Estimating Discrete Stochastic Volatility Models & Application To Risk Measures Estimation

Loïc Cantin

ENSAE (MiE)

September, 27th 2022

Why Should We Care About Volatility?

Predicting **volatility** (i.e. a measure of the intensity of fluctuation of a time series) is one of the major topics in **financial econometrics**. Indeed, its applications include **portfolio optimisation**, **risk management**, **derivatives pricing** and is an indispensable tool for the theoretical study of market mechanisms.

e.g. Portfolio Management:

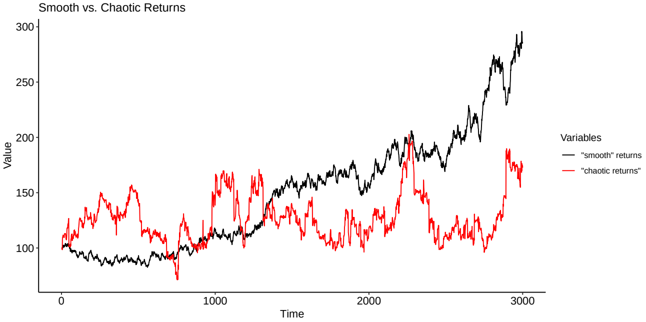
It is well known that according to the modern portfolio management theory of **Markowitz**, the optimal portfolio [1] is the one that maximizes the so-called **Sharpe ratio**:

$$\boxed{SR = \frac{r_p - r_f}{\sigma_p}}$$

- r_p: expected return of the portfolio
- r_f: return of the risk free asset
- σ_p: standard deviation of the portfolio

[1] assuming that there is no constraints on risk and expected returns

Volatility in Portfolio Management - Illustration



- $\mathbb{E}(\epsilon_{1t}) = \mathbb{E}(\epsilon_{2t}) = 0.04\%$
- $\sigma_1 \approx 0.010$, $\sigma_2 \approx 0.025$
- $\boldsymbol{CAGR_1 = 9.2\%}$, $\boldsymbol{CAGR_2 = 4.7\%}$, with $CAGR = \left(\frac{p_T}{p_1}\right)^{\left(\frac{250}{T}\right)} - 1$

Thus the objective is to "smooth the returns" by trying to maximise them while minimising the variance (i.e. volatility).

Volatility Models

There are mainly **two types of models** to study and forecast volatility:

- **Conditional** Volatility Models (e.g. GARCH models)
- **Stochastic** Volatility Models (SV)

A good volatility model should verify stylized facts such as **volatility clustering** - *"large changes tend to be followed by large changes and small changes to be followed by small changes"* [2], and **"conditional independence"** of the log-returns [3].

[2] Mandelbrot, 1963

[3] returns are not assumed independent but only independent conditionally on $\mathcal{F}_{t-1}$

GARCH vs. SV Model

Volatility - Conditional vs. Stochastic

$$\begin{cases} \textbf{GARCH} & h_t(\boldsymbol{\theta_G}) = \omega + \alpha\, \epsilon_{t-1}^2 + \beta\, h_{t-1}(\boldsymbol{\theta_G}) \in \mathcal{F}_{t-1} \\ \textbf{SV} & h_t(\boldsymbol{\theta_{SV}}) = \omega + \beta\, h_{t-1}(\boldsymbol{\theta_{SV}}) + \sigma\, v_t \in \mathcal{F}_t \end{cases}$$

- with $v_t \sim \mathcal{N}(0,1)$
- $\boldsymbol{\theta}_G = (\omega, \alpha, \beta)'$, $\boldsymbol{\theta}_{SV} = (\omega, \beta, \sigma)'$
- $\sqrt{h_t}(\boldsymbol{\theta_0})$ refers to the volatility

Contrary to the **conditional volatility**, the **stochastic volatility** is composed of an additional stochastic process $\boldsymbol{v_t} \notin \mathcal{F}_{t-1}$.

SV Model

Discrete Stochastic Volatility (1)

$$(1)\begin{cases} \epsilon_t & = \sqrt{h_t}\eta_t \\ log(h_t) & = \omega + \beta log(h_{t-1}) + \sigma v_t \end{cases}$$

- with $\eta_t \sim \mathcal{N}(0,1)$ and $v_t \sim \mathcal{N}(0,1)$

which can be re-written:

Discrete Stochastic Volatility (2)

$$(2)\begin{cases} \epsilon_t & = exp\left(\frac{1}{2}\alpha_t\right)\eta_t \\ \alpha_t & = \omega + \beta\alpha_{t-1} + \alpha v_t \end{cases}$$

- with $\eta_t \sim \mathcal{N}(0,1)$ and $v_t \sim \mathcal{N}(0,1)$
- under this formulation, volatility is represented by $\sqrt{exp(\alpha_t)}$

Comparison of the GARCH vs. SV model

GARCH		SV	
Pros	**Cons**	**Pros**	**Cons**
- Satisfy Empirical Stylized Facts	- No direct link with continuous-time theory	- Can be considered as a **discretization of a coninuous process**	- **Difficult to estimate**
- Easy to Estimate	- No natural economic interpretation ("mixture of distributions hypothesis") [4]	- Enables an economic interpretation	
		- Is more flexible than the GARCH model	

[4] states that returns are driven by a mixture of two random variables; an independent noise term and a stochastic process representing the inflow of new information

Link with continuous-time models

A popular continuous-time model of stock returns is as follows:

$$\begin{cases} dlog(S_t) & = \mu dt + \sqrt{h_t} dW_{1t} \\ dlog(h_t) & = \{\omega + (\beta - 1) log(h_t)\} dt + \sigma dW_{2t} \end{cases}$$

- (W_{1t}) and (W_{2t}) are two independent Brownian motions
- the log-volatility $dlog(h_t)$ follows an Ornstein-Uhlenbeck process
- S_t is the stock price and μ the drift of the process

We can discretize [5] it to get our "canonical" SV model [6]:

$$\begin{cases} r_t := log(S_t) - log(S_{t-1}) = log\left(\frac{S_t}{S_{t-1}}\right) = \mu + \sqrt{h_t}\eta_t \\ h_t = \omega + \beta h_{t-1} + \sigma \eta_t \end{cases}$$

[5] using Euler discretization method

[6] with $\epsilon_t := r_t - \mu$

Introduction to the Estimation methods (1)

To understand the problems involved in estimating the stochastic volatility model, it may be appropriate to compare it with the **QML estimation** of a GARCH [7] model. It has been proved that:

QML Estimator of GARCH(p,q) process

$$\hat{\theta}_T = argmax_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{n} \left(log(h_t(\theta)) + \frac{\epsilon_t^2}{h_t(\theta)} \right)$$
$$= argmax_{\theta \in \Theta} log(\mathcal{L}_T)(\theta)$$

- With $h_t(\theta) = \omega + \sum_{i=1}^{p} \alpha_i \, \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \, h_{t-j}(\theta)$
- $\theta_0 = (\omega, \alpha, \beta)'$, $\alpha = (\alpha_1, ..., \alpha_p)$, $\beta = (\beta_1, ..., \beta_q)$

is a CAN estimator of θ_0.

