ARMA-GARCH-NIG Risk Model and Implementation Details in Python3

Haochen Jiang
Master in Quantitative Finance track

2022.05.03

Outline

Stony Brook University

- **Preliminaries — from quants' view**
  - **Distributions**
  - **ARMA(1,1) – GARCH(1,1) model**
  - **Backtesting**
- **Empirical Study — providing managers with some ideas**
  - **Full procedure of analysis – take IBM & VXIBM as an example**
- **Future work**
- **Reference**

Part I.

Stony Brook University

Preliminaries

— from quants' view

Gaussian Distribution

scipy.stats.norm = <scipy.stats._continuous_distns.norm_gen object>

- A normal continuous random variable.
- The *location (loc)* keyword specifies the mean. The *scale (scale)* keyword specifies the standard deviation.
- As an instance of the rv_continuous class, norm object inherits from it a collection of generic methods, and completes them with details specific for this particular distribution.
- For a real number x, the probability density function for norm is:

$$\mathbb{f}(x,\ \mu,\ \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\mathbb{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- The location-scale-based parameterization implemented in SciPy is the same as the definition above, where $loc = \mu,\ scale = \sigma$.

Reference:
https://scipy.github.io/devdocs/reference/generated/scipy.stats.norm.html

Generalized Hyperbolic Distribution

scipy.stats.genhyperbolic = <scipy.stats._continuous_distns.genhyperbolic_gen object>

- A generalized hyperbolic continuous random variable.
- The probability density function for genhyperbolic is:

$$\mathbb{f}(x,\ p,\ a,\ b) = \frac{\left(a^2-b^2\right)^{0.5p}}{\sqrt{2\pi}a^{p-0.5}\mathbb{K}_p\left(\sqrt{a^2-b^2}\right)}\mathbb{e}^{bx} \cdot \frac{\mathbb{K}_{p-0.5}\left(a\sqrt{1+x^2}\right)}{\left(\sqrt{1+x^2}\right)^{0.5-p}}$$

  for $x, p \in (-\infty, +\infty)$. $|b| < a$ if $p \geq 0$; $|b| \leq a$ if $p < 0$. $\mathbb{K}_p(\cdot)$ denotes the modified Bessel function of the second kind and order p.

- Genhyperbolic takes a as a ***shape*** parameter, b as a ***skewness*** parameter. Parameter p affects ***tail heaviness*** and allows us to navigate through different subclasses. Particularly, for $\lambda = -0.5$ we get the normal inverse Gaussian and for $\lambda = 1$ we get the hyperbolic distribution.
- As an instance of the rv_continuous class, genhyperbolic object inherits from it a collection of generic methods, and completes them with details specific for this particular distribution.

Reference:
https://scipy.github.io/devdocs/reference/generated/scipy.stats.genhyperbolic.html#scipy.stats.genhyperbolic

Generalized Hyperbolic Distribution - cont'd

- The original parameterization of the Generalized Hyperbolic Distribution is as follows:

$$\mathbb{g}(x,\lambda,\alpha,\beta,\delta,\mu) = \frac{\left(\frac{\sqrt{\alpha^2-\beta^2}}{\delta}\right)^{\lambda}}{\sqrt{2\pi}\mathbb{K}_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)} \mathbb{e}^{\beta(x-\mu)} \cdot \frac{\mathbb{K}_{\lambda-0.5}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\frac{\sqrt{\delta^2+(x-\mu)^2}}{\alpha}\right)^{0.5-\lambda}}$$

  for $x,\lambda,\mu \in (-\infty,+\infty)$. $\delta \geq 0$, $|\beta| < \alpha$ if $\lambda \geq 0$; $\delta > 0$, $|\beta| \leq \alpha$ if $\lambda < 0$. $\mathbb{K}_{\lambda}(\cdot)$ denotes the modified Bessel function of the second kind and order λ.
- The following formulas bridge the original definition shown above and the location-scale-based parameterization implemented in SciPy: $a = \alpha\delta,\ b = \beta\delta,\ p = \lambda,\ scale = \delta,\ loc = \mu.$
- For the distributions that are a special case such as Student-T, it is not recommended to rely on the implementation of genhyperbolic. To avoid potential numerical problems and for performance reasons, the methods of the specific distributions should be used.

Normal Inverse Gaussian Distribution

scipy.stats.norminvgauss = <scipy.stats._continuous_distns.norminvgauss_gen object>

- A Normal Inverse Gaussian continuous random variable.
- The probability density function for norminvgauss is:

  $$\mathbb{f}(x, a, b) = \frac{a\mathbb{K}_1\left(a\sqrt{1+x^2}\right)}{\pi\sqrt{1+x^2}}\mathbb{e}^{\sqrt{a^2-b^2}+bx}$$

  for $x \in (-\infty, +\infty)$. $|b| \leq a$ and $a > 0$. $\mathbb{K}_1(\cdot)$ is the modified Bessel function of the second kind and order 1.
- Normivgauss takes a as the ***tail heaviness*** parameter, b as the ***asymmetry*** parameter.
- As an instance of the rv_continuous class, norminvgauss object inherits from it a collection of generic methods, and completes them with details specific for this particular distribution.

Reference:

https://scipy.github.io/devdocs/reference/generated/scipy.stats.norminvgauss.html#scipy.stats.norminvgauss

Normal Inverse Gaussian Distribution - cont'd

- Another common parametrization of the distribution is given by the following expression of the pdf:

  $$\mathbb{g}(x,\alpha,\beta,\delta,\mu)=\frac{\alpha\delta\mathbb{K}_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}}\mathbb{e}^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)}$$

  for $x\in(-\infty,+\infty)$. $\delta>0$, $|\beta|\leq\alpha$. $\mathbb{K}_1(\cdot)$ denotes the modified Bessel function of the second kind and order 1.
- The following formulas bridge the original definition shown above and the location-scale-based parameterization implemented in SciPy:
  $a=\alpha\delta,\ b=\beta\delta,\ scale=\delta,\ loc=\mu.$

- A normal inverse Gaussian random variable Y with parameters a and b can be expressed as a normal mean-variance mixture:
  Y = b * V + sqrt(V) * X where X is norm(0,1) and V is invgauss(mu=1/sqrt(a**2 - b**2)). This representation is used to generate random variates.

In order to analyze the distribution of log returns, the impact of a random variable's linear transformation of on characteristic function is the key issue.

- **Characteristic functions**
  - For a real number u, the Gaussian characteristic function is:

$$\varphi(u,\ \mu,\ \sigma) = \mathbb{e}^{iu\mu-\frac{1}{2}u^2\sigma^2}$$

The location-scale-based parameterization implemented in SciPy is the same as the definition above, where $loc = \mu,\ scale = \sigma$.

  - The Generalized Hyperbolic characteristic function is:

$$\varphi(u,\lambda,\alpha,\beta,\delta,\mu) = \mathbb{e}^{iu\mu}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+iu)^2}\right)^{\frac{\lambda}{2}}\frac{\mathbb{K}_\lambda\left(\delta\sqrt{\alpha^2-(\beta+iu)^2}\right)}{\mathbb{K}_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$$

for $u,\lambda,\mu \in (-\infty,+\infty)$. $\delta \geq 0$, $|\beta| < \alpha$ if $\lambda \geq 0$; $\delta > 0$, $|\beta| \leq \alpha$ if $\lambda < 0$. $\mathbb{K}_\lambda(\cdot)$ denotes the modified Bessel function of the second kind and order λ. The following formulas bridge the original definition shown above and the location-scale-based parameterization implemented in SciPy: $a = \alpha\delta$, $b = \beta\delta,\ p = \lambda,\ scale = \delta,\ loc = \mu$.

Characteristic functions - cont'd

- Particularly, the Normal Inverse Gaussian characteristic function is:

  $\varphi(u,\alpha,\beta,\delta,\mu) = \mathbb{e}^{iu\mu+\delta\sqrt{\alpha^2-\beta^2}-\delta\sqrt{\alpha^2-(\beta+iu)^2}}$

for $x \in (-\infty,+\infty)$. $\delta > 0$, $|\beta| \le \alpha$.

The following formulas bridge the original definition shown above and the location-scale-based parameterization implemented in SciPy: $a = \alpha\delta$, $b = \beta\delta,\ scale = \delta,\ loc = \mu$.

Linear transformation on characteristic functions

- Consider a linear transformation: $\mathcal{L}(x) = \mathcal{A}x + \mathcal{B}$, where $\mathcal{A} \neq 0$.
- For a random variable $\xi \sim \mathbb{N}(\mu,\sigma^2)$, the linear transformation of it $\mathcal{L}(\xi) \sim \mathbb{N}[\mathcal{A}\mu+\mathcal{B},\ (\mathcal{A}\sigma)^2]$.

  Proof. $\varphi_{\mathcal{L}(\xi)}(u,\tilde{\mu},\tilde{\sigma}) = \varphi_{\mathcal{A}\xi+\mathcal{B}}(u) \coloneqq \mathbb{E}\left[\mathbb{e}^{iu(\mathcal{A}\xi+\mathcal{B})}\right] = \mathbb{e}^{iu\mathcal{B}}\mathbb{E}\left[\mathbb{e}^{i(\mathcal{A}u)\xi}\right]$

  $\Rightarrow\ \mathbb{E}\left[\mathbb{e}^{i(\mathcal{A}u)\xi}\right] = \varphi_{\xi}(\mathcal{A}u,\mu,\sigma) = \mathbb{e}^{i(\mathcal{A}u)\mu-\frac{1}{2}(\mathcal{A}u)^2\sigma^2} = \mathbb{e}^{iu(\mathcal{A}\mu)-\frac{1}{2}u^2(\mathcal{A}\sigma)^2}$

  $\Rightarrow\ \varphi_{\mathcal{L}(\xi)}(u,\tilde{\mu},\tilde{\sigma}) = \mathbb{e}^{iu\mathcal{B}}\mathbb{E}\left[\mathbb{e}^{i(\mathcal{A}u)\xi}\right] = \mathbb{e}^{iu\mathcal{B}}\mathbb{e}^{iu(\mathcal{A}\mu)-\frac{1}{2}u^2(\mathcal{A}\sigma)^2} = \mathbb{e}^{iu(\mathcal{A}\mu+\mathcal{B})-\frac{1}{2}u^2(\mathcal{A}\sigma)^2}$

  $\Rightarrow\ \tilde{\mu} = \mathcal{A}\mu+\mathcal{B},\ \ \tilde{\sigma} = \mathcal{A}\sigma.$

  $\Rightarrow\ \widetilde{\boldsymbol{loc}} = \tilde{\mu} = \boldsymbol{\mathcal{A}loc} + \boldsymbol{\mathcal{B}},\ \ \widetilde{\boldsymbol{scale}} = \tilde{\sigma} = \boldsymbol{\mathcal{A}} \cdot \boldsymbol{scale}.$

Linear transformation on characteristic functions - cont'd

- For a random variable $\xi \sim \mathbb{GH}(\lambda, \alpha, \beta, \delta, \mu)$, the linear transformation of it $\mathcal{L}(\xi) \sim \mathbb{GH}\left(\lambda, \frac{\alpha}{\mathcal{A}}, \frac{\beta}{\mathcal{A}}, \mathcal{A}\delta, \mathcal{A}\mu + \mathcal{B}\right)$. Particularly, it is also true of $\mathbb{NIG}$.

