real numbers) and the following defining properties:

- Definition

  1. $B_0 = 0$
  2. Independent increments: $B_t - B_s$ is independent of $\{B_r : r \leq s\}$, where $s < t$
  3. Stationary increments: Distribution of $B_t - B_s$ depends only on $(t - s)$, where $s < t$;the change in the value of the process over any two non- overlapping periods are statistically independent
  4. Gaussian increments: $B_t - B_s \sim N(0, t - s)$
  5. Continuity: B_t has continuous sample paths

Brownian Motion

Definition

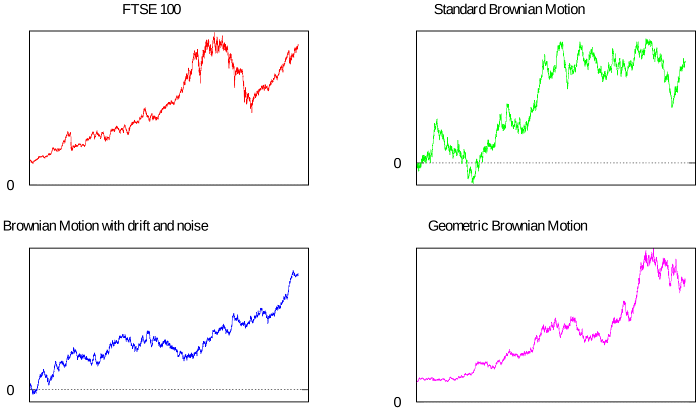
Brownian Motion, BM, is a stochastic process W_t, with state space $S = R$ (set of real numbers) and the following defining properties:

1. Independent increments: $W_t - W_s$ is independent of $\{W_r : r \leq s\}$, where $s < t$.
2. Stationary increments: Distribution of $W_t - W_s$ depends only on $(t - s)$, where $s < t$.
3. Gaussian increments: $W_t - W_s \sim N(\mu(t-s), \sigma^2(t-s))$.
4. Continuity: W_t has continuous sample paths.

Relationship between SBM and BM

- W_t (BM) can be obtained from B_t (SBM) by $W_t = W_0 + \mu t + \sigma B_t$
- μ - drift parameter and σ- volatility
- SBM can be obtained from BM by setting $\mu = 0$, $\sigma = 1$ and $W_0 = 0$.
- A **Geometric Brownian Motion** (GBM) is $S_t = \exp\left(W_t\right) = S_0 \exp\left(\mu t + \sigma B_t\right)$

Modelling Stock Prices



Geometric Brownian Motion (GBM) revisited

Consider the stock S_t with the stochastic differential equation:

$$dS_t = \alpha S_t dt + \sigma S_t dB_t$$

I want to find and expression for S_t
Standard Brownian Motion is nowhere differentiable despite the fact that it is continuous everywhere

1. - SBM is not a smooth function
   - Can I use stochastic calculus to find an explicit formula for S_t?

Ito's Lemma

Ito's Lemma Let X_t be a stochastic process satisfying $dX_t = Y_t dB_t + Z_t dt$ and let $f(t, X_t)$ be a real-valued function, twice partially differentiable with respect to x and once with respect to t. Then $f(t, X_t)$ is also a stochastic process and is given by:

$$df(t, X_t) = \frac{\partial f}{\partial X_t} Y_t dB_t + \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial X_t} Z_t + \frac{1}{2} \frac{\partial^2 f}{\partial X_t^2} Y_t^2 \right] dt.$$

Geometric Brownian Motion (GBM) revisited

Some intuition:

$$\frac{1}{S_t} dS_t = \alpha dt + \sigma dB_t$$

If we were dealing with an ordinary integral integration would lead to:

$$\ln\left(\frac{S_t}{S_0}\right) = \alpha t + \sigma B_t$$

So

$$S_t = S_0 \exp\left(\alpha t + \sigma B_t\right)$$

Applying Ito's lemma to $f\left(t, S_t\right) = \ln S_t$:

Geometric Brownian Motion (GBM) revisited

Let $Y_t = \sigma S_t$ and $Z_t = \alpha S_t$

$$\begin{aligned} d\ln S_t &= \frac{1}{S_t}\sigma S_t dB_t + \left[0 + \frac{1}{S_t}\alpha S_t + \frac{1}{2}\left(-\frac{1}{S_t^2}\right)\sigma^2 S_t^2\right] dt \\ &= \left(\alpha - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t \end{aligned}$$

Geometric Brownian Motion (GBM) revisited

Thus:

$$\begin{aligned} \ln S_t &= \ln S_0 + \left(\alpha - \frac{1}{2}\sigma^2\right) t + \sigma B_t \\ S_t &= S_0 \exp\left[\left(\alpha - \frac{1}{2}\sigma^2\right) t + \sigma B_t\right] \end{aligned}$$

Earlier we defined GBM as $S_t = S_0 \exp\left(\mu t + \sigma B_t\right)$, thus:

- S_t Geometric Brownian Motion with drift parameter $\mu = \alpha - \frac{1}{2}\sigma^2$ and **volatility** σ.

Another name for GBM: For $T > t$:
$\log(S_T) - \log(S_t) \sim N(\mu(T-t), \sigma^2(T-t))$

- μ and σ specific to the investment/security

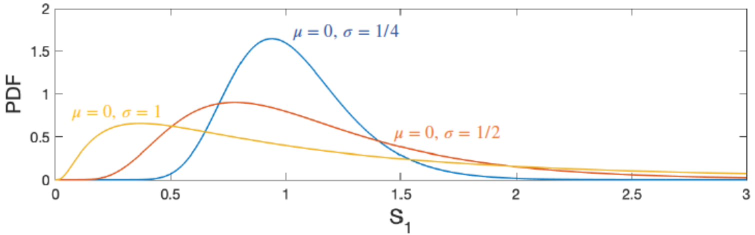
A reminder of the lognormal distribution

If $r_1 \sim N(\mu, \sigma^2)$ then $S_1 = \exp(r_1) S_0$ then
$S_1 / S_0 \sim$Lognormal$(\mu, \sigma^2) \Leftrightarrow \ln(S_1/S_0) \sim N(\mu, \sigma^2)$

$$E(S_1) = S_0 \exp(\mu + \sigma^2/2)$$

$$Var(S_1) = \exp(\sigma^2 - 1)\exp(2\mu + \sigma^2)$$

For $S_0 = 1$:



A reminder of the lognormal distribution

Example (past exam exercise)
If $\log(S_T) - \log(S_t) \equiv \log\left(\frac{S_T}{S_t}\right) \sim N(\mu(u-t), \sigma^2(u-t))$ then

$$S_T = S_t \frac{S_T}{S_t} = S_t e^{\log\left(\frac{S_T}{S_t}\right)}$$

$$E(S_T) = S_t E\left(e^{\log\left(\frac{S_T}{S_t}\right)}\right)$$

Then $E\left(e^{\log\left(\frac{S_T}{S_t}\right)}\right) = \exp(T-t)\left(\mu + \sigma^2/2\right)$ and hence
$E(S_T) = S_t E\left(e^{\log\left(\frac{S_T}{S_t}\right)}\right) = S_t \exp(T-t)\left(\mu + \sigma^2/2\right)$

Simulation of Stock Prices

Given N iid variables X_i with an assumed distribution, e.g. $\mathcal{N}(\mu, \sigma^2)$ estimate the model parameters (here μ, σ^2)

Here: X_i daily log-returns;
model parameters: mean value μ and variance σ^2.

Convenient approximation: Empirical mean & Variance:

$$\mu \approx \bar{X} = (X_1 + \ldots + X_N)/N$$

$$\sigma^2 \approx \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$$

Excel:
use `AVERAGE` & `STDEV.S`

- Parameter estimate $\hat{\theta}_N$ of parameter θ called
  - **unbiased**, iff $\mathbb{E}(\widehat{\theta}_N) = \theta$
  - **consistent**, iff $P(\widehat{\theta}_N \xrightarrow{N\to\infty} \theta) = 1$
- Mean Square Error (MSE):

$$\text{MSE} = \mathbb{E}\left((\hat{\theta}_\text{N} - \theta)^2\right) = \text{bias}(\hat{\theta}_\text{N})^2 + \text{var}(\hat{\theta}_\text{N})$$