[7] same notations as in the previous definition

Introduction to the Estimation methods (2)

The estimator has the following asymptotic property:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \sim \mathcal{N}(\mathbf{0}, (\kappa_n - 1)\boldsymbol{J}^{-1})$$

- $\boldsymbol{J} := \mathbb{E}\left[\frac{\delta^2 l_t(\boldsymbol{\theta}_0)}{\delta\boldsymbol{\theta}\delta\boldsymbol{\theta}'}\right] = \mathbb{E}\left[\frac{1}{h_t^2(\boldsymbol{\theta}_0)}\frac{\delta h_t(\boldsymbol{\theta}_0)}{\delta\boldsymbol{\theta}}\frac{\delta h_t(\boldsymbol{\theta}_0)}{\delta\boldsymbol{\theta}'}\right]$.
- $\boldsymbol{l}_t = \boldsymbol{l}_t(\boldsymbol{\theta}) = h_t(\boldsymbol{\theta}) + \frac{\epsilon_t^2}{h_t(\boldsymbol{\theta})}$; $\kappa_n = \mathbb{E}\left[\eta_t^4\right]$

However, QML approach is not applicable as is to the SV framework:

Likelihood of the SV Model

$$\mathcal{L}(\theta, \underline{\epsilon}_T) \propto \int f(\underline{\epsilon}_T | \underline{h}_T; \boldsymbol{\theta}) f(\underline{h}_T | \boldsymbol{\theta}) d\underline{h}_T$$

- $\boldsymbol{\theta} = (\omega, \beta, \sigma)$; $\underline{\epsilon}_T = (\epsilon_1, ..., \epsilon_T)$
- $\underline{h}_T = (h_1, ..., h_T)$

Here we have a **T-dimensional integral** as $h_t \notin \mathcal{F}_t$, $\forall t \in (1, \dots T)$, which was the case for those of the GARCH model.

Linearization of the Model

Linearized Model

$$\begin{aligned} y_t := log(\epsilon_t^2) &= log(h_t) + log(\eta_t^2) \\ &= log(h_t) + \mu_Z + u_t \end{aligned}$$

- with $\mu_Z := \mathbb{E}\left[log(\eta_t^2)\right]$

Relying on this linearization we can write the following state-space model in order to implement the **Kalman filter**, which gives estimate $\in \mathcal{F}_{t-1}$ and enables us to apply the **QML approach**:

- **A.1** $log(\eta_t^2)$ follows a Gaussian distribution (not the case in general but we assume it to be close enough).

State-Space Model

$$\begin{cases} y_t := log(\epsilon_t^2) = log(h_t) + \mu_Z + u_t \\ log(h_t) = \omega + \beta log(h_{t-1}) \sigma v_t \end{cases}$$

- with u_t and $v_t \sim \mathcal{N}(0,1)$

Kalman Filter Algorithm

A few notations:

- $\alpha_{t|t-1} = \mathbb{E}\left(log(h_t)|\epsilon_1^2, ..., \epsilon_{t-1}^2\right)$
- $P_{t|t-1} = \mathbb{V}\left(log(h_t)|\epsilon_1^2, ..., \epsilon_{t-1}^2\right)$

Algorithm

1. $\alpha_{1|0} = \beta_0 a_0 + \omega, \qquad P_{1|0} = \beta^2 P_0 + \sigma^2$
2. $F_{t-1|t-2} = P_{t-1|t-2} + \sigma_Z^2, \qquad K_t = \beta P_{t-1|t-2} F_{t-1|t-2}^{-1}$
3. $P_{t|t-1} = \beta^2 P_{t-1|t-2} - K_t^2 F_{t-1|t-2} + \sigma^2$
4. $\alpha_{t|t-1} = \beta \alpha_{t-1|t-2} + K_t(y_{t-1} - \alpha_{t-1|t-2} - \mu_Z) + \omega$

Remark: The steps 1-3 can be run apart as they are independent from the y_t.

Estimation by QML

Log-Likelihood

Once we have applied the previous algorithm we just have to maximize the log-likelihood of the new "linear" problem:

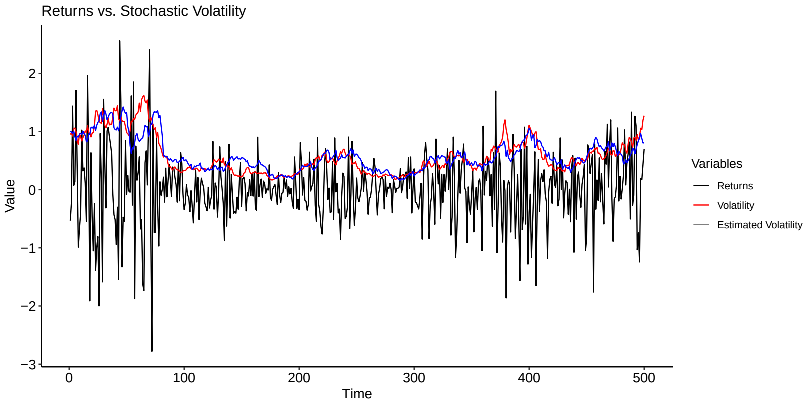
$$\widehat{\theta_T} \in argmax_{\boldsymbol{\theta} \in \Theta} log\, \mathcal{L}(\epsilon_1, ..., \epsilon_T; \boldsymbol{\theta})$$

$$\text{with } log\, \mathcal{L}(\epsilon_1, ..., \epsilon_T; \boldsymbol{\theta}) = -\frac{1}{T} log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \left(log F_{t|t-1} + \frac{\left(log(\epsilon_t^2) - \alpha_{t|t-1} - \mu_Z \right)^2}{F_{t|t-1}} \right)$$

Contrary to the estimation of a GARCH model, having access to parameter estimators is not enough to make volatility **predictions**. For this we need generally the use of a **smoother**. In this case, however, we can just take $\sqrt{exp(\alpha_{t|t-1})}$ as a $\mathcal{F}_{t-1}$-measurable approximation of the volatility $\sqrt{h_t}$.

Volatility Estimation Illustration

Below we compare the **true volatility** of simulated date and the **estimated volatility** by QML and forecasted using Kalman approximation.