Proof. $\varphi_{\mathcal{L}(\xi)}(u, \tilde{\lambda}, \tilde{\alpha}, \tilde{\beta}, \tilde{\delta}, \tilde{\mu}) = \varphi_{\mathcal{A}\xi+\mathcal{B}}(u) \coloneqq \mathbb{E}\left[e^{iu(\mathcal{A}\xi+\mathcal{B})}\right] = e^{iu\mathcal{B}}\mathbb{E}\left[e^{i(\mathcal{A}u)\xi}\right]$

$$\Rightarrow \quad \mathbb{E}\left[e^{i(\mathcal{A}u)\xi}\right] = \varphi_\xi(\mathcal{A}u, \mu, \sigma) = e^{i(\mathcal{A}u)\mu}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+i(\mathcal{A}u))^2}\right)^{\frac{\lambda}{2}} \frac{\mathbb{K}_\lambda\left(\delta\sqrt{\alpha^2-(\beta+i(\mathcal{A}u))^2}\right)}{\mathbb{K}_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$$

$$= e^{iu(\mathcal{A}\mu)}\left(\frac{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\frac{\beta}{\mathcal{A}}\right)^2}{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\left(\frac{\beta}{\mathcal{A}}\right)+iu\right)^2}\right)^{\frac{\lambda}{2}} \frac{\mathbb{K}_\lambda\left(\mathcal{A}\delta\sqrt{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\left(\frac{\beta}{\mathcal{A}}\right)+iu\right)^2}\right)}{\mathbb{K}_\lambda\left(\mathcal{A}\delta\sqrt{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\frac{\beta}{\mathcal{A}}\right)^2}\right)}$$

$$\Rightarrow \varphi_{\mathcal{L}(\xi)}(u, \tilde{\lambda}, \tilde{\alpha}, \tilde{\beta}, \tilde{\delta}, \tilde{\mu}) = e^{iu\mathcal{B}}\mathbb{E}\left[e^{i(\mathcal{A}u)\xi}\right]$$

$$= e^{iu(\mathcal{A}\mu+\mathcal{B})}\left(\frac{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\frac{\beta}{\mathcal{A}}\right)^2}{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\left(\frac{\beta}{\mathcal{A}}\right)+iu\right)^2}\right)^{\frac{\lambda}{2}} \frac{\mathbb{K}_\lambda\left(\mathcal{A}\delta\sqrt{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\left(\frac{\beta}{\mathcal{A}}\right)+iu\right)^2}\right)}{\mathbb{K}_\lambda\left(\mathcal{A}\delta\sqrt{\left(\frac{\alpha}{\mathcal{A}}\right)^2-\left(\frac{\beta}{\mathcal{A}}\right)^2}\right)}$$

$$\Rightarrow \tilde{\alpha} = \frac{\alpha}{\mathcal{A}}, \quad \tilde{\beta} = \frac{\beta}{\mathcal{A}}, \quad \tilde{\delta} = \mathcal{A}\delta, \quad \tilde{\mu} = \mathcal{A}\mu + \mathcal{B}, \quad \tilde{\lambda} = \lambda.$$

$\Rightarrow \widetilde{\boldsymbol{a}} = \tilde{\alpha}\tilde{\delta} = \boldsymbol{a}, \ \ \widetilde{\boldsymbol{b}} = \tilde{\beta}\tilde{\delta} = \boldsymbol{b}, \ \ \widetilde{\boldsymbol{scale}} = \tilde{\delta} = \boldsymbol{\mathcal{A}} \cdot \boldsymbol{scale}, \ \ \widetilde{\boldsymbol{loc}} = \tilde{\mu} = \boldsymbol{\mathcal{A}loc} + \boldsymbol{\mathcal{B}}, \ \ \widetilde{\boldsymbol{p}} = \tilde{\lambda} = \boldsymbol{p}.$

ARMA(1, 1) Model

$\mathbb{r}_t = \mu + \phi \mathbb{r}_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$

- $\{\epsilon_t\}$ is called ***innovations***. And ϵ_t are uncorrelated random variables with $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{Var}[\epsilon_t] = \sigma^2$, which will be fitted with Levy models or a Gaussian model later;
- $\{\mathbb{r}_t\}$ stands for the asset log return process;
- $\mathbb{E}[\mathbb{r}_t] \coloneqq \tilde{\mu}$ satisfies $(1-\phi)\tilde{\mu} = \mu$ under the ***stationarity assumption*** that the root of $1 - \phi z = 0$ is outside the unit circle, which gives the condition of $-1 < \phi < 1$;
- $\widetilde{\mathbb{r}_t} = \phi \widetilde{\mathbb{r}_{t-1}} + \theta \epsilon_{t-1} + \epsilon_t$ can be reached after a transformation that $\widetilde{\mathbb{r}_t} \coloneqq \mathbb{r}_t - \tilde{\mu}$. Next, consider ***invertibility condition***, the root of $1 + \theta z = 0$ should be outside the unit circle, which gives the condition of $-1 < \theta < 1$.

In conclusion, ARMA(1, 1) model can be rewritten as:
$\widetilde{\mathbb{r}_t} = \phi \widetilde{\mathbb{r}_{t-1}} + \theta \epsilon_{t-1} + \epsilon_t$, where $\widetilde{\mathbb{r}_t} \coloneqq \mathbb{r}_t - \mathbb{E}[\mathbb{r}_t]$ and $-1 < \phi \,\&\, \theta < 1$.

- **GARCH(1, 1) Model**

$$\begin{cases} \sigma_t^2 = \omega + \psi\sigma_{t-1}^2 + \beta\epsilon_{t-1}^2 \\ \epsilon_t = \mathbb{u}_t\sigma_t \end{cases}$$

  - $\{\epsilon_t\}$ is called ***innovations***. And ϵ_t are uncorrelated random variables with $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{Var}[\epsilon_t] = \sigma^2$, which will be fitted with Levy models or a Gaussian model by using process $\{u_t\}$;
  - $\{\mathbb{u}_t\}$ stands for a certain random process. And u_t are *i.i.d.* random variables with with $\mathbb{E}[\mathbb{u}_t] = 0$ and $\mathbb{Var}[\mathbb{u}_t] = 1$, which will be also fitted with Levy models or a Gaussian model;
  - $\psi > 0$ and $\beta > 0$ are always assumed in this model;
  - To ensure that ϵ_t are ***covariance stationary***, it is required that the root of $1 - (\psi + \beta)z = 0$ is outside the unit circle, which gives the condition of $\psi + \beta < 1$;
  - Once ϵ_t are covariance stationary ($\mathbb{E}[\sigma_t^2] \equiv \sigma^2$), we can get the following relationship: $\sigma^2 = \frac{\omega}{1-\psi-\beta}$.

- **GARCH(1, 1) Model - cont'd**

$$\begin{cases} \sigma_t^2 = \omega + \psi\sigma_{t-1}^2 + \beta\epsilon_{t-1}^2 \\ \epsilon_t = \mathbb{u}_t\sigma_t \end{cases}$$

➢ From ARMA(1, 1) model $\widetilde{\mathbb{r}_t} = \phi\widetilde{\mathbb{r}_{t-1}} + \theta\epsilon_{t-1} + \epsilon_t$ on previous page. Take unconditional variance on the two sides and we can get :

$\mathbb{Var}[\widetilde{\mathbb{r}_t}] = \frac{(1-\phi^2)+(\phi+\theta)^2}{1-\phi^2}\sigma^2$, thus $\sigma^2 = \frac{(1-\phi^2)\mathbb{Var}[\widetilde{\mathbb{r}_t}]}{(1+2\phi\theta+\theta^2)}$.

Proof. $\mathbb{Var}[\widetilde{\mathbb{r}_t}] = \lim_{n\to\infty} \mathbb{Var}\left[\phi^n(\widetilde{\mathbb{r}_{t-n}} - \epsilon_{t-n}) + \epsilon_t + (\phi+\theta)\sum_{i=1}^n \phi^{i-1}\epsilon_{t-i}\right]$

$\approx \lim_{n\to\infty} \mathbb{Var}\left[\epsilon_t + (\phi+\theta)\sum_{i=1}^n \phi^{i-1}\epsilon_{t-i}\right]$

$= \mathbb{Var}[\epsilon_t] + (\phi+\theta)^2 \lim_{n\to\infty} \sum_{i=1}^n \phi^{2i-2}\mathbb{Var}[\epsilon_{t-i}]$

$= \sigma^2 + \frac{(\phi+\theta)^2}{1-\phi^2}\sigma^2$ (covariance stationary)

In conclusion, GARCH(1, 1) model can be rewritten as:

$$\begin{cases} \sigma_t^2 = \frac{(1-\psi-\beta)(1-\phi^2)\mathbb{Var}[\mathbb{r}_t]}{1+2\phi\theta+\theta^2} + \psi\sigma_{t-1}^2 + \beta\epsilon_{t-1}^2 \\ \epsilon_t = \mathbb{u}_t\sigma_t \end{cases}$$

where $\mathbb{Var}[\widetilde{\mathbb{r}_t}] = \mathbb{Var}[\mathbb{r}_t]$, $\psi > 0$, $\beta > 0$ and $\psi + \beta < 1$.

In order to parameterize the ARMA(1,1) - GARCH(1,1) model, we can rely on the Log-likelihood function of GARCH(1,1) model. Since the calculation of $\{\epsilon_t\}$ are based on ARMA(1, 1) as well, parameterization can be done simultaneously, where we can get a precise result in the end.

- **Log-likelihood function in GARCH(1, 1) Model**
  - Let $\mathbf{\Theta} = (\omega, \psi, \beta)^T$, however, in this case, the parameters should be extended, i.e. $\widetilde{\mathbf{\Theta}} = (\omega, \phi, \theta, \psi, \beta)^T$.
  - Then, the log-likelihood function in the GARCH(1, 1) model is given by: $l(\widetilde{\mathbf{\Theta}}) = \sum_{t=1}^{n} \log[\mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_t | \epsilon_{t-1})]$.

Proof. To perform maximum-likelihood estimation, we analyze in the joint distribution $\mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_1, \cdots, \epsilon_n)$ where $\widetilde{\boldsymbol{\Theta}}$ is the parameter vector. Using iteratively that the joint distribution is equal to the product of the conditional and the marginal density, we obtain

$$\begin{aligned}
\mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_1, \cdots, \epsilon_n) &= \mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_0, \epsilon_1, \cdots, \epsilon_n) \\
&= \mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_0) \cdot \mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_1, \cdots, \epsilon_n | \epsilon_0) \\
&= \mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_0) \cdot \prod_{t=1}^{n} \mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_t | \epsilon_{t-1}, \cdots, \epsilon_0) \\
&= \mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_0) \cdot \prod_{t=1}^{n} \mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_t | \epsilon_{t-1})
\end{aligned}$$

For deterministic function $\mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_0)$, dropping this term and taking logs, we obtain the log-likelihood function $l(\widetilde{\mathbf{\Theta}}) = \sum_{t=1}^{n} \log[\mathbb{f}(\widetilde{\boldsymbol{\Theta}}; \epsilon_t | \epsilon_{t-1})]$.

