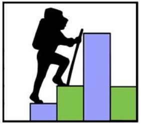
Ostrum ASSET MANAGEMENT

Evaluating density forecasts

Antoine Lepeltier

26 août 2021



Introduction

Evaluating Density Forecasts, International Economic Review volume 39 (1997),

Co-écrit par Francis X. Diebold, Todd A. Gunther et Anthony S. Tay.

Implémentation en *Python* pour faciliter la représentation graphique.

Plan

1. Cadre théorique
2. Résultats théoriques
3. Application pratique
4. Applications sur données réelles
   - S&P500
   - Euro Stoxx 50
5. Extensions
   - Amélioration de l'estimation
   - Multi-dimensionelle

Cadre

$(y_t)_{t=1}^m$ générées selon $(f_t|\Omega_t)_{t=1}^m$ où $\Omega_t = \{y_{t-1}, y_{t-2}, ..., y_1\}$.

Soit $(p_t|\Omega_t)_{t=1}^m$ l'estimation de la densité f.

$L(a, y)$ la fonction de perte de l'utilisateur avec

$$a^*(p(y)) = \operatorname*{argmin}_{a \in \mathcal{A}} \int L(a, y) p(y) dy$$

Résultats théoriques

L'utilisateur préfère p_j à p_k si

$$\int L(a_j^*, y) f(y) dy \leq \int L(a_k^*, y) f(y) dy$$

Proposition 1 : Soit f la vraie densité de y et a_j^* (resp. a_k^*) l'action optimale basée sur p_j (resp. p_k). Sous l'hypothèse que p_j et p_k sont différentes de f, il n'est pas possible d'établir un classement entre les 2.

Le choix de p_j ou de p_k dépend de la fonction de perte L.

Résultats théoriques

Proposition 2 : Si $p_j \sim f$ alors l'action a_j^* basée sur p_j minimise toujours les pertes moyennes

$$\mathbb{E}[L(a_j^*, y)] = \int L(a_j^*, y) f(y) dy$$

Si l'estimation p_j est bonne elle sera toujours préférée peu importe la fonction de perte de l'utilisateur.

Résultats théoriques

L'objectif est d'évaluer la qualité de notre estimation.

Probability integral transform $z_t = \int_{-\infty}^{y_t} p_t(u)du = P_t(y_t)$.

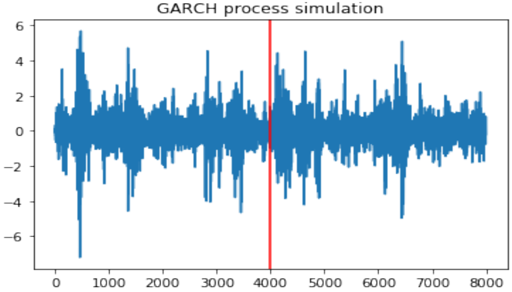
Proposition 3 : Sous l'hypothèse que $(p_t(y_t|\Omega_t))_{t=1}^m$ coincïde avec $(f_t(y_t|\Omega_t))_{t=1}^m$ et que le Jacobien de P_t^{-1} est continu et différent de 0 alors

$$(z_t)_{t=1}^m \overset{\text{iid}}{\sim} \mathcal{U}(0,1)$$

Evaluation à l'aide d'un histogramme et de corrélogrammes

Application pratique

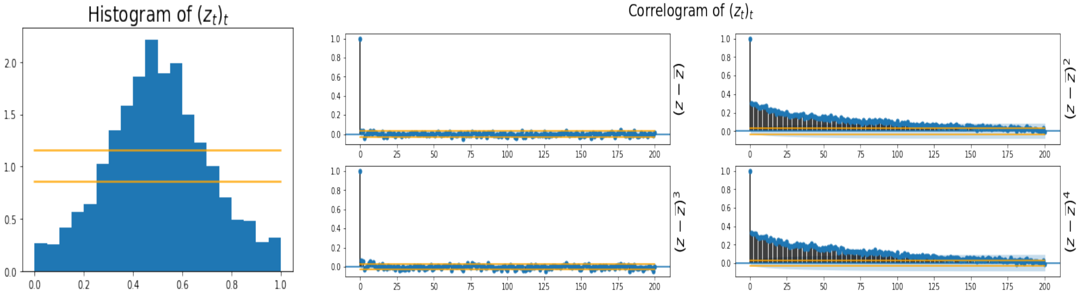
$$\text{t-GARCH(1,1)} : \begin{cases} y_t = \sqrt{\frac{h_t^2(\nu-2)}{\nu}}\mathcal{T}(\nu) \\ h_t^2 = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2 \end{cases} \quad \omega = 0,01 \ \alpha = 0,13 \ \beta = 0,86.$$



4000 observations pour estimer le modèle et les 4000 suivantes pour l'évaluation.