Monte Carlo Experiment

We run a Monte Carlo experiment with **M=1,000 independent draws** to illustrate the consistency and the speed of convergence of the estimator. We represent for each coefficient ω, β, σ its asymptotic behavior: $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)$ for different sample size T=500, T=1,000, T=3,000, T=5,000.

Figures 3 to 5 seem to indicate the **consistency of the estimator** and a **normal asymptotic speed of convergence**, which proves the relatively good performance of this method of estimation, which we will therefore use in the application section.

Figure: $\sqrt{T}\left(\hat{\omega}_T - \omega_0\right)$

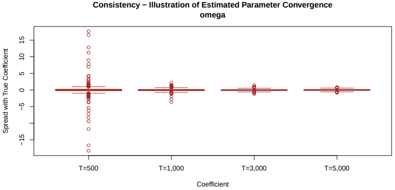


Figure: $\sqrt{T}\left(\hat{\beta}_T - \beta_0\right)$

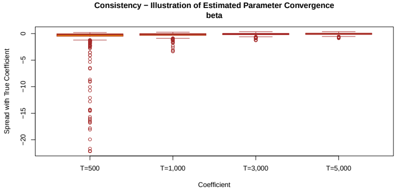
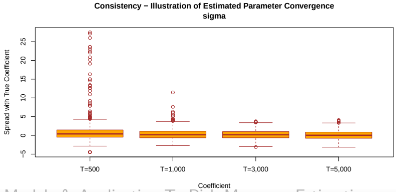


Figure: $\sqrt{T}\left(\hat{\sigma}_T - \sigma_0\right)$



GMM Approach - Principle of the Method

For this second "frequentist" approach we rely on the paper from Andersen & Sorensen (1996)

The general idea of the **GMM approach** is to compare the **empirical moments** ($M_T(\boldsymbol{\theta})$) to the **theoretical moments** $A(\boldsymbol{\theta})$ of the model based on a particular value of the parameter $\boldsymbol{\theta} \in \Theta$, and find the one which minimizes a particular criterion:

$$(\boldsymbol{M_T}(\theta) - \boldsymbol{A}(\theta))' \boldsymbol{\Lambda_T^{-1}} (\boldsymbol{M_T}(\theta) - \boldsymbol{A}(\theta))$$

with $\boldsymbol{\Lambda}_T$ a positive definite random weighting matrix.

We rely on the choice of the paper to select Q=24 different moments.

Presentation of the Moments

Vector of "Analytical" Moments

$$A(\theta) := m_t(\theta) = (m_{1t}(\theta), ..., m_{Qt}(\theta))' \text{ with } Q = 24$$

- $m_{1t} := \mathbb{E}\left(|\epsilon_t|\right) = \left(\frac{1}{\pi}\right)^{0.5}\mathbb{E}\left(\sqrt{h_t}\right)$
- $m_{2t} := \mathbb{E}\left(\epsilon_t^2\right) = \mathbb{E}\left(h_t\right)$
- $m_{3t} := \mathbb{E}\left(|\epsilon_t^3|\right) = 2\sqrt{\frac{2}{\pi}}\mathbb{E}(h_t^{3/2})$
- $m_{4t} := \mathbb{E}\left(\epsilon_t^4\right) = 3\mathbb{E}(h_t^2)$
- $m_{j+4,t} := \mathbb{E}\left(|\epsilon_t \epsilon_{t-j}|\right) = \frac{2}{\pi}\mathbb{E}(\sqrt{h_t}\sqrt{h_{t-j}})$ for $j \in (1, ..., 10)$
- $m_{j+14,t} := \mathbb{E}\left(\epsilon_t^2 \epsilon_{t-j}^2\right) = \mathbb{E}(h_t h_{t-j})$ for $j \in (1, ..., 10)$

Empirical Moments

$$M_T(\theta) = (M_{1T}(\theta), ..., M_{QT}(\theta))$$

- $M_{it}(\theta) = \sum_{t=j+1}^{T} \frac{m_{it}(\theta)}{T-j}$ for $i \in (1, \ldots, Q)$ and j is the maximum lag between the variables defining the sample moments.

Moments

In order to compute the theoretical value of the moments we can rely on the following formulae:

$$\mathbb{E}\left(\sqrt{h_t}^r\right) = exp\left(\frac{r\mu}{2} + \frac{r^2\sigma_h^2}{8}\right) \text{ for j a positive integer and r,s positive constants}$$

$$\mathbb{E}\left(\sqrt{h_t}^r\sqrt{h_t}^s\right) = \mathbb{E}\left(\sqrt{h_t}^r\sqrt{h_t}^s exp(\frac{rs\beta^j\sigma_h^2}{4}\right)$$

$$\mu = \frac{\omega}{1-\beta}$$

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

The idea is then to minimize the following quantity and find the optimal parameter $\hat{\theta}_T$ doing so:

$$\boxed{\hat{\theta}_T = argmin_{\theta\in\Theta}(M_T(\theta) - A(\theta))'\Lambda_T^{-1}(M_T(\theta) - A(\theta))}$$

Weighting Matrix

Let's give a few details about the method of selection for the weighting matrix. The optimal $\mathbf{\Lambda}$ is given by:

$$\mathbf{\Lambda} = lim_{T\to\infty}\mathbb{E}\left(\sum_{t,\tau}^{T}\frac{(m_t - A(\boldsymbol{\theta}_\mathbf{0}))\,(m_\tau - A(\boldsymbol{\theta}_\mathbf{0}))'}{T}\right)$$

which can be approximated by:

$$\sum_{j=-T+1}^{T} k(j)\hat{\Gamma}_T(j)$$

- with k(j) weights that may become 0 for $|j| > L_T$ (we can select L_T=10 for instance - this is a lag truncation parameter)
- $\hat{\Gamma}_T(j) = \frac{1}{T}\sum_{t=j+1}^{T}\left(m_t(\hat{\boldsymbol{\theta}}) - A(\hat{\boldsymbol{\theta}})\right)\left(m_{t-j}(\hat{\boldsymbol{\theta}}) - A(\hat{\boldsymbol{\theta}})\right)'$

Monte Carlo Results

For the GMM method, the convergence of the estimator to the true parameter seems to be **less efficient** than in the previous case with the QML approach.

However, we find **similar results to those of Andersen & Sorensen** (see table below). These results are obtained for T=2,000, $\boldsymbol{\theta_0} = (-0.736, 0.900, 0.363)$ and M=1,000:

Table: Consistency Result

	Our Results	Results from Andersen & Sorensen
$\lvert\omega_0 - \hat{\omega_T}\rvert$	0.108	0.103
$\lvert\beta_0 - \hat{\beta_T}\rvert$	0.016	0.013
$\lvert\sigma_0 - \hat{\sigma_T}\rvert$	0.105	0.054

Indirect Inference

The **indirect inference** method has been introduced in Gouriéroux, Monfort and Renault (1993) and is applied by Monfardini (1998) to the issue of **stochastic volatility model estimation**.