Log-likelihood function in GARCH(1, 1) Model - cont'd

- If $\{\mathbb{u}_t\}$ follows a Gaussian distribution, and based on zero mean, unit variance assumption, we can get $\mathbb{u}_t \sim \mathbb{N}(0,\ 1)$. Then conditioning on filtration $\mathcal{F}_{t-1}$, ϵ_t can be seen as a const. Finally, we deduce that $[\epsilon_t|\mathcal{F}_{t-1}] = \sigma_t \mathbb{u}_t \sim \mathbb{N}(0, \sigma_t^2)$, i.e., $\mathbb{f}_{\epsilon_t\,|\,\mathcal{F}_{t-1}}(x) = \frac{1}{\sqrt{2\pi}\sigma_t} \mathbb{e}^{-\frac{x^2}{2\sigma_t^2}}$.
- Thus, the log-likelihood function in the GARCH(1, 1) model is given by:
$$l_{\mathbb{N}}(\widetilde{\boldsymbol{\Theta}}) = \sum_{t=1}^{n} \log\left[\mathbb{f}_{\epsilon_t\,|\,\mathcal{F}_{t-1}}(\epsilon_t)\right] = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{n}\frac{\epsilon_t^2}{\sigma_t^2}$$
- Similarly, if $\{\mathbb{u}_t\}$ follows a Generalized Hyperbolic distribution, calibrating distribution parameters to satisfy zero mean, unit variance assumption is required.
- However, mean and variance are complicated to calibrate as shown below since location-scale parameters cannot be explicitly expressed by other parameters.
$$\boldsymbol{\mu}_{\mathbb{GH}}(\lambda, \alpha, \beta, \delta, \mu) = \mu + \frac{\delta\beta\mathbb{K}_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\sqrt{\alpha^2-\beta^2}\mathbb{K}_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$$
$$\boldsymbol{\sigma}^2_{\mathbb{GH}}(\lambda, \alpha, \beta, \delta, \mu) = \frac{\delta\mathbb{K}_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\sqrt{\alpha^2-\beta^2}\mathbb{K}_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)} + \frac{\beta^2\delta^2}{\alpha^2-\beta^2}\cdot\left(\frac{\mathbb{K}_{\lambda+2}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\mathbb{K}_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)} - \frac{\mathbb{K}^2_{\lambda+1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\mathbb{K}^2_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\right)$$
- Thus, we give up the idea to fit $\{\mathbb{u}_t\}$ with a Generalized Hyperbolic distribution.

Stony Brook University

Log-likelihood function in GARCH(1, 1) Model - cont'd

- If $\{\mathbb{u}_t\}$ follows a Normal Inverse Gaussian distribution, calibrating distribution parameters to satisfy zero mean, unit variance assumption is also required.
- Mean and variance are listed below, we have to make location and scale parameters explicitly expressed by other parameters.

$$\boldsymbol{\mu}_{\mathbb{NIG}}(\alpha,\beta,\delta,\mu) = \mu + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}} = 0;\ \ \boldsymbol{\sigma}^2_{\mathbb{NIG}}(\alpha,\beta,\delta,\mu) = \frac{\delta\alpha^2}{\sqrt{(\alpha^2-\beta^2)^3}} = 1.$$

$$\Rightarrow \mu = -\frac{\beta(\alpha^2-\beta^2)}{\alpha^2};\ \ \delta = \frac{\sqrt{(\alpha^2-\beta^2)^3}}{\alpha^2}$$

- Thus, we can get $u_t \sim \mathbb{NIG}\left(\alpha,\beta,\delta = \frac{\sqrt{(\alpha^2-\beta^2)^3}}{\alpha^2}, \mu = -\frac{\beta(\alpha^2-\beta^2)}{\alpha^2}\right)$.
- According to linear transformation of Normal Inverse Gaussian random variable. And similarly, we can deduce that

$[\epsilon_t|\mathcal{F}_{t-1}] = \sigma_t u_t \sim \mathbb{NIG}\left(\frac{\alpha}{\sigma_t}, \frac{\beta}{\sigma_t}, \delta\sigma_t, \mu\sigma_t\right)$, i.e.,

$$\mathbb{f}_{\epsilon_t\,|\,\mathcal{F}_{t-1}}(x) = \frac{\frac{\alpha}{\sigma_t}\delta\sigma_t\mathbb{K}_1\left(\frac{\alpha}{\sigma_t}\sqrt{\delta^2\sigma_t^2+(x-\mu\sigma_t)^2}\right)}{\pi\sqrt{\delta^2\sigma_t^2+(x-\mu\sigma_t)^2}}\mathbb{e}^{\delta\sigma_t\sqrt{\left(\frac{\alpha}{\sigma_t}\right)^2-\left(\frac{\beta}{\sigma_t}\right)^2}+\frac{\beta}{\sigma_t}(x-\mu\sigma_t)}$$

$$= \frac{1}{\sigma_t}\cdot\frac{\alpha\delta\mathbb{K}_1\left(\alpha\sqrt{\delta^2+\left(\frac{x}{\sigma_t}-\mu\right)^2}\right)}{\pi\sqrt{\delta^2+\left(\frac{x}{\sigma_t}-\mu\right)^2}}\mathbb{e}^{\delta\sqrt{\alpha^2-\beta^2}+\beta\left(\frac{x}{\sigma_t}-\mu\right)}$$

Log-likelihood function in GARCH(1, 1) Model - cont'd

- Thus, the log-likelihood function in the GARCH(1, 1) model is given by:

$$l_{\mathbb{NIG}}(\widetilde{\boldsymbol{\Theta}}) = \sum_{t=1}^{n} \log\left[\mathbb{f}_{\epsilon_t \mid \mathcal{F}_{t-1}}(\epsilon_t)\right] = \sum_{t=1}^{n} \log\left[\frac{1}{\sigma_t} \cdot \frac{\alpha\delta\mathbb{K}_1\left(\alpha\sqrt{\delta^2+\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)^2}\right)}{\pi\sqrt{\delta^2+\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)^2}} \mathbb{e}^{\delta\sqrt{\alpha^2-\beta^2}+\beta\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)}\right]$$

$$= \sum_{t=1}^{n}\left\{\log\left(\alpha\delta\mathbb{K}_1\left(\alpha\sqrt{\delta^2+\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)^2}\right)\right) - \frac{1}{2}\cdot\log\left(\delta^2+\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)^2\right)\right.$$
$$\left. -\log(\pi\sigma_t) + \delta\sqrt{\alpha^2-\beta^2} + \beta\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)\right\}$$

$$= n\cdot\delta\sqrt{\alpha^2-\beta^2} - n\cdot\beta\cdot\mu + \beta\sum_{t=1}^{n}\left(\frac{\epsilon_t}{\sigma_t}\right) - \sum_{t=1}^{n}\log(\pi\sigma_t)$$
$$+ \sum_{t=1}^{n}\log\left(\alpha\delta\mathbb{K}_1\left(\alpha\sqrt{\delta^2+\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)^2}\right)\right) - \frac{1}{2}\sum_{t=1}^{n}\log\left(\delta^2+\left(\frac{\epsilon_t}{\sigma_t}-\mu\right)^2\right)$$

where $\mathbb{K}_1(\cdot)$ denotes the modified Bessel function of the second kind and order 1, and location and scale parameters have following relationship between α and β:

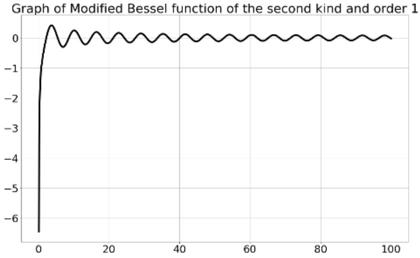
$\mu = -\frac{\beta(\alpha^2-\beta^2)}{\alpha^2}$; $\delta = \frac{\sqrt{(\alpha^2-\beta^2)^3}}{\alpha^2}$.

- This way, we can fit the ARMA-GARCH and NIG parameters simultaneously, rather than using two-step procedure, which gives us a more accurate estimation (But it is not always true!).
- However, the time of handling MLE might become much more than before.

- **Difficulty of fitting MLE with NIG (log) likelihood function**
  - $l_{\mathbb{NIG}}(\widetilde{\boldsymbol{\Theta}}) = n \cdot \delta\sqrt{\alpha^2 - \beta^2} - n \cdot \beta \cdot \mu + \beta \sum_{t=1}^{n} \left(\frac{\epsilon_t}{\sigma_t}\right) - \sum_{t=1}^{n} \log(\pi\sigma_t)$
    $+ \sum_{t=1}^{n} \log\left(\alpha\delta\mathbb{K}_1\left(\alpha\sqrt{\delta^2 + \left(\frac{\epsilon_t}{\sigma_t} - \mu\right)^2}\right)\right) - \frac{1}{2}\sum_{t=1}^{n} \log\left(\delta^2 + \left(\frac{\epsilon_t}{\sigma_t} - \mu\right)^2\right)$

    where $\mathbb{K}_1(\cdot)$ denotes the ***modified Bessel function of the second kind and order 1***, and location and scale parameters have following relationship between α and β:

    $\mu = -\frac{\beta(\alpha^2-\beta^2)}{\alpha^2}$; $\delta = \frac{\sqrt{(\alpha^2-\beta^2)^3}}{\alpha^2}$.
  - However, when analyzing the modified Bessel function of the second kind and order 1, there comes a severe problem: it can have negative values.
  - Notice that there are terms in form of $\log(\mathbb{K}_1(\cdot))$. Thus, on the one hand, when we use the log likelihood function, it is very likely to cause undefined issues. On the other hand, when we try to use likelihood function as an alternative, it is still problematic: existence of negative values in product operations make itself very sensitive to changes of variables, thus hard to converge.
    Meanwhile, the computational cost would be extremely high.



Stony Brook University

- **ARMA(1, 1) - GARCH(1, 1) Model**

$$\begin{cases} \widetilde{\mathbb{r}_t} = \phi \widetilde{\mathbb{r}_{t-1}} + \theta \epsilon_{t-1} + \epsilon_t \\ \sigma_t^2 = \omega + \psi \sigma_{t-1}^2 + \beta \epsilon_{t-1}^2 \\ \epsilon_t = \mathbb{u}_t \sigma_t \end{cases}$$

where $\widetilde{\mathbb{r}_t} \coloneqq \mathbb{r}_t - \mathbb{E}[\mathbb{r}_t]$ and $-1 < \phi \,\&\, \theta < 1$, $\psi > 0$, $\beta > 0$ and $\psi + \beta < 1$.

- **Forecasting the distribution of log return one step ahead**

  - $$\begin{aligned} [\mathbb{r}_{t+1} \mid \mathcal{F}_t] &= [\widetilde{\mathbb{r}_{t+1}} + \mathbb{E}[\mathbb{r}_{t+1}] \mid \mathcal{F}_t] \\ &\approx [\widetilde{\mathbb{r}_{t+1}} + \mathbb{E}[\mathbb{r}_t] \mid \mathcal{F}_t] \\ &= \mathbb{E}[\mathbb{r}_t] + [\widetilde{\mathbb{r}_{t+1}} \mid \mathcal{F}_t] \\ &= \mathbb{E}[\mathbb{r}_t] + [\phi \widetilde{\mathbb{r}_t} + \theta \epsilon_t + \epsilon_{t+1} \mid \mathcal{F}_t] \\ &= \mathbb{E}[\mathbb{r}_t] + \phi \widetilde{\mathbb{r}_t} + \theta \epsilon_t + [\epsilon_{t+1} \mid \mathcal{F}_t] \\ &= \mathbb{E}[\mathbb{r}_t] + \phi \widetilde{\mathbb{r}_t} + \theta \epsilon_t + [\mathbb{u}_{t+1} \sqrt{\omega + \psi \sigma_t^2 + \beta \epsilon_t^2} \mid \mathcal{F}_t] \\ &= \mathbb{E}[\mathbb{r}_t] + \phi \widetilde{\mathbb{r}_t} + \theta \epsilon_t + \sqrt{\omega + \psi \sigma_t^2 + \beta \epsilon_t^2} [\mathbb{u}_{t+1} \mid \mathcal{F}_t] \\ &\coloneqq \mathbb{E}[\mathbb{r}_{t+1} \mid \mathcal{F}_t] + \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t][\mathbb{u}_{t+1} \mid \mathcal{F}_t] \end{aligned}$$

where $\mathbb{E}[\mathbb{r}_{t+1} \mid \mathcal{F}_t] = \mathbb{E}[\mathbb{r}_t] + \phi \widetilde{\mathbb{r}_t} + \theta \epsilon_t$, $\mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t] = \sqrt{\omega + \psi \sigma_t^2 + \beta \epsilon_t^2}$.