Application pratique

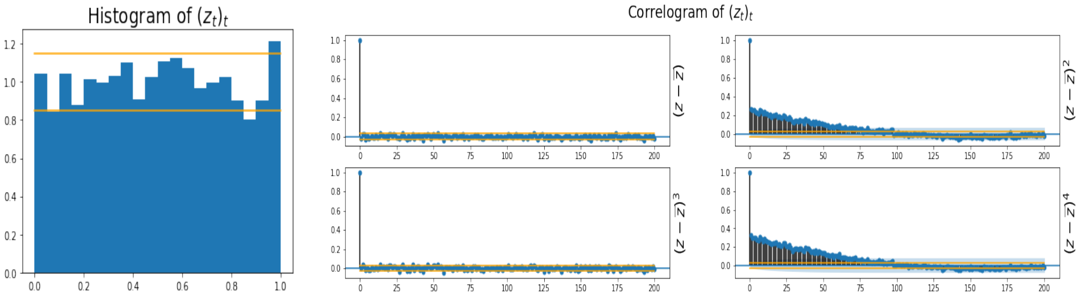
$$(y_t)_{t=1}^{m} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$



La suite $(z_t)_t$ n'est ni distribuée selon une $\mathcal{U}(0,1)$ ni indépendante.

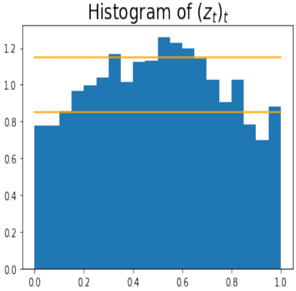
Application pratique

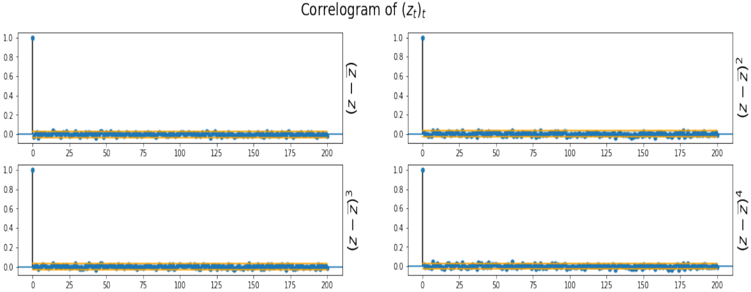
$$y_t \sim \hat{p}_h(y) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{y-y_i}{h}\right) \text{ et } h = 1.06n^{-\frac{1}{5}} \text{ (Silverman B.W, 1986).}$$



La suite $(z_t)_t$ est maintenant distribuée selon une $\mathcal{U}(0,1)$ mais n'est toujours pas indépendante.

$$y_t \sim \text{Gaussian GARCH(1,1)}\ \omega = 0,01\ \alpha = 0,13\ \beta = 0,86.$$

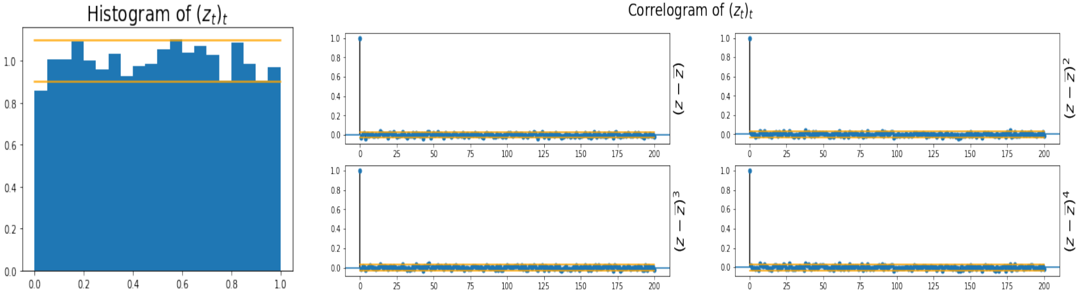




La suite $(z_t)_t$ n'est plus distribuée selon une $\mathcal{U}(0,1)$ mais est indépendante.

Application pratique

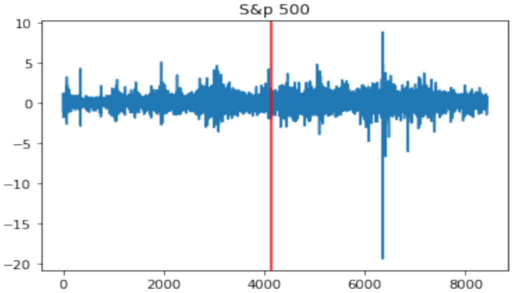
$$y_t \sim \text{Student GARCH(1,1)}, \ \omega = 0,01 \ \alpha = 0,13 \ \beta = 0,86.$$



La suite $(z_t)_t$ est distribuée selon une $\mathcal{U}(0,1)$ et est indépendante. D'après nos critères c'est une bonne estimation.