The method allows to estimate the parameter of the **true model** $M_{\boldsymbol{\theta_0}}$ using an **auxiliary model** $M^a_{\boldsymbol{\beta}}$ with which there is a certain relation of "**injectivity**" and that is **easy to estimate**. Another key condition is that we are able to simulate the **true model**.

Thanks to this injectivity relation, the knowledge of the true parameter of the auxiliary model $\boldsymbol{\theta_{aux}}$ allows to know the true parameter of interest $\boldsymbol{\theta_0}$.

It suffices to simulate the true model a large number of times and see which parameter $\boldsymbol{\theta} \in \Theta$ allows to reduce to the minimum the distance between $\boldsymbol{\theta_{aux}}$ and what will be $\hat{\boldsymbol{\theta}}_{\boldsymbol{aux}}$ in order to find our estimator $\hat{\boldsymbol{\theta}}$ of the true parameter $\boldsymbol{\theta_0}$.

Introduction of the Method

For this example we will use by convenience the second form of the *canonical* model that we have introduced previously, namely:

Discrete Stochastic Volatility

$$\begin{cases} y_t & = exp\left(\frac{1}{2}\alpha_t\right)\eta_t \\ \alpha_t & = \omega + \beta\alpha_{t-1} + \alpha v_t \end{cases}$$

- with $\eta_t \sim \mathcal{N}(0,1)$ and $v_t \sim \mathcal{N}(0,1)$
- under this formulation, volatility is represented by $\sqrt{exp(\alpha_t)}$

and let's denote $x_t := log(y_t^2)\ \forall t \in 1, ..., T$.

We can divide the approach into **two consecutive steps**:

1. Obtain an **estimator of the auxiliary parameter**
2. Obtain an approximation of the **binding function** and retrieve an estimator of the **parameter of interest**

Indirect Inference Method - First Step

In the **first step**, one can get an estimator of $\boldsymbol{\beta_0}$, denoted $\hat{\boldsymbol{\beta}}_{\boldsymbol{T}}$, from the T observations $\underline{y_T} = (y_1, ..., y_T)$ using the auxiliary criterion Q_T (e.g. the log-likelihood of M^a_β) :

$$\hat{\beta}_{\boldsymbol{T}} = \arg\max_{\beta} Q_T(\underline{y_T}, \beta).$$

The observations $\underline{y_T}$ are assumed to be generated by the initial true model M_{θ_0}. Let's define the binding function:

$$b(\theta) = \arg\max_{\beta} Q_\infty(\theta, \beta).$$

It verifies $\boldsymbol{\beta_0} = b(\boldsymbol{\theta_0})$. From here, one could define an estimator of $\boldsymbol{\theta_0}$ as the solution $\hat{\boldsymbol{\theta}}_{\boldsymbol{T}}$ of $\hat{\boldsymbol{\beta}}_{\boldsymbol{T}} = b(\hat{\boldsymbol{\theta}}_{\boldsymbol{T}})$.

Yet, the binding function may be either unknown or at least difficult to compute. The 2nd step of the indirect estimation consists in obtaining a functional estimator of $b(\cdot)$.

Indirect Inference Method - Second Step (1)

In the **second step**, one can simulate H times the initial model $M_{\boldsymbol{\theta}}$ for a given value of $\boldsymbol{\theta}$ and collect the corresponding simulated data $\{\underline{y_T^h}(\boldsymbol{\theta}) = (y_1^h, ..., y_T^h), h = 1, ..., H\}$. From this TH simulated data, one can get H estimators of $b(\boldsymbol{\theta})$, denoted $\{\hat{\boldsymbol{\beta}}_{\boldsymbol{T}}^h(\boldsymbol{\theta}), h = 1, ..., H\}$, using the same auxiliary criterion Q_T :

$$\hat{\beta}_T^h(\theta) = \arg\max_{\beta} Q_T(\underline{y_T^h}(\theta), \beta).$$

Then, one gets an estimator of $b(\boldsymbol{\theta})$, denoted $\widehat{\boldsymbol{\beta}}_{HT}(\boldsymbol{\theta})$, by averaging the H estimators $\widehat{\boldsymbol{\beta}}_T^h(\boldsymbol{\theta})$:

$$\widehat{\boldsymbol{\beta}}_{HT}(\boldsymbol{\theta}) = \frac{1}{H} \sum_{h=1}^{H} \hat{\boldsymbol{\beta}}_T^h(\boldsymbol{\theta}).$$

Indirect Inference Method - Second Step (2)

If, for each $\boldsymbol{\theta}$, $\hat{\beta}_{HT}(\boldsymbol{\theta})$ is a consistent estimator of $b(\boldsymbol{\theta})$ then $\hat{\beta}_{\boldsymbol{HT}}(\cdot)$ is a consistent functional estimator of $b(\cdot)$. In particular $\hat{\beta}_{\boldsymbol{HT}}(\boldsymbol{\theta_0})$ is a consistent estimator of $b(\boldsymbol{\theta_0}) = \beta_{\mathbf{0}}$.

Then, $\hat{\theta}_{HT}$ is defined as the solution of a minimum distance problem:

Indirect Inference Estimator

$$\widehat{\boldsymbol{\theta}_{\boldsymbol{HT}}} = argmin_{\boldsymbol{\theta}\in\Theta}[\hat{\beta}_T - \widehat{\beta_{HT}}(\boldsymbol{\theta})]'\hat{\boldsymbol{\Omega}}_T[\hat{\beta}_T - \widehat{\beta_{HT}}(\boldsymbol{\theta})]$$

- where $\hat{\boldsymbol{\Omega}}_{\boldsymbol{T}}$ is a positive definite matrix converging to a deterministic positive definite matrix $\boldsymbol{\Omega}$

Choice of the Auxiliary Model

Before Monfardini, we could find in the literature the choice of the GARCH(1,1) as an auxiliary model. This choice seems to go against the spirit of the method which **advocates an auxiliary model that is easy to estimate**.

Thus, we focus on the auxiliary model based on the **ARMA(1,1)** representation:

$$x_t = \alpha_0^* + \alpha_1^* x_{t-1} + \omega_t - \alpha_2^* \omega_{t-1};\ \omega_t \sim I.I.N(0, \nu^2).$$

This auxiliary model $M_{\boldsymbol{\alpha}}^{ARMA}$ has a 4-dimensional parameter α_0 with true value $\boldsymbol{\alpha} = (\alpha_0^*, \alpha_1^*, \alpha_2^*, \nu^2)'$, the auxiliary parameter (i.e. $\boldsymbol{\beta}$ in the general presentation of the method).