Forecasting the distribution of log return one step ahead - cont'd

Now, we should refer to the conclusions of linear transformation on characteristic functions to determine the distribution.

- $[\mathbb{r}_{t+1} \mid \mathcal{F}_t] = \mathbb{E}[\mathbb{r}_{t+1} \mid \mathcal{F}_t] + \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t] \cdot [\mathbb{u}_{t+1} \mid \mathcal{F}_t]$ (from previous page).

- If $\{\mathbb{u}_t\}$ follows a Gaussian distribution, i.e. $[\mathbb{u}_{t+1} \mid \mathcal{F}_t] \sim \mathbb{N}(\hat{\mu},\ \hat{\sigma}^2)$. Then, $[\mathbb{r}_{t+1} \mid \mathcal{F}_t] \sim \mathbb{N}[\mathbb{E}[\mathbb{r}_{t+1} \mid \mathcal{F}_t] + \hat{\mu} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t],\ \ (\hat{\sigma} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t]\ )^2]$.
- Thus, the location-scale-based parameterization implemented in SciPy: $\widetilde{\widehat{loc}} = \widehat{loc} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t] + \mathbb{E}[\mathbb{r}_{t+1} \mid \mathcal{F}_t]$, $\widetilde{\widehat{scale}} = \widehat{scale} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t]$.

- If $\{\mathbb{u}_t\}$ follows a Generalized Hyperbolic distribution, i.e. $[\mathbb{u}_{t+1} \mid \mathcal{F}_t] \sim \mathbb{GH}\left(\hat{\lambda}, \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu}\right) \Leftrightarrow [\mathbb{u}_{t+1} \mid \mathcal{F}_t] \sim \mathbb{GH}\left(\hat{p}, \hat{a}, \hat{b}, \widehat{loc}, \widehat{scale}\right)$.
- Then, $[\mathbb{r}_{t+1} \mid \mathcal{F}_t]$
$$\sim \mathbb{GH}\left(\hat{p}, \hat{a}, \hat{b}, \widehat{loc} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t] + \mathbb{E}[\mathbb{r}_{t+1} \mid \mathcal{F}_t],\ \widehat{scale} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t]\right).$$
- Thus, the location-scale-based parameterization implemented in SciPy: $\tilde{\hat{a}} = \hat{a}$, $\tilde{\hat{b}} = \hat{b}$, $\tilde{\hat{p}} = \hat{p}$, $\widetilde{\widehat{loc}} = \widehat{loc} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t] + \mathbb{E}[\mathbb{r}_{t+1} \mid \mathcal{F}_t]$, $\widetilde{\widehat{scale}} = \widehat{scale} \cdot \mathbb{E}[\sigma_{t+1} \mid \mathcal{F}_t]$.

Stony Brook University

- **Assumption Tests**
  - Time series (i.e., log return series $\{\mathbb{r}_t\}$) stationary assumption

    Implemented with ADF test and KPSS test
  - Unit process (i.e., $\{\mathbb{u}_t\}$) zero mean and unit variance assumption

    Implemented with student-T tests

- **Unconditional Coverage Tests[2]**
  - Backtesting is a statistical procedure where actual profits and losses are systematically compared to corresponding VaR estimates. For example, if the confidence level used for calculating daily VaR is 99%, we expect an exception to occur once in every 100 days on average. *In the backtesting process we could statistically examine whether the frequency of exceptions over some specified time interval is in line with the selected confidence level.* These types of tests are known as tests of ***unconditional coverage.***

  - Binomial Test
  - Kupiec's POF Test (proportion of failures)

Conditional Coverage Tests[2]

- In theory, however, a good VaR model not only produces the ‘correct’ amount of exceptions but also *exceptions that are evenly spread over time i.e. are independent of each other*. Clustering of exceptions indicates that the model does not accurately capture the changes in market volatility and correlations. Tests of ***conditional coverage*** therefore examine also conditioning, or time variation, in the data. (Jorion, 2001)

- Duration - based tests (Christoffersen and Pelletier, 2004; Berkowitz et al., 2011)

Augmented Dickey-Fuller (ADF) Test[3]

- Statistical tests make strong assumptions about your data. They can only be used to inform the degree to which a null hypothesis can be rejected or fail to be rejected. The result must be interpreted for a given problem to be meaningful.
- However, they provide a quick check and confirmatory evidence that the time series is stationary or non-stationary.

- The Augmented Dickey-Fuller test is a type of statistical test called a unit root test.
- In probability theory and statistics, a unit root is a feature of some stochastic processes (such as random walks) that can cause problems in statistical inference involving time series models. In a simple term, the unit root is non-stationary but does not always have a trend component.

- ADF test is conducted with the following assumptions.
  - Null hypothesis ($\mathbb{H}_0$): Series is non-stationary or series has a unit root.
  - Alternate hypothesis($\mathbb{H}_a$): Series is stationary or series has no unit root.

  If the null hypothesis is failed to be rejected, this test may provide evidence that the series is non-stationary.
- Conditions to reject null hypothesis($\mathbb{H}_0$)

  If test statistic < critical value or p-value < 0.05 – Reject null hypothesis($\mathbb{H}_0$) i.e., time series does not have a unit root, meaning it is stationary. It does not have a time-dependent structure.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test[3]

- The KPSS test, short for, Kwiatkowski-Phillips-Schmidt-Shin (KPSS), is a type of Unit root test that tests for the stationarity of a given series around a deterministic trend.
- In other words, the test is somewhat similar in spirit with the ADF test.

- A common misconception, however, is that it can be used interchangeably with the ADF test. This can lead to misinterpretations about the stationarity, which can easily go undetected causing more problems down the line.

- A key difference from ADF test is the null hypothesis of the KPSS test is that the series is stationary. In other words, hypothesis is reversed in KPSS test compared to ADF test. ***So practically, the interpretation of p-value is just the opposite to each other.***

- KPSS test is conducted with the following assumptions.
  · Null hypothesis ($\mathbb{H}_0$): Series is trend stationary or series has no unit root.
  · Alternate hypothesis($\mathbb{H}_a$): Series is non-stationary or series has a unit root.
  If the null hypothesis is failed to be rejected, this test may provide evidence that the series is trend stationary.
- Conditions to Fail to Reject Null Hypothesis($\mathbb{H}_0$)
  If test statistic > critical value and p-value > 0.05 – Fail to reject null hypothesis($\mathbb{H}_0$) i.e., time series does not have a unit root, meaning it is trend stationary.

Binomial Test[1]

- It states that if $\{\mathbb{I}_t(\alpha)\}$ are independently and identically distributed and $\mathbb{Pr}[\mathbb{I}_{t+1}(\alpha) = 1] = \alpha$, then the total number of exceedances $\mathbb{H}$ has a binomial distribution $\mathbb{B}(n, \alpha)$ with mean $\mathbb{E}[\mathbb{H}] = n\alpha$ and $\mathbb{Var}[\mathbb{H}] = n\alpha(1-\alpha)$.
- If the number of observations is large enough, the central limit theorem can approximate the binomial distribution by the normal distribution.
- As outlined by Jorion (2001), an immediate test statistic is:

$$\mathbb{T} = \frac{\mathbb{H} - n\alpha}{\sqrt{n\alpha(1-\alpha)}}$$

  Its asymptotic null distribution is the standard normal distribution.
- An alternative is to use the likelihood ratio statistic:

$$\mathbb{LR}_{Bin} = -2\ln[(1-\alpha)^{n-\mathbb{H}}\alpha^{\mathbb{H}}] + 2\ln\left[\left(1-\frac{\mathbb{H}}{n}\right)^{n-\mathbb{H}}\left(\frac{\mathbb{H}}{n}\right)^{\mathbb{H}}\right]$$

  Its asymptotic null distribution is the chi-squared distribution with one degree of freedom.

Python code for Binomial Test

scipy.stats.binomtest (k, n, p=0.5, alternative='two-sided')

- Perform a test that the probability of success is p.
- The binomial test is a test of the null hypothesis that the probability of success in a Bernoulli experiment is p.

- **Parameters:**
- (Int)k: The number of successes.
- (Int)n: The number of trials.
- (float)p: Optional. The hypothesized probability of success, i.e. the expected proportion of successes. The value must be in the interval 0 <= p <= 1. The default value is p = 0.5.
- Alternative{'two-sided', 'greater', 'less'}: Optional. Indicates the alternative hypothesis. The default value is 'two-sided'.

- **Returns:**

  (instance)resultBinomTestResult: The return value is an object with the following attributes:
- (Int)k: The number of successes (copied from binomtest input).
- (Int)n: The number of trials (copied from binomtest input).
- (Str)alternative: Indicates the alternative hypothesis specified in the input to binomtest. It will be one of 'two-sided', 'greater', or 'less'.
- (float)pvalue: The p-value of the hypothesis test.
- (float)proportion_estimate: The estimate of the proportion of successes.

Reference:
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.binomtest.html

Kupiec's POF Test (proportion of failures)[2]

- Kupiec's test measures whether the number of exceedances is consistent with the confidence level. Under null hypothesis of the model being 'correct', the number of exceptions follows the binomial distribution discussed in the last page. Hence, the only information required to implement a POF-test is the number of observations(n), number of exceedances($\mathbb{H}$) and the confidence level(c). (Dowd, 2006).
- The null hypothesis for the POF-test is $\mathbb{H}_0: \alpha = \hat{\alpha} = \frac{\mathbb{H}}{n}$. The idea is to find out whether the observed failure rate $\hat{\alpha}$ is significantly different from α, the failure rate suggested by the confidence level. According to Kupiec(1995), the POF-test is best conducted as a likelihood-ratio test, The test statistic takes the form: $\mathbb{LR}_{POF} = 2\ln\left[\frac{\left(1-\frac{\mathbb{H}}{n}\right)^{n-\mathbb{H}}\left(\frac{\mathbb{H}}{n}\right)^{\mathbb{H}}}{(1-\alpha)^{n-\mathbb{H}}\alpha^{\mathbb{H}}}\right]$.
- Under the null hypothesis that the model is correct, $\mathbb{LR}_{POF}$ is asymptotically χ^2 distributed with one degree of freedom. If the value of the $\mathbb{LR}_{POF}$-statistic exceeds the critical value of the χ^2 distribution, the null hypothesis will be rejected and the model is deemed as inaccurate.

- Kupiec's POF-test is hampered by two ***shortcomings.*** First, the test is statistically weak with sample sizes consistent with current regulatory framework (one year). This lack of power has already been recognized by Kupiec himself. Secondly, POF-test considers only the frequency of losses and not the time when they occur. As a result, it may fail to reject a model that produces clustered exceptions. Thus, model backtesting should not rely solely on tests of unconditional coverage. (Campbell, 2005).