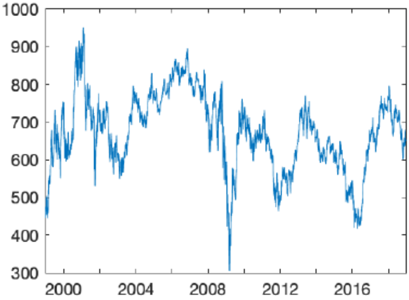
- Note: empirical mean and variance are unbiased and consistent
- Uncorrected sample variance $\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$ is biased but consistent

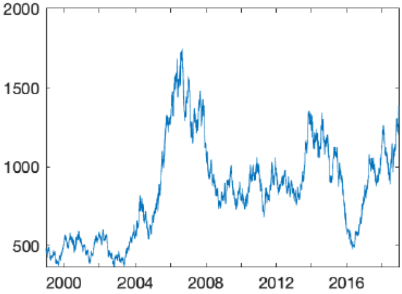
Simulation of Stock Prices

1. Download data (e.g. Yahoo Finance)
2. Read data in Excel and clean data
3. Compute daily log-returns using LN()
4. Estimate parameters μ, σ using AVERAGE() and STDEV.S()
5. Simulate X_t by samples of normal random variables; evaluate $S_t = S_0 \exp(\sum_{t=0}^{t-1} X_t)$ - see previous slides.
6. Investigate the model, e.g. plot values, returns or a histogram

Simulation of Stock Prices

Visual comparison (HSBC):

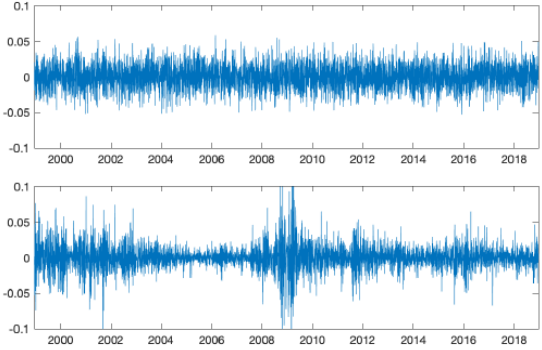




Spot the real data?

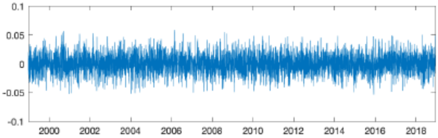
Empirical Data vs. Simulated Data

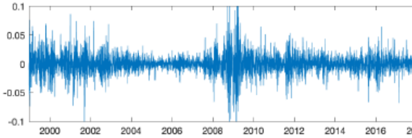
Visual comparison (HSBC log-returns):



Spot the real data?

Empirical Data vs. Simulated Data



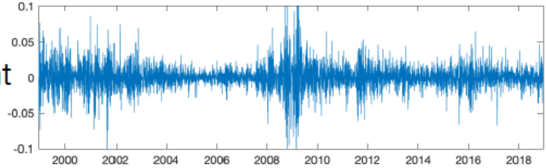


- **Spikes:** single large gains & losses
  - indicator agains a normal distribution

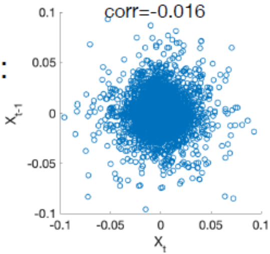
- **Clustering** of high returns in absolute values
  - indicator for dependence of subsequent returns

Clustering

Subsequent returns seem to be dependent



But we noticed no autocorrelation:



How does this work? No correlation ≠ Independence

Compare $\mathrm{Cov}(X, X^2)$, for $X \sim \mathcal{N}(0,1)$

Log-Normal Model

Consider EMH:

- Subsequent returns uncorrelated,
  but their magnitude might be correlated

- Empirical values for HSBC (20 years, i.e. 5k values)

  $\mathrm{corr}(X_t, X_{t+1}) \approx -0.0160$, no statistical significance;