Then, we proceed as previously mentioned, for simplicity we take the identity matrix as weighting.

Results

Our application of the indirect inference method based on the ARMA auxiliary model **was not successful** during its implementation.

Indeed, although we could not conduct a Monte Carlo experiment of reasonable size to show this (the computation time of our algorithm being quickly too large), **our tests do not seem to indicate the consistency of the estimator**. We find consistency in some cases if we reduce the parameter space $\hat{\Theta} \subset \Theta$ in which our estimator $\hat{\theta}$ can live.

Possible reasons for this result could be: the **use of a bad weighting matrix** (we took the identity matrix for simplicity), a **too small sample size**, an **algorithmic difficulty** to solve the minimization problem.

Why and How to Model Risks?

Although financial institutions have been looking at risk models since the **1980s-1990s**, it was mainly the **global financial crisis** that led them to refocus closely on the issue of risk measurement.

In portfolio management, risk is often measured by **volatility** itself, but it can be interesting to look at **more extreme risks** (i.e. risks that have a lower probability of occurring $\leq \alpha$, but whose occurrence would cause large losses and even endanger the company or the system it describes).

Let's recall for the following sections that $\epsilon_t|\mathcal{F}_{t-1} \sim \mathcal{D}$ denotes the log-returns of an asset following an unknown distribution.

For simplicity, *quants* have long relied on the assumption that $\mathcal{D} = \mathcal{N}(0, h_t)$, which has been identified as one of the major causes of the magnitude of the 2008 crisis. Indeed, the **gaussian distribution has a much lower mass in the tails of distributions than what is empirically observed**.

Value at Risk

The **Value at Risk** (VaR) is the most commonly used measure to account for the risk of a financial time series:

$$VaR(\alpha) = -q_{\alpha}\left(\epsilon_t\right);\ \epsilon_t \sim \mathcal{D}$$

In order to estimate it, we will focus on what we call the **Conditional Value at Risk (CVaR)** at risk level α (conditional to the information set available at *t-1*), which is defined as follows:

$$VaR_{t-1}(\alpha) = -q_{\alpha}\left(\epsilon_t|\mathcal{F}_{t-1}\right)$$

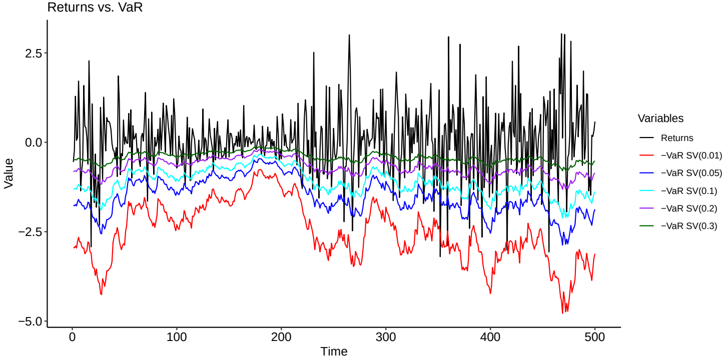
It is **minus the quantile of order α for the log-returns at date t** knowing all the past information $\mathcal{F}_{t-1}$.
Its greatest advantage is that it does not require the existence of any moment to exist, which is not the case for some other risk measures such as the (Conditional) **Expected Shortfall** for instance.

(Conditional) Value at Risk - Illustration

(Conditional) VaR Characterization

$$\mathbb{P}_{t-1}\left(\epsilon_t < -VaR_{t-1}(\alpha)\right) = \alpha$$

Illustration of VaR with multiple levels of risk revels.



Model & Estimation - Case for GARCH

GARCH(1,1) Conditional Volatility Model

$$\begin{cases} \epsilon_t = -\sqrt{h_t}(\boldsymbol{\theta_0})\,\eta_t \\ h_t(\boldsymbol{\theta_0}) = \omega + \alpha\epsilon_{t-1}^2 + \beta\, h_{t-1}(\boldsymbol{\theta_0}) \end{cases}$$

- $\sqrt{h_t(\boldsymbol{\theta_0})}$ represents the **conditional volatility** at time t ($\mathcal{F}_{t-1}$ measurable) of the series under consideration.
- η_t is the **innovation** at time t verifying $\mathbb{E}[\eta_t] = 0$, $\mathbb{V}[\eta_t] = 1$ and the sequence (η_t) is iid, but its **distribution is supposed unknown**.
- True parameter: $\theta_0 = (\omega_0, \alpha_0, \beta_0)$.
- Estimator: $\hat{\boldsymbol{\theta}}_T = (\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)$, with T the in-sample size.

VaR Estimation via GARCH Model

$$VaR_{t-1}(\alpha) = -\sqrt{h_t(\theta_0)}\,\xi_\alpha, \ \textit{with}\ \xi_\alpha\ \textit{the}\ \alpha - \textit{quantile of the}\ \eta_t$$

$$\widehat{VaR}_{t-1}(\alpha) = -\sqrt{h_t(\hat{\theta}_T)}\,\hat{\xi}_\alpha, \ \textit{with}\ \hat{\xi}_\alpha\ \textit{emp. quantile of the}\ \hat{\eta}_t$$

Adaptation to the case of SV models

However, this method cannot be applied as is to the stochastic volatility model since **volatility $\sqrt{h_t}$ is not $\mathcal{F}_{t-1}$-measurable** in this case.

However, we have seen previously when applying the QML method to the state-space model and using the Kalman filter, that we could obtain **an approximation of the volatility $\sqrt{exp(\alpha_{t|t-1})}$, which is $\mathcal{F}_{t-1}$-measurable.** We can then reconstruct an estimator of the innovations $(\hat{\eta}_1, ..., \hat{\eta_T})$ with:

$$\boxed{\hat{\eta}_t := \epsilon_t / \sqrt{exp(\alpha_{t|t-1})} \text{ for } t \in (1, ..., T)}$$

and thus apply the **two-step method** as in the case of a GARCH model. Below is an example of the VaR of the log-returns of the S&P500 at risk level, 1%, 5% and 20%, centered around the period of the "Covid crisis".