Duration - based tests (Christoffersen and Pelletier, 2004; Berkowitz et al., 2011)[1]

- The main idea behind duration based tests is that clustering of violations will result in a large number of relatively short and long no hit durations, corresponding to market turbulence and market calm.
- Denote the number of days between two $VaR(\alpha)$ violations by $d_i = t_i - t_{i-1}$, where t_i denotes the day of violation i.
- Suppose the $\{d_i\}$ follows the exponential distribution specified by the probability density function $f(d) = p\mathrm{e}^{-pd}$, for d > 0 and p > 0. Then, the null hypothesis that the risk model is correctly specified corresponds to $p = \alpha$ (Christoffersen and Pelletier, 2004).
- Suppose now that the $\{d_i\}$ follow the Weibull distribution specified by the probability density function $f(d) = bp^b d^{b-1}\mathrm{e}^{-(pd)^b}$, for d > 0, b > 0 and p > 0. Then, the null hypothesis that the risk model is correctly specified corresponds to $b = 1$ and $p = \alpha$ (Berkowitz et al., 2011). The null hypothesis of independence of violations corresponds to $b = 1$ (Berkowitz et al., 2011).
- All of these tests can be based on the likelihood ratio principle. The implementation of these tests is very straightforward and provides a clear interpretation of parameters. However, these tests are not very popular among practitioners. The main drawback of these tests is that they have relatively small power for realistic sample sizes (Haas, 2007) and normally one would struggle in computing the standard LR duration based statistics.

Part II.

Stony Brook University

Empirical Study

— providing managers with some ideas

Part II. Empirical Study – Full procedure of analysis

Step1. Download Data

- **Stocks and Indices Data**
  - Mainly obtained via Yahoo Finance Python Interface
  - Link: https://www.yahoo.com/

- **Volatility Indices Data**
  - Mainly obtained on CBOE (the Chicago Board Options Exchange) Official Website.
  - Link: https://www.cboe.com/us/indices/indicessearch/
  - Details of ***Cboe Equity VIX on IBM, Goldman Sachs, Google***.

Cboe applies its proprietary Cboe Volatility Index® (VIX®) methodology to options on select individual stocks. The indexes are designed to measure the expected volatility of the respective individual equities.

  - Details of ***Cboe Options on AMZN, DJIA, AAPL***.

The Cboe ~ VIX IndexSM (VX~) is a VIX®-style estimate of the expected 30-day volatility of ~ stock returns. Like VIX, VX~ is calculated by interpolating between two weighted sums of option midquote values, in this case options on ~. The two sums essentially represent the expected variance of the ~ returns up to two option expiration dates that bracket a 30-day period of time. VX~ is obtained by annualizing the interpolated value, taking its square root and expressing the result in percentage points.

Part II. Empirical Study – Full procedure of analysis

Step2. Log Return Series Stationary Test

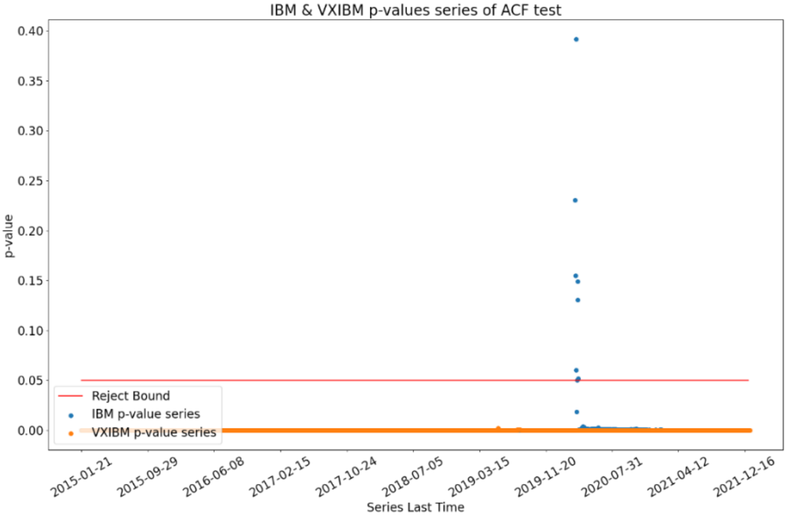
Stony Brook University

Asset: IBM stock (Sample Size : 2000)		Notes
ADF Test p-value	KPSS Test p-value	ADF Test Null hypothesis : Series is non-stationary or series has a unit root. Reject null hypothesis when p-value < 0.05, which may provide evidence that the series is non-stationary. KPSS Test Null hypothesis : Series is trend stationary or series has no unit root. Reject null hypothesis when p-value < 0.05, which may provide evidence that the series is trend stationary. If p-value > 0.05, then fail to reject null hypothesis, i.e., time series does not have a unit root, meaning it is trend stationary.
6.867591e-26	0.100000	
Asset: VXIBM index (Sample Size : 2000)		
ADF Test p-value	KPSS Test p-value	
0.000000	0.100000	

First, I carry out the tests in the whole length. For these two kinds of assets, both ADF test and KPSS test conclude that the log return series are stationary.
Then, I will analyze how do the p-values behave when we carry out the rolling window analysis, where window size is 250.

Part II. Empirical Study – Full procedure of analysis

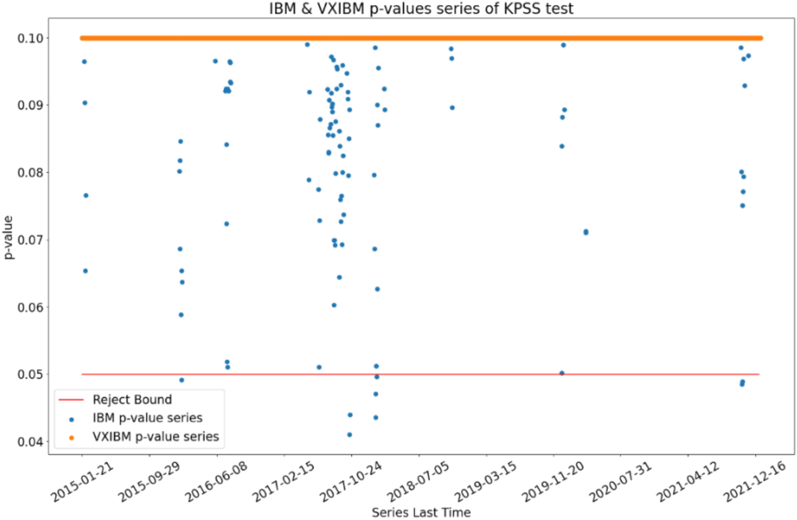
Step2. Log Return Series Stationary Test



This diagram shows the p-values of ACF test on the log return series of previous 250 time steps. We can see that all of VXIBM samples satisfy the stationary assumption, while there are 7 exceptions (out of 1750) in IBM, but it still can be regarded to satisfy the assumption.

Part II. Empirical Study – Full procedure of analysis

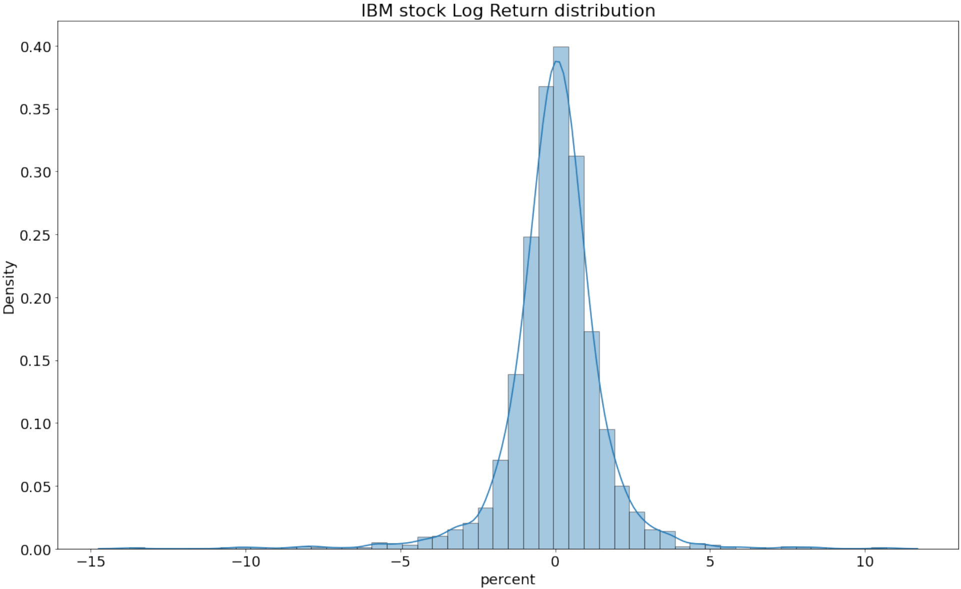
Step2. Log Return Series Stationary Test



This diagram shows the p-values of KPSS test on the log return series of previous 250 time steps. We can see that all of VXIBM samples satisfy the stationary assumption, while there are 8 exceptions (out of 1750) in IBM, but it still can be regarded to satisfy the assumption.