Application sur données réelles

Daily returns du *S&P 500* de 1962 à 1995.



4100 observations pour estimer le modèle et les 4300 suivantes pour l'évaluation.

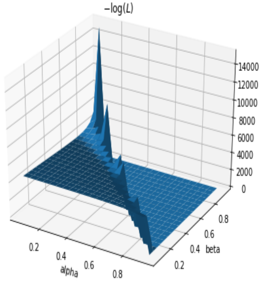
Application sur données réelles

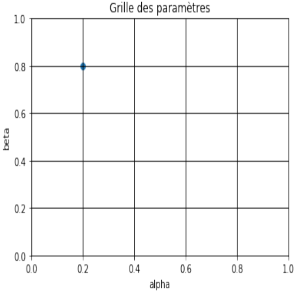
$$\text{ARMA}(p_m, q_m)\text{GARCH}(p_v, q_v) :=$$
$$\begin{cases} y_t &= \mu_t + h_t \xi_t \\ \mu_t &= \mu + \sum_{i=1}^{p_m} \phi_i y_{t-1} + \sum_{i=1}^{q_m} \theta_i \xi_{t-1} \\ h_t^2 &= \omega + \sum_{i=1}^{p_v} \alpha_i y_{t-1}^2 + \sum_{i=1}^{q_v} \beta_i h_{t-1}^2 \end{cases}$$

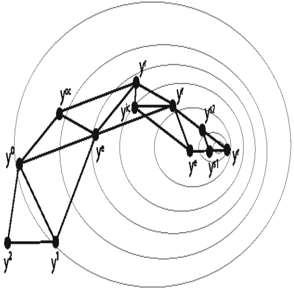
Bayesian information criterion, $BIC = k \ln N - 2 \ln(L)$

- k : nombre de paramètres
- N : taille de l'échantillon
- L : vraisemblance

Estimation des paramètres par la méthode du maximum de vraisemblance.







Pour que le processus soit stationnaire il y a des bornes sur les paramètres.

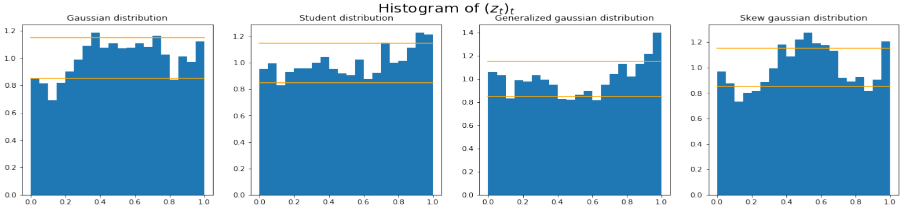
Recherche du minimum sur une grille des paramètres puis avec la méthode de **Nelder-Mead**.

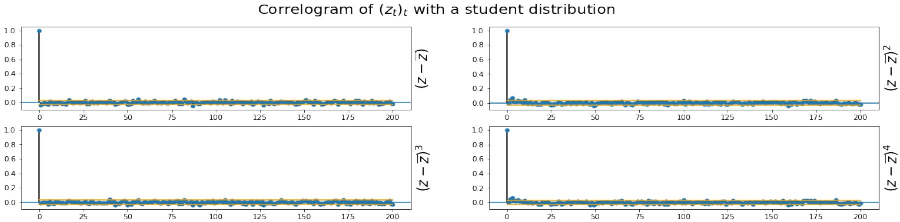
Application sur données réelles

Le critère BIC sélectionne un ARMA(0,1)-GARCH(1,1).

```
                 ARMA(0,1)-GARCH(1,1)
================================================
Distribution            Normal
Standard deviation      0.96
Method                  Maximum Likelihood
Log Likelihood          -3967.042020987171
AIC                     7944.084041974342
BIC                     7967.391077232389
================================================
                        Coeffs
================================================
--------------------Mean Model------------------
theta[0]    0.2555273437499999
----------------Volatility Model----------------
omega       0.006953475275568938
alpha[0]    0.11913311641542426
beta[0]     0.8808667833316336
```

Application sur données réelles

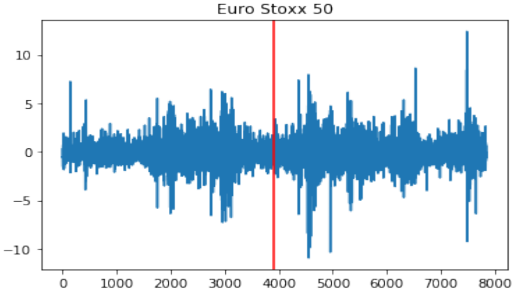




D'après nos critères le modèle Student ARMA(0,1)-GARCH(1,1) est une bonne estimation des données du S&P 500.