Expected Shortfall

Expected Shortfall (ES) - Definition

$$ES(\alpha) := \mathbb{E}\left[\epsilon_t | \epsilon \leq -VaR(\alpha)\right]$$

(Conditional) ES - Characterization

$$\begin{aligned} ES_{t-1}(\alpha) &= \mathbb{E}_{t-1}\left[\epsilon_t \,|\, \epsilon_t < -VaR_{t-1}(\alpha)\right] \\ &= \mathbb{E}_{t-1}\left[\epsilon_t \,|\, \eta_t < \xi_\alpha\right] \\ &= \sqrt{h_t}(\boldsymbol{\theta_0})\,\mathbb{E}\left[\eta_t \,|\, \eta_t < \xi_\alpha\right] \end{aligned}$$

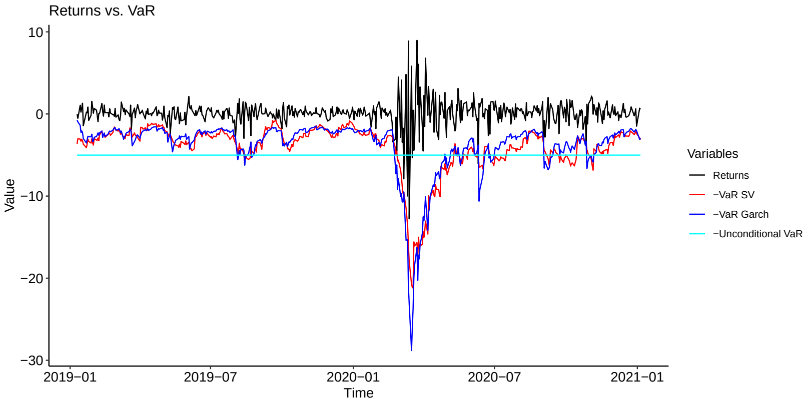
ES Empirical Estimator

$$\widehat{ES}_{t-1}(\alpha) = \frac{\sqrt{h_t}(\hat{\boldsymbol{\theta}}_T)}{\tilde{T}} \sum_{t=1}^{T} \hat{\eta}_t \, \mathbb{1}\{\hat{\eta}_t \,|\, \hat{\eta}_t < \hat{\xi}_T\}$$

- $\tilde{T} = \sum_{t=1}^{T} \mathbb{1}\{\hat{\eta}_t \,|\, \hat{\eta}_t < \hat{\xi}_T\}$

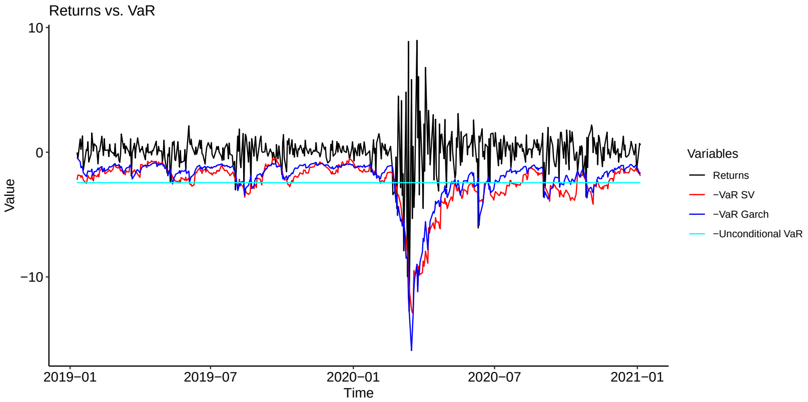
GARCH vs. SV VaR(1%)

Figure: Illustration of the (Conditional) VaR(1%) of the log-returns of the S&P 500 during the Covid crisis



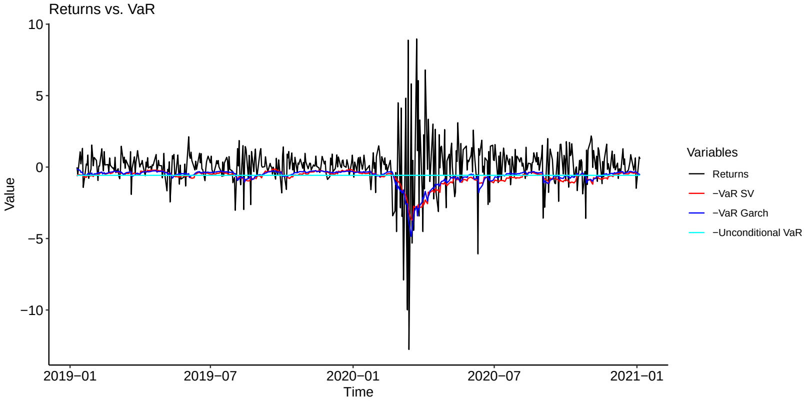
GARCH vs. SV VaR(5%)

Figure: Illustration of the (Conditional) VaR(5%) of the log-returns of the S&P 500 during the Covid crisis



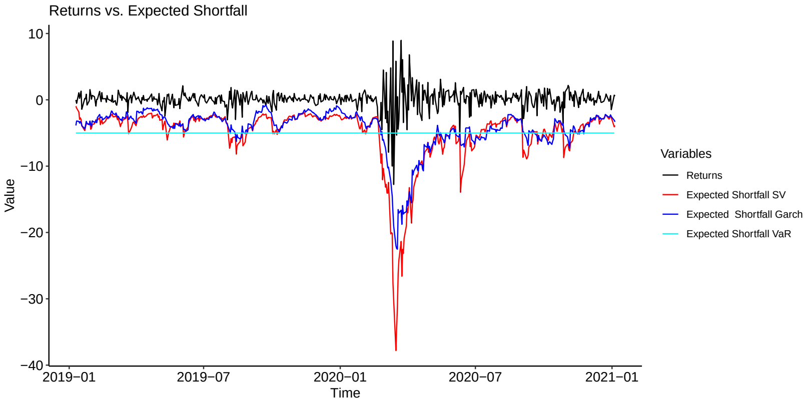
GARCH vs. SV VaR(20%)

Figure: Illustration of the (Conditional) VaR(20%) of the log-returns of the S&P 500 during the Covid crisis



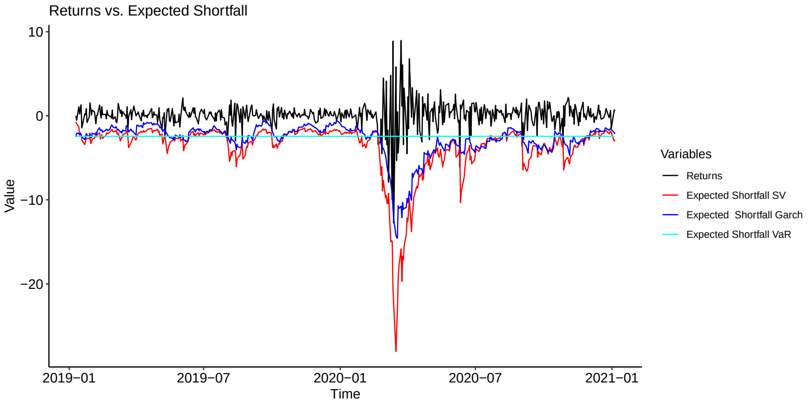
GARCH vs. SV ES(1%)