Part II. Empirical Study – Full procedure of analysis

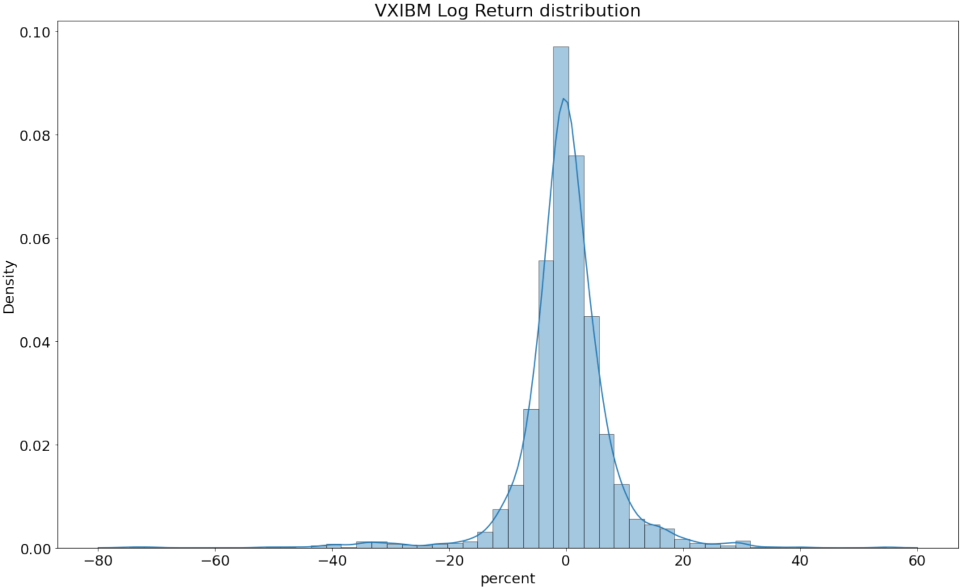
Step3. Plot Log Return Series graphs



Part II. Empirical Study – Full procedure of analysis

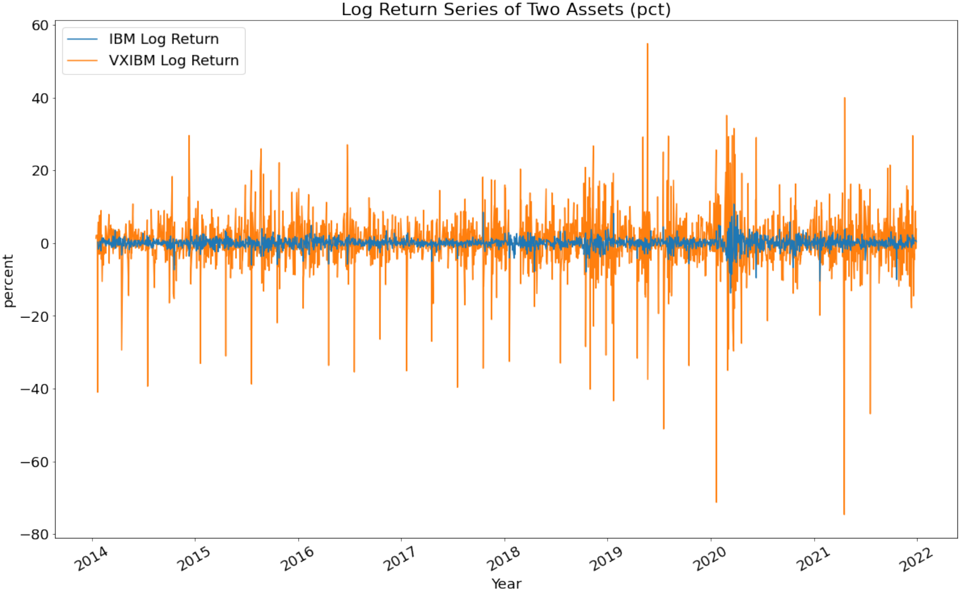
Stony Brook University

Step3. Plot Log Return Series graphs



Part II. Empirical Study – Full procedure of analysis

Step3. Plot Log Return Series graphs

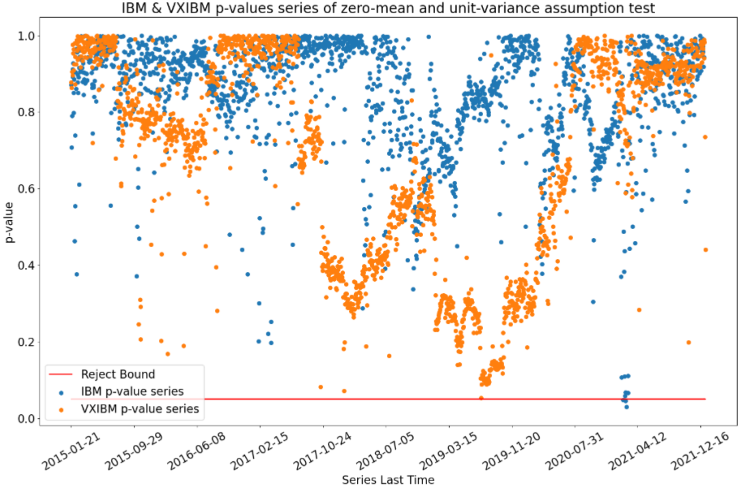


Step4. Run the code

Part II. Empirical Study – Full procedure of analysis

Step5. Residual zero mean and unit variance assumption test

Stony Brook University

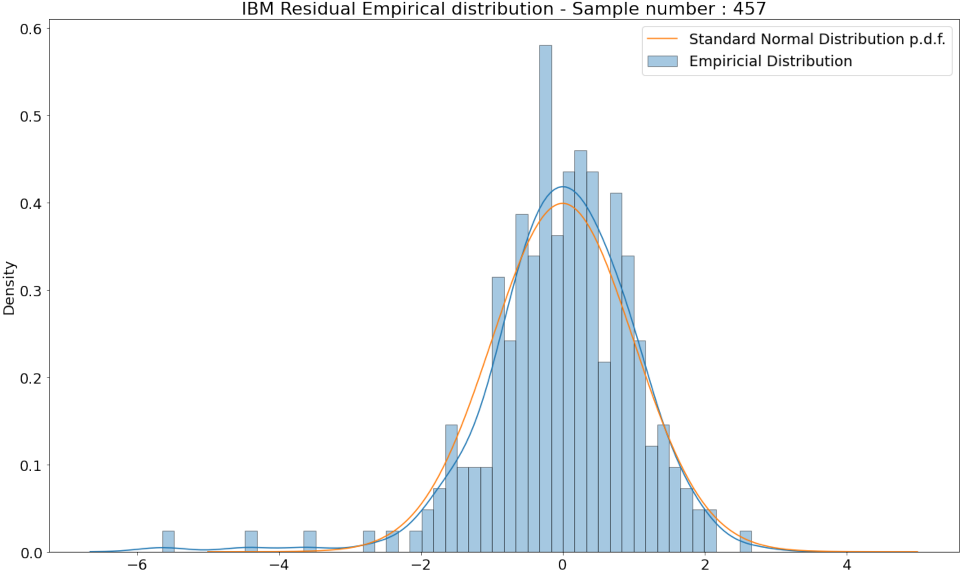


We can see that there are only 3 exception in the case of IBM stock, so the two assets satisfy this zero mean and unit variance very well. Then I randomly choose four residual series to illustrate that.

Notes: Null hypothesis: Mean of distribution is 0 and variance of distribution is 1. Reject null hypothesis when p-value < 0.05, which indicates that it does not satisfy the assumption, thus the model is not robust.

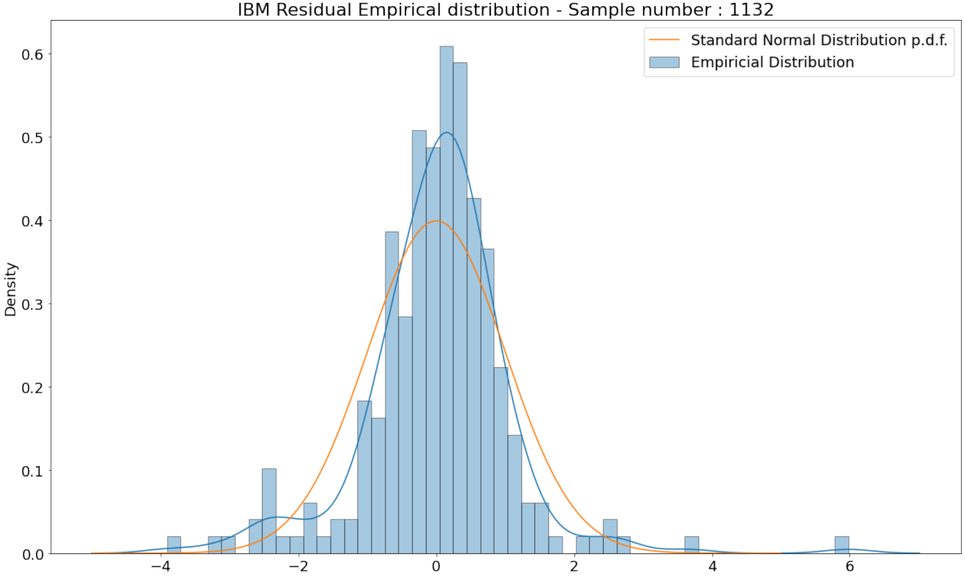
Part II. Empirical Study – Full procedure of analysis

Step5. Residual zero mean and unit variance assumption test



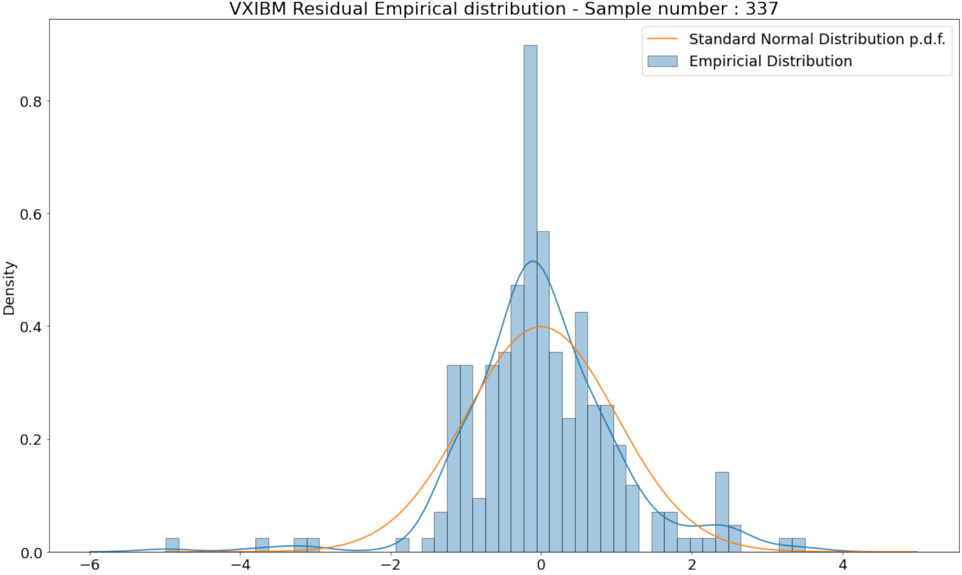
Part II. Empirical Study – Full procedure of analysis

Step5. Residual zero mean and unit variance assumption test



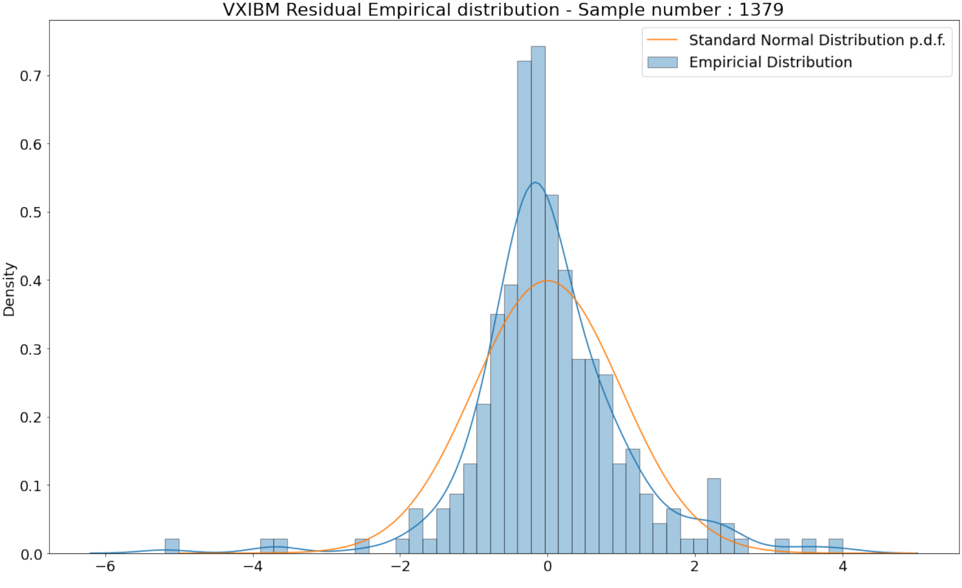
Part II. Empirical Study – Full procedure of analysis

Step5. Residual zero mean and unit variance assumption test



Part II. Empirical Study – Full procedure of analysis

Step5. Residual zero mean and unit variance assumption test



Main further study of midterm project:

- Add GH case to this model

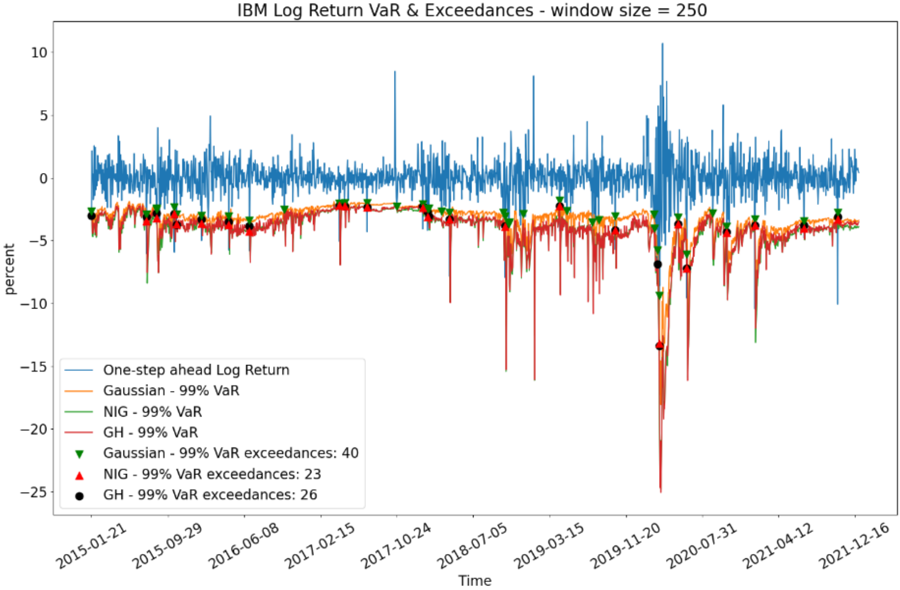
- Add the risk measure of CVaR.