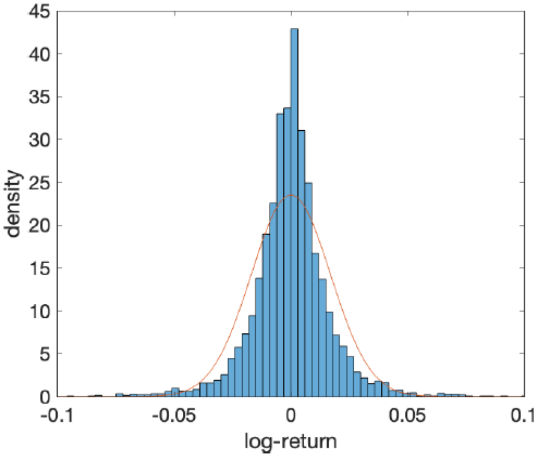
  $\mathrm{corr}(X_t^2, X_{t+1}^2) \approx 0.1661$, statistically significant

  (Note: In lognormal model X_t iid $\Rightarrow \mathrm{corr}(X_t^2, X_{t-1}^2) = 0$)

- Magnitude of returns: standard deviation σ (volatility)
  Clusters known as „volatility clusters“

Sample Distribution

- Sampled log-return values for 20 years of HSBC (blue)
- Fitted normal distribution (red)



Stylised facts

Let's collect these stylised facts:

- No linear autocorrelation: $\text{corr}(R_t, R_{t+1}) \approx 0$; (confirms weak form of EMH)
- Volatility clustering: $\text{corr}(R_t^2, R_{t+1}^2) > 0$. We can observe periods of large volatility and of small volatility;
- Heavy tails: High losses and gains much more likely than for normally distributed random variables.

Stochastic Volatility Models

Observation: Time-dependent volatility;

not available in Lognormal model

Approach $X_t \sim \mathcal{N}(\mu, \sigma_t^2)$ (no longer iid),

with stochastic process σ_t

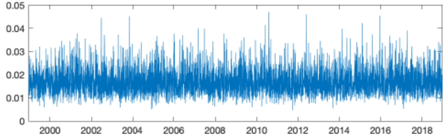
$$X_t = \mu + \sigma_t Z_t, \quad Z_t \sim \mathcal{N}(0,1)$$

With σ_t and Z_t independent

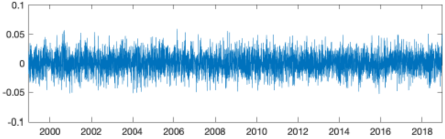
First Idea - Log-normal Volatility

Using $\sigma_t \sim \text{Lognormal}(\mu_{\text{vol}}, \sigma_{\text{vol}}))$

Volatility (σ_t)



Log-returns:



Observation: still no volatility clustering with iid volatility

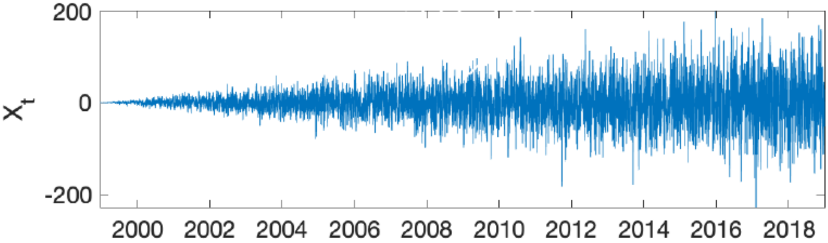
Autoregressive Volatility

- σ_t shall correlate with σ_{t-1} , e.g. with $\alpha, \nu > 0$, $\epsilon_t \sim \mathcal{N}(0,1)$ iid $(\sigma_t)_{t\in\mathbb{N}}$ as

$$\sigma_t = \alpha + \sigma_{t-1} + \nu\epsilon_t, \quad \sigma_0 = \alpha$$

= innovation + old value + noise

Resulting log-return:



What went wrong?