Figure: Illustration of the (Conditional) Expected Shortfall (1%) of the log-returns of the S&P 500 during the Covid crisis



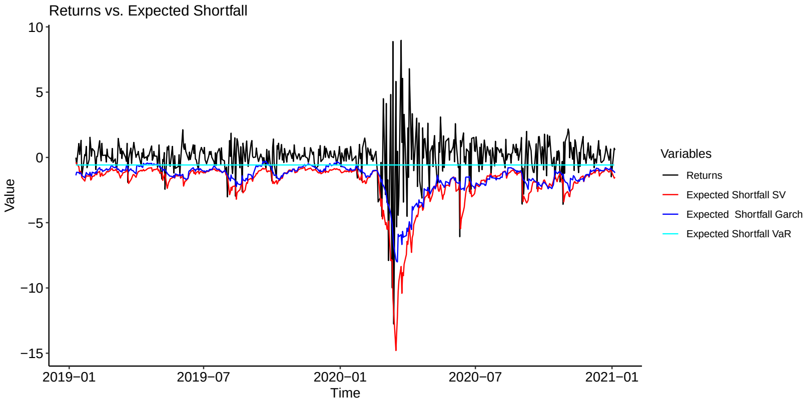
GARCH vs. SV ES(5%)

Figure: Illustration of the (Conditional) Expected Shortfall (5%) of the log-returns of the S&P 500 during the Covid crisis



GARCH vs. SV ES(20%)

Figure: Illustration of the (Conditional) Expected Shortfall (20%) of the log-returns of the S&P 500 during the Covid crisis



Quick Overview of the Existing Tests

We can divide the main existing tests into 3 categories:

- The tests based on the **number of violations** and their **random** appearance for a particular risk level.
- The tests based on **multiple risk levels** (see Multivariate Portmanteau Test from Hurlin (2007) for example.
- The "tests" based on the **distance between the VaR and the log-returns** (see what we call the "α-criterion").

In this presentation we will focus on the most traditional Kupiec, Christoffersen tests, let's define the *Hit variable*:

Hit Variable

$$Hit_t(\alpha) = \mathbb{1}\{\epsilon_t < -VaR_{t-1}(\alpha)\}$$

Selected Tests

We will focus on the Christofersen's test here and then compare the α-criteria of the two models (GARCH vs. SV)

Let's recall that these tests are not taking into account the estimation risk [8].

Christoffersen has established 3 different tests:

- The **unconditional coverage** test, testing the consistency between the frequency of violations and the theoretical value α.
- The **independence** test, testing the random appearance of the violations.
- The **conditional coverage** test, testing both.

[8] we do as if we know the true parameter

Christoffersen's Tests

Unconditional Coverage Test:

- Null hypothesis: $H_0^{UC} : \mathbb{P}[Hit_t = 1] = \alpha$
- Test Statistics: $LR_{UC} = 2log\frac{\pi_{exp}^{n_1}(1-\pi_{exp})^{n_0}}{\pi_{obs}^{n_1}(1-\pi_{obs})^{n_0}}$

Independence Test:

- $H_0^{Ind} : \mathbb{P}[Hit_t = 1|Hit_{t-1} = 0] = \mathbb{P}[Hit_t = 1|Hit_{t-1} = 1]$
- Test Statistics: $LR_{Ind} = 2log\frac{\pi_{obs}^{n_1}(1-\pi_{obs})^{n_0}}{\pi_{01}^{n_{01}}(1-\pi_{01})^{n_{00}}\,\pi_{11}^{n_{11}}(1-\pi_{11})^{n_{10}}}$

Conditional Coverage Test:

- Test Statistics: $LR_{cc} = 2log\frac{\pi_{exp}^{n_1}(1-\pi_{exp})^{n_0}}{\pi_{01}^{n_{01}}(1-\pi_{01})^{n_{00}}\,\pi_{11}^{n_{11}}(1-\pi_{11})^{n_{10}}}$
    - π_{exp} is the expected proportion of violations.
    - π_{obs} is the observed proportion of violations.
    - n_1 is the number of violations and $n_0 = n - n_1$ is the sample size.
    - n_{ij} is the number of indicator i followed by indicator j.
    - $\pi_{01} = \frac{n_{01}}{(n_{00}+n_{01})}$ and $\pi_{11} = \frac{n_{11}}{(n_{10}+n_{11})}$

α-criterion

A good model is not only evaluated on the comparison between the observed frequency of violations and the chosen risk level.

Indeed, a good model **should be as close as possible to the true distribution**, so that when volatility is low, one wants to get as close as possible to the returns.

This requirement comes from the fact that **banks calibrate their reserves on the VaR**, so they want to keep as few reserves as possible while minimising the risk of having losses that would exceed the reserves.

Definition ("α - Comparison Criterion")

$$\mathbb{E}[(1-\alpha).(\epsilon - q)^{-} + \alpha.(\epsilon - q)^{+}]$$

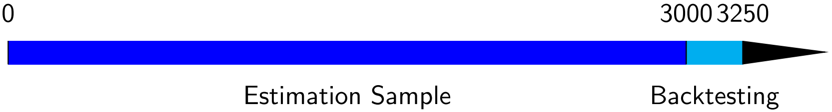
Definition ("Empirical α-Comparison Criterion")

$$T^{-1}.\sum_{t=1}^{T}(1-\alpha).(\epsilon_t - VaR_t)^{-} + \alpha(\epsilon_t - VaR_t)^{+}$$

Backtesting Methodology

We perform a backtest on **real data**. We use a dataset that is composed of the log-returns of basket of **1,489 US stocks** (that we have selected to have less than 1% missing data compared to the yearly market trading days).

Each series is divided into **two periods**: an **estimation period** (called "in-sample") which is used, to estimate the parameters of the model as well as to estimate the quantile of the innovations (here from 02/05/2009 to 01/05/2021), and a **backtesting period** (called "out-of-sample" - here from 01/06/2021 to 12/31/2021), which is used to evaluate the performance of the predictions resulting from the model.



Results (1)

Table: Backtesting Results
Frequency of Null Rejection (over 1,489 stocks)

Risk level $\alpha = 1\%$	UC	Ind.	α-criterion
SV Model	0.208	0.017	0.171
GARCH Model	0.107	0.013	0.199
Unconditional Model	0.410	0.017	0.155

Results (2)

Table: Backtesting Results
Frequency of Null Rejection (over 1,489 stocks)

Risk level $\alpha = 5\%$	UC	Ind.	α-criterion
SV Model	0.187	0.056	0.467
GARCH Model	0.105	0.044	0.486
Unconditional Model	0.349	0.083	0.464

Results (3)

Table: Backtesting Results
Frequency of Null Rejection (over 1,489 stocks)

Risk level $\alpha = 20\%$	UC	Ind.	α-criterion
SV Model	0.148	0.093	0.960
GARCH Model	0.123	0.077	0.966
Unconditional Model	0.303	0.091	0.965

Results - Conclusion

We observe that the **GARCH model is rejected less often** than the stochastic volatility model and therefore seems to perform better. This can be explained by two reasons:

1. the **GARCH model is indeed more efficient**, notably because the method chosen for the SV is less rigorous
2. the estimation risk for the stochastic volatility model may be higher

Nevertheless, the **rejection rate of the stochastic volatility model turns out to be much better than that of the unconditional volatility** and is therefore worth considering.