- Plot the distribution surface.

- Analyze the relationship between window size and exceedance ratio.

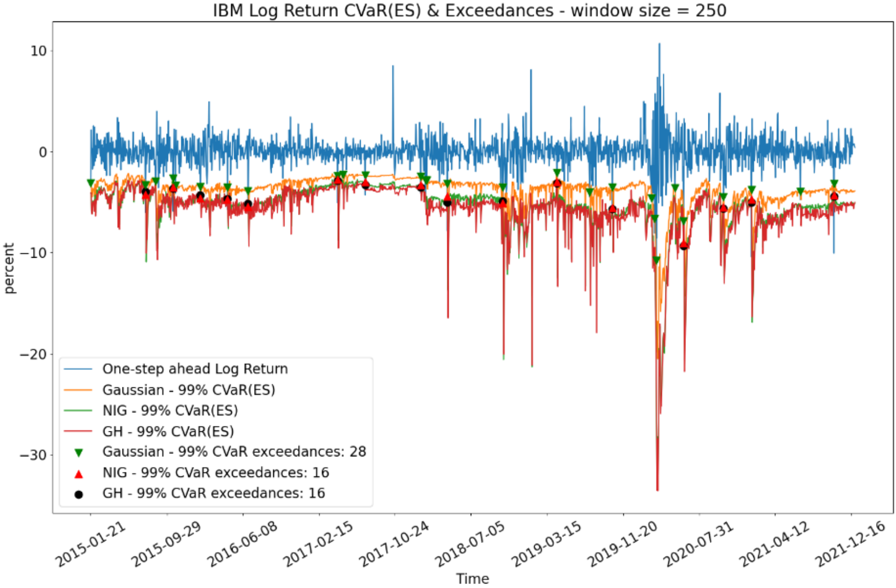
Part II. Empirical Study – Full procedure of analysis

Step6. VaR and Exceedances Analysis



Part II. Empirical Study – Full procedure of analysis

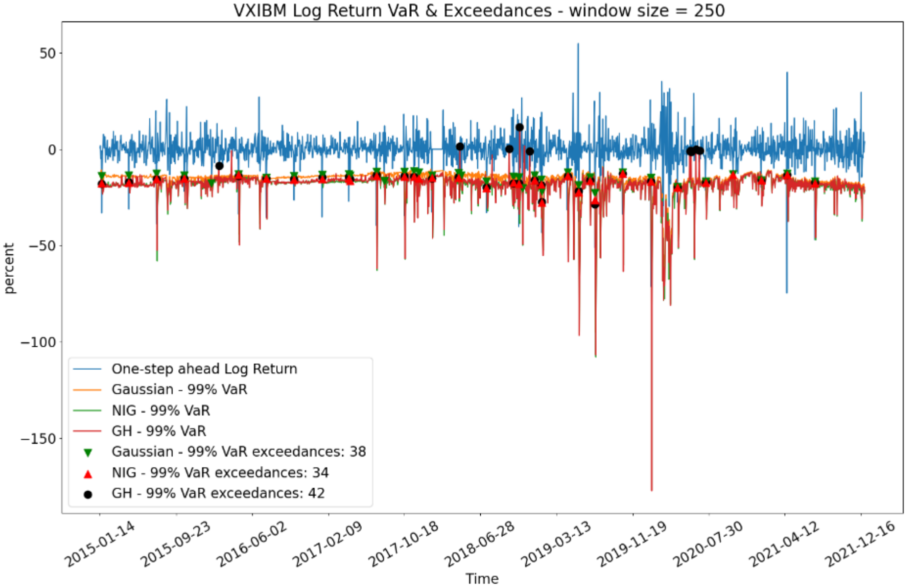
Step6. VaR and Exceedances Analysis



Part II. Empirical Study – Full procedure of analysis

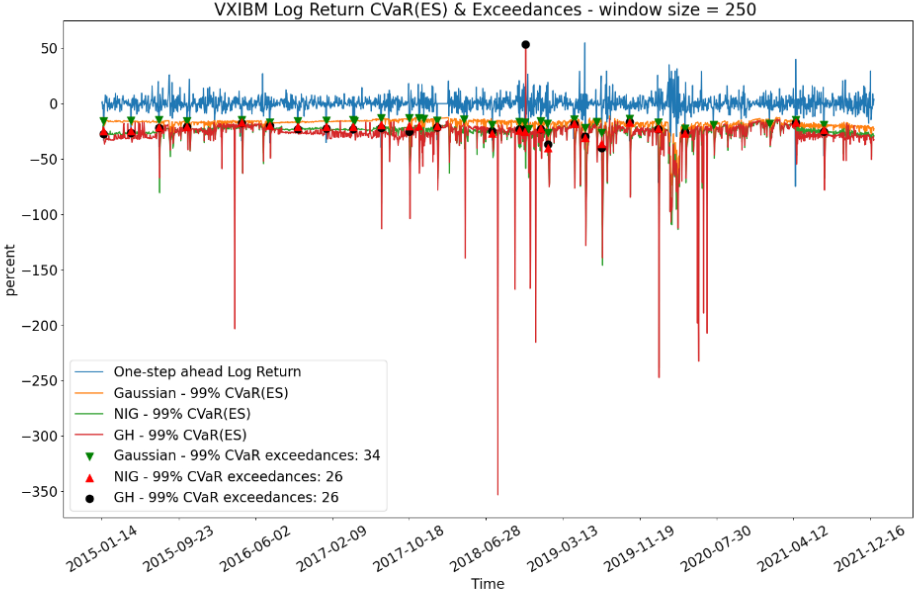
Stony Brook University

Step6. VaR and Exceedances Analysis



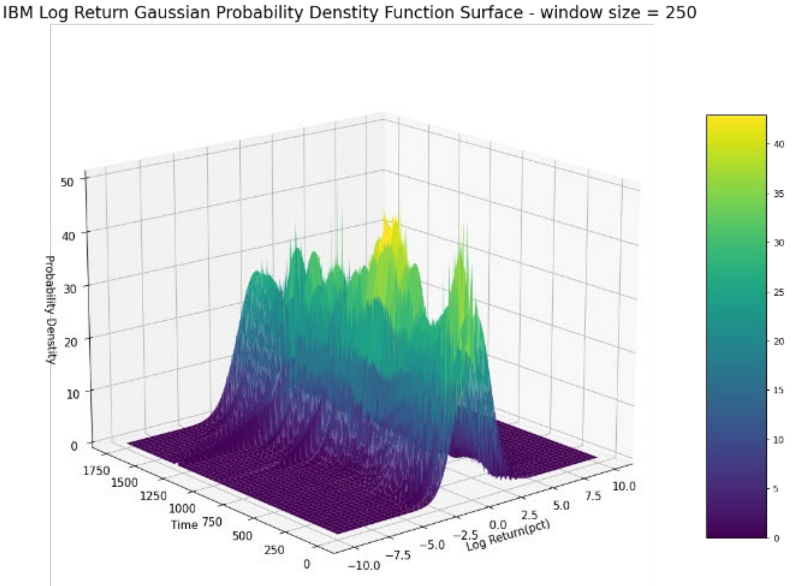
Part II. Empirical Study – Full procedure of analysis

Step6. VaR and Exceedances Analysis



Part II. Empirical Study – Full procedure of analysis

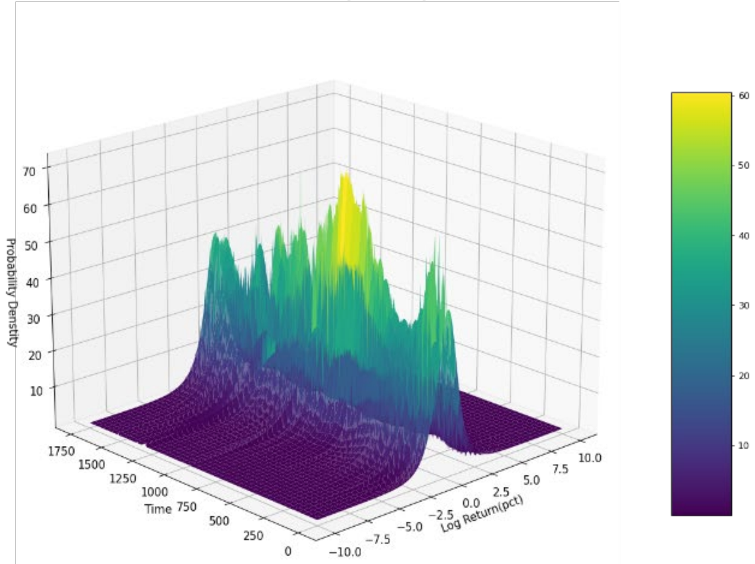
Step7. Plot Distribution Surface graphs



Part II. Empirical Study – Full procedure of analysis

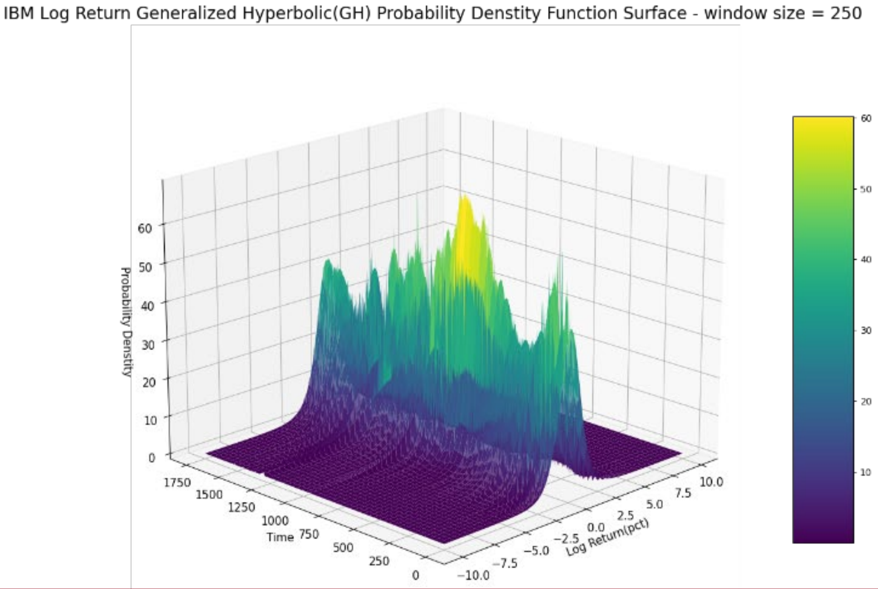
Step7. Plot Distribution Surface graphs

IBM Log Return Normal Inverse Gaussian(NIG) Probability Denstity Function Surface - window size = 250