Application sur données réelles

Daily returns *Euro Stoxx 50* disponibles sur *www.stoxx.com* de 1992 à aujourd'hui.



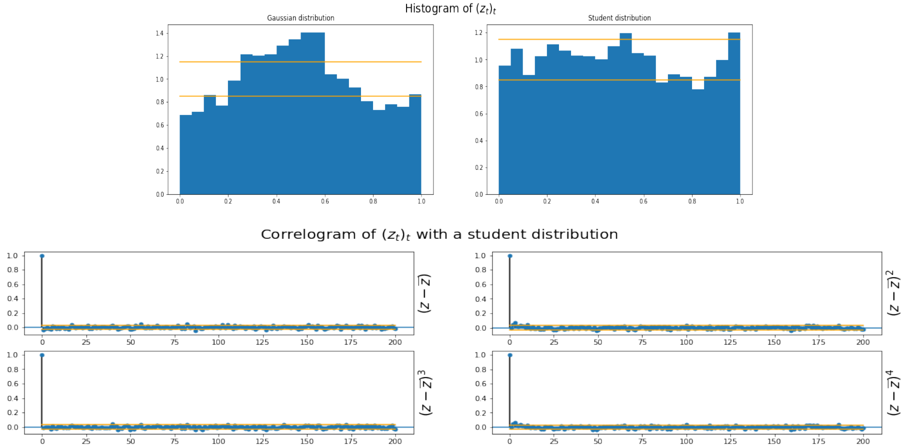
3900 observations pour estimer le modèle et les 3946 suivantes pour l'évaluation.

Application sur données réelles

Le critère BIC sélectionne un modèle GARCH(2,2)

```
                 ARMA(0,0)-GARCH(2,2)
================================================
Distribution               Student
Degree of freedom          6
Method                     Maximum Likelihood
Optimization method        Nelder-Mead
Sucess                     True
Log Likelihood             -5571.390988829947
AIC                        11164.781977659894
BIC                        11225.46929598107
================================================
                       Coeffs
================================================
-----------------Volatility Model---------------
omega       0.01271650622175058
alpha[0]    0.03463121646664299
alpha[1]    0.06934018936033681
beta[0]     0.787582559626089
beta[1]     0.10818839476527395
```

Application sur données réelles

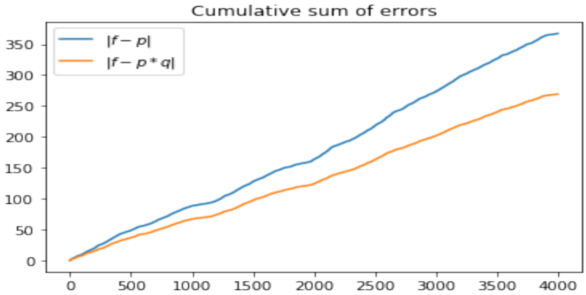


D'après nos critères le modèle Student GARCH(2,2) est une bonne estimation des données de l'Euro Stoxx 50.

Extensions

Extension n^o **1 :** Amélioration de l'estimation de densité.

D'après le *Lemme 1* on a $f_t(y_t) = p_t(y_t)q_t(z_t)$.



La somme des erreurs est plus faible pour $p * q$ que pour p. Autrement dit q a corrigé légèrement l'estimation p.

Extensions

Extension $n^{\mathbf{o}}$ 2 : Estimation de densité multi-dimensionelle.

On consière un Gaussian GARCH(1,1) n-dimensionelle.

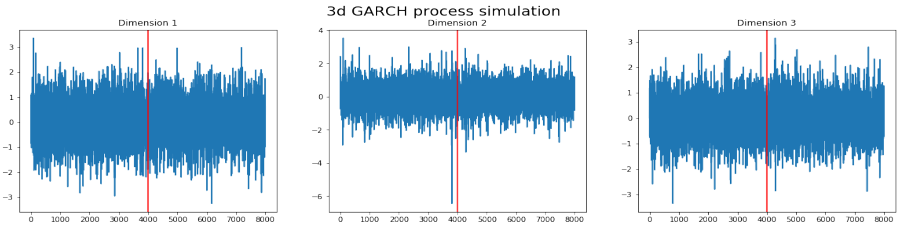
$$\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{n,t} \end{pmatrix} = \begin{pmatrix} h_{1,t} \times \xi_{1,t} \\ \vdots \\ h_{n,t} \times \xi_{