Moreover, the model based on stochastic volatility seems to have a **lower α-criterion than the GARCH model**, which is a positive point for a company wishing to reduce its reserves, especially during a period of calm in the markets.

Volatility Risk Premium

To illustrate the potential use of the stochastic volatility model, we can also look at the so-called **"volatility risk premium"** - i.e. the difference between the **"implied volatility"** and the **"realized volatility"** [9].

Volatility Risk Premium

$$VRP_t := ImpV_t - RV_t$$

Conditional and Stochastic Volatility Risk Premia

$$(C)VRP_t := \sqrt{h_t^C} - RV_t;\ (S)VRP_t := \sqrt{h_t^S} - RV_t$$

- where $\sqrt{h_t^C}$ and $\sqrt{h_t^S}$ refer respectively to the **conditional** and **stochastic** volatilities.

[9] We will define these two notions in the next slides

Realized Volatility vs. Conditional volatility

These premia can be seen as **market indicators of the risk aversion of the investors** as the **implied volatility is the "price" of the volatility** in the market and the other volatilities can be seen as technical prediction or historical realization in the market.

Realized Volatility

$$(\textbf{Realized Volatility})\ RV_t = \sum_{i=t-N}^{t} \epsilon_i^2$$

where N is chosen by the practitioner (e.g. 252 for realized volatility computed over the past trading year, 21 when it is computed over the last month etc.)

Realized Volatility is therefore a "lagging indicator" it overweights past events, compared to their actual impact on current volatility as it is nothing else but a rolling mean which **weights uniformly past volatilities** over a period.

Conditional or Stochastic Volatilities however put **more weights on recent events**, which enables to take into account volatility clustering and is much less lagging. Thus, we believe (C)VRP and (S)VRP are more interesting indicators to look at.

Implied volatility and VIX

The **implied volatility** is the volatility induced by the market option prices. **It is therefore by definition a forward indicator**. The **(CBOE) VIX** is the most common measure used to track the implied volatility of the market and is **based on the price of options whose underlying is the S&P 500**. The calculation of the price of a large number of options is based on the **Black-Scholes model**, whose only "real" unknown is the volatility.

Example of the Black-Scholes formula for a call option:

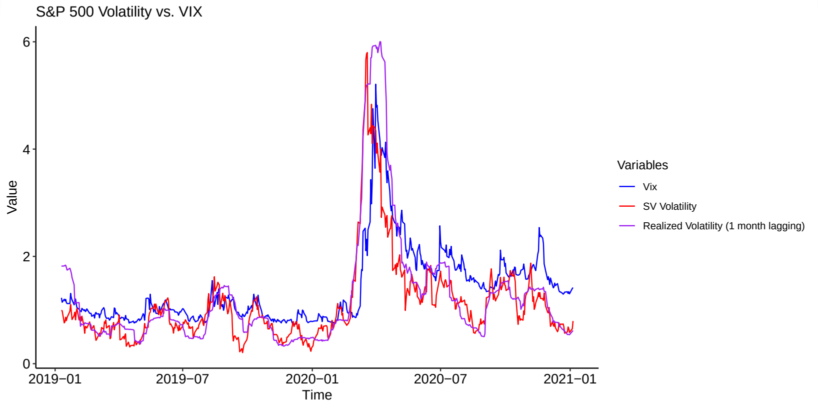
$$C(S_t, t) = N(d1)S_t - N(d_2)Ke^{-r(T-t)}$$

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right]$$

$$d_2 = d_1 - \sigma\sqrt{T-t}$$

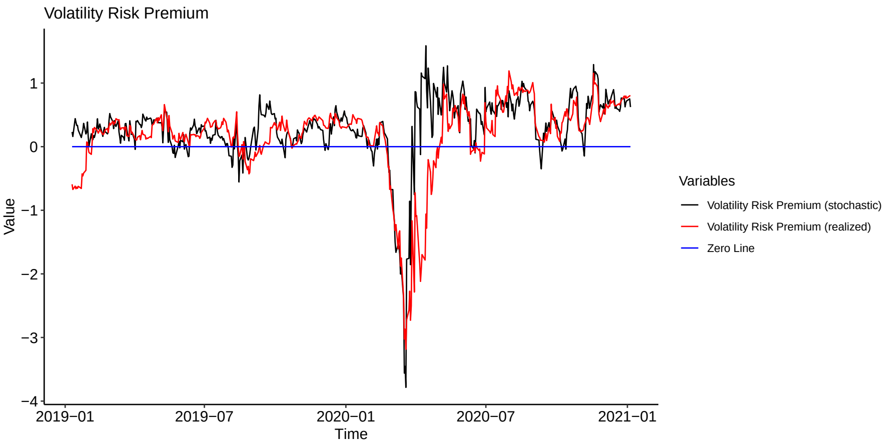
- With C the price of the call option depending on the underlying asset, whose price at date t is denoted S_t
- K is the strike price (price at which we can buy at maturity)
- σ refers to the volatility (our implied volatility of interest)
- r refers to the interest rate and T the time of the maturity
- N(.) refers to the standard normal cumulative distribution function

Comparison of Different Volatilities



We see here that **implied volatility** (which should be a "forward" indicator) **is actually lagging about a month** it is aligned on the 1-month Realized volatility, which seems to indicate that **investors are looking backward to predict coming risks and not forward**.

Volatility Risk Premia



Comments on next slide.

VRP Conclusions

We can see that, as expected, the **VRP is generally positive** because investors tend to **keep a margin of risk in excess of the actual or predicted market risk**.

However, as we saw in the previous figure, **implied volatility is lagging**, which explains a **negative VRP at the beginning of the health crisis**, due to the fact that investors took time to realize the risk of the pandemic and to effectively price this risk.

Interestingly, the **VRP also tends to increase in the period following a crisis**, which may indicate increased risk aversion on the part of investors following a crisis (i.e. after having potentially experiencing losses).

We have seen that unlike returns, which investors tend to try to anticipate, it seems that when it comes to risk, they **look more at past risks to make their decisions**.

End of the Presentation
Q&A