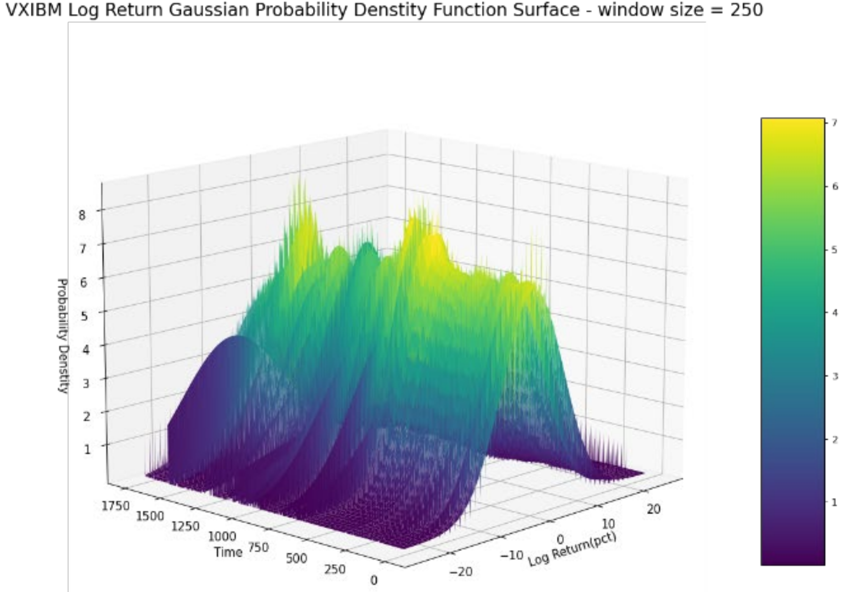
Part II. Empirical Study – Full procedure of analysis

Step7. Plot Distribution Surface graphs



Part II. Empirical Study – Full procedure of analysis

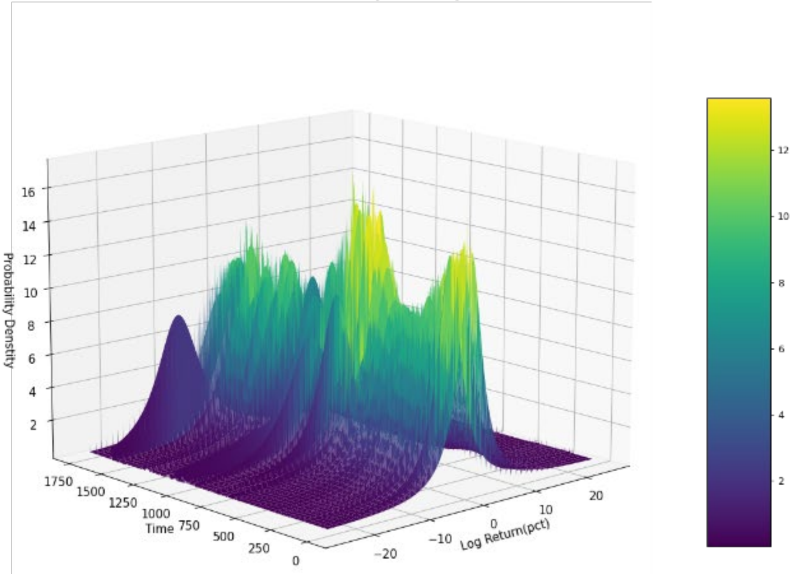
Step7. Plot Distribution Surface graphs



Part II. Empirical Study – Full procedure of analysis

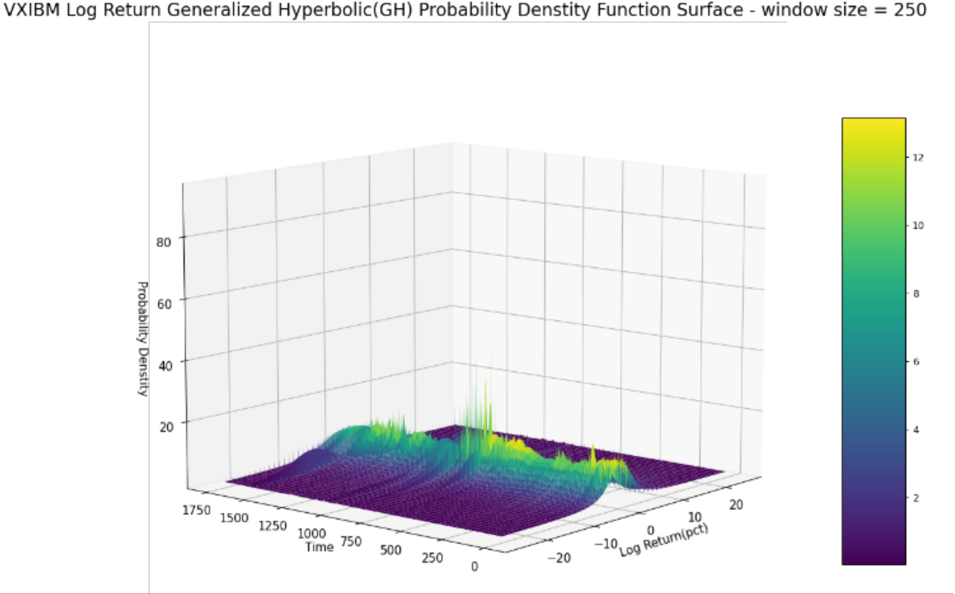
Step7. Plot Distribution Surface graphs

VXIBM Log Return Normal Inverse Gaussian(NIG) Probability Denstity Function Surface - window size = 250



Part II. Empirical Study – Full procedure of analysis

Step7. Plot Distribution Surface graphs



Part II. Empirical Study – Full procedure of analysis

Step8. Backtest - Unconditional Coverage Tests

Asset: IBM stock (Sample Size : 1750)						
Distribution	Gaussian	NIG	GH	Gaussian	NIG	GH
Risk Measure	99% VaR	99% VaR	99% VaR	99% CVaR	99% CVaR	99% CVaR
Exceedances	40	23	26	28	16	16
KupiecTest p-value	**3.6741e-06**	0.2075	0.05680	**0.02032**	0.7146	0.7146
BinomialTest p-value	**2.8472e-06**	0.1848	0.05282	**0.01595**	0.8105	0.8105

We can see that in this case, both Gaussian distributions failed. And other models can give a relatively more accurate estimate of risk. And CVaR models perform better than VaR, since they have larger p-values.

Notes: Null hypothesis: Probability of exceedances is 0.01, and reject null hypothesis when p-value < 0.05, which indicates that it is not a good model. Binomial Test Alternative Hypothesis: Two-sided.

Part II. Empirical Study – Full procedure of analysis

Step8. Backtest - Unconditional Coverage Tests

Stony Brook University

Asset: VXIBM index (Sample Size : 1750)						
Distribution	Gaussian	NIG	GH	Gaussian	NIG	GH
Risk Measure	99% VaR	99% VaR	99% VaR	99% CVaR	99% CVaR	99% CVaR
Exceedances	38	34	42	34	26	26
KupiecTest p-value	**2.01738e-05**	**0.0004480**	**6.07757e-07**	**0.0004480**	0.05680	0.05680
BinomialTest p-value	**1.7104e-05**	**0.0003999**	**4.2848e-07**	**0.0003999**	0.05282	0.05282

We can see that in this case, where log return series behave more unstably, all of the VaR models and CVaR gaussian model failed. And NIG and GH CVaR models can give a not bad estimate of risk.

Notes: Null hypothesis: Probability of exceedances is 0.01, and reject null hypothesis when p-value < 0.05, which indicates that it is not a good model. Binomial Test Alternative Hypothesis: Two-sided.

Part II. Empirical Study – Full procedure of analysis

Step8. Backtest - Conditional Coverage Tests

Stony Brook University

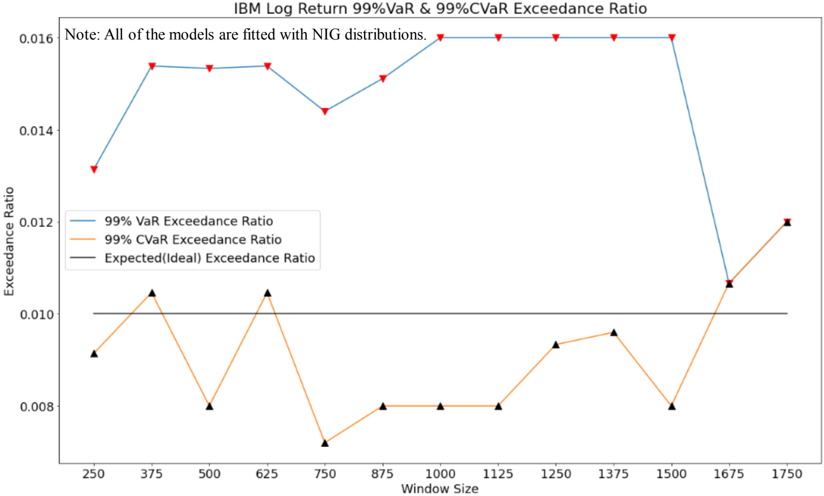
Asset: IBM stock						
Distribution	Gaussian	NIG	GH	Gaussian	NIG	GH
Risk Measure	99% VaR	99% VaR	99% VaR	99% CVaR	99% CVaR	99% CVaR
Duration Test p-value	1(fail to converge)	0.07430	1(fail to converge)	1(fail to converge)	**0.000504 91**	**0.000504 91**
Asset: VXIBM index						
Distribution	Gaussian	NIG	GH	Gaussian	NIG	GH
Risk Measure	99% VaR	99% VaR	99% VaR	99% CVaR	99% CVaR	99% CVaR
Duration Test p-value	**0.002767**	**0.001881**	0.4259	**0.01631**	**0.006786**	**0.02676**

We can see that only NIG VaR model in the first case and GH VaR model in the second case perform well. But considering the latter one will give an abnormal estimate of risk, it cannot be regarded as a good model.

Notes: Null hypothesis: Duration between exceedances have no memory (Weibull b=1 = Exponential), and reject null hypothesis when p-value < 0.05, which indicates that the model is likely to be suffering from volatility clustering.

Part II. Empirical Study – Full procedure of analysis

Step9. Hyperparameter analysis - window size & Exceedance Ratio



The idea of exceedance ratio is the similar to the unconditional coverage test.

We can see that as the window size increases, the exceedance ratio is not likely to change. So there is no problem to set the window size to 250 (by convention).

Finally, we can conclude NIG CVaR models always overestimate the risk, while NIG VaR underestimate.

Conclusions:

- For stock assets, NIG 99% VaR model performs best. Not only can it accurately estimate the risk, it can also avoid suffering from volatility clustering, while NIG 99% CVaR is a good choice too, though it did not pass the duration test.
- For VIX on stocks indices assets, where there are volatile behaviors in log return series, GH 99% CVaR may be a good choice, though it did not pass the duration test as well. It can give a good estimate of risk at such an extreme situation. However, GH distribution is very sensitive to outliers, which might be a latent problem.
- These are the conclusions I draw from previous analysis. But it is not convincing enough. We have to repeat this procedure to different assets from stocks or volatility indices.

Future work

- **Carry out this produce on different assets**
- **Introduce some other risk measures, like CoVaR and so on, which can better reflect the systemic risk.**

Reference

- [1] Y. Zhang and S. Nadarajah (2017). A review of backtesting for value at risk.
- [2] Kansantaloustiede, Maisterin tutkinnon tutkielma and Olli Nieppola (2009). Backtesting Value-at-Risk Models.
- [3] https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/
- Tze Leung Lai, Haipeng Xing(2008). Statistical Models and Methods for Financial Markets.