- Increment of σ_t: $\sigma_t - \sigma_{t-1} = \alpha + \nu\epsilon_t \sim \mathcal{N}(\alpha, \nu^2)$ iid

- What does this mean for σ_t?

$$\sigma_t \sim \mathcal{N}\left(\alpha(t+1), \nu^2 t\right)$$

We need a stationary process

i.e. σ_t has the same distribution for each t

Stationary AR(1) Model

AR(1): autoregressive with dependency on one past value

$$(\sigma_t)_{t\in\mathbb{Z}}: \sigma_t = \alpha + \beta\,\sigma_{t-1} + \nu\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1) \text{ iid}$$

is weakly stationary for $|\beta| < 1.$

- Expectation value: $\mathbb{E}(\sigma_t) = \frac{1}{1-\beta}\alpha$
- Variance: $\mathrm{Var}(\sigma_t) = \frac{1}{1-\beta^2}\nu^2$
- Autocorrelation: $\mathrm{corr}(\sigma_t, \sigma_{t-1}) = \beta$

See next slides

Stationary AR(1) Model

we have $\sigma_t = \alpha + \beta \sigma_{t-1} + \nu \epsilon_t$

$$\implies \mathbb{E}(\sigma_t) = \alpha + \beta \mathbb{E}(\sigma_{t-1}) + \nu \mathbb{E}(\varepsilon_t)$$

$= \mu$ $= \mu$ $= 0$

$$\implies \mu = \frac{\alpha}{1-\beta} \quad (\text{and } \alpha = \mu(1-\beta))$$

and $\mathrm{Var}(\sigma_t) = \mathrm{Var}(\alpha + \beta\sigma_{t-1} + \nu\varepsilon_t) = \beta^2 \mathrm{Var}(\sigma_{t-1}) + \nu^2 \mathrm{Var}(\varepsilon_t)$

$= \sigma^2$ $= \sigma^2$ $= 1$

$$\implies \sigma^2 = \frac{\nu^2}{1-\beta^2}$$

Stationary AR(1) Model

$\sigma_t = \alpha + \beta\,\sigma_{t-1} + \nu\epsilon_t$ **How to compute the autocorrelation?**

$$\mathrm{Cov}(\sigma_t, \sigma_{t-1}) = \mathbb{E}[\sigma_t \sigma_{t-1}] - \mathbb{E}[\sigma_t]\mathbb{E}[\sigma_{t-1}]$$

insert eq for σ_t

$$= \mathbb{E}[(\alpha + \beta\,\sigma_{t-1} + \nu\epsilon_t)\sigma_{t-1}] - \mathbb{E}[\sigma_{t-1}]^2$$

indep. of $\sigma_{t-1}, \varepsilon_t$

$$= \alpha\mathbb{E}[\sigma_{t-1}] + \beta\mathbb{E}[\sigma_{t-1}^2] + \nu\mathbb{E}[\epsilon_t]\mathbb{E}[\sigma_{t-1}] - \mathbb{E}[\sigma_{t-1}]^2$$

$\alpha = \mathbb{E}[\sigma_{t-1}](1-\beta)$,
$\mathbb{E}[\varepsilon_t] = 0$

$$= (1 - \beta - 1))\mathbb{E}[\sigma_{t-1}]^2 + \beta\mathbb{E}[\sigma_{t-1}^2]$$

$$= \beta\left(\mathbb{E}[\sigma_{t-1}^2] - \mathbb{E}[\sigma_{t-1}]^2\right) = \beta\,\mathrm{Var}(\sigma_{t-1}) = \beta\sigma^2$$

$$\Longrightarrow \mathrm{Corr}(\sigma_t, \sigma_{t-1}) = \frac{\mathrm{Cov}(\sigma_t, \sigma_{t-1})}{\sqrt{\mathrm{Var}(\sigma_t)\mathrm{Var}(\sigma_{t-1})}} = \frac{\beta\sigma^2}{\sigma^2} = \beta$$

First Test with AR(1)

Computational approximation:

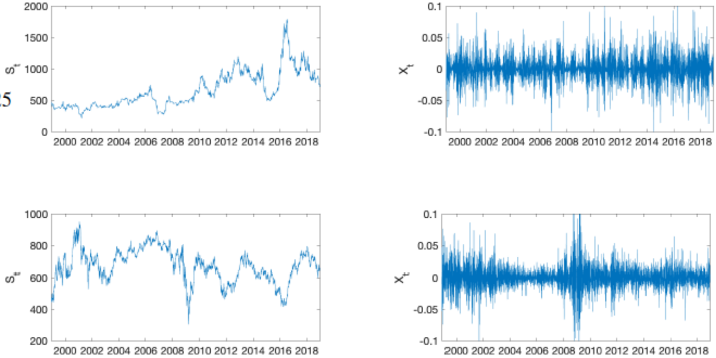
$$\sigma_0 = \frac{\alpha}{1+\beta}, \quad \sigma_t = \alpha + \beta\sigma_{t-1} + \nu\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1) \text{ iid}, \quad t \in \mathbb{N}$$

AR(1)

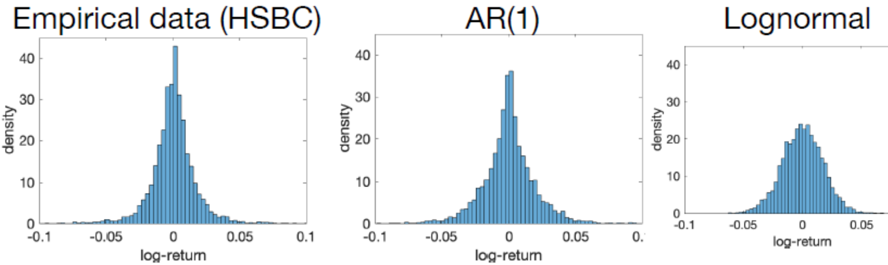
$\beta = 0.96, \nu = 0.0025$

$\mathbb{E}(\sigma_t) \approx \sqrt{\mathrm{Var}(X_t)}$

Empirical (HSBC)



- Volatility clusters can be observed
- Improvement can be seen in histogram:



Heavy tails observed in AR(1)

How to fit parameters?

- Model parameters α, β, v : $\sigma_t = \alpha + \beta\sigma_{t-1} + v\epsilon_t$
- Fit
  - expectation value $\mathbb{E}(\sigma_t)$
  - variance $\mathrm{Var}(\sigma_t)$
  - and autocorrelation $\mathrm{corr}(\sigma_t, \sigma_{t-1})$

  and compute parameters:

$$\beta = \mathrm{corr}(\sigma_t, \sigma_{t-1})$$
$$\alpha = (1-\beta)\,\mathbb{E}(\sigma_t)$$
$$v^2 = (1-\beta^2)\ \mathrm{Var}(\sigma_t)$$

How to fit parameters?

Problem: how to find empirical volatility?

Recall parameter estimation (as for lognormal model):

- E.g. for X_i iid, estimate $\sigma^2 \approx \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$
- With stochastic volatility models X_i are not independent
- σ_t different for each X_t (impossible to estimate variance with a single data point)

As a compromise, local estimates are used

Note:
Estimation of parameters in financial models is a hard task and daily returns may not be sufficient for a reliable estimate!

How to fit parameters? Naive parameter fit

1. Estimate local variance using 5 neighbouring values of the log-return:

$$\sigma_t^2 \approx 1/4 \sum_{i=t-2}^{t+2} \left(X_i - \bar{X}\right)^2$$

Use this time-series to estimate

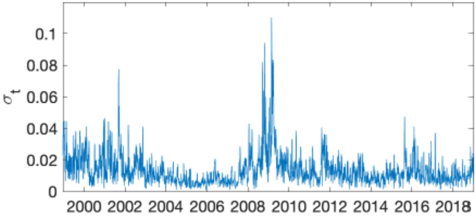
$$\mathbb{E}(\sigma_t), \ \mathrm{Var}(\sigma_t), \ \text{and} \ \mathrm{corr}(\sigma_t, \sigma_{t-1})$$

Estimated local volatility (HSBC):

$\mathbb{E}(\sigma_t) \approx 0.0138$

$\mathrm{Var}(\sigma_t) \approx 1.05 \cdot 10^{-4}$

$\mathrm{corr}(\sigma_t, \sigma_{t-1}) \approx 0.9013$



Comparison with fitted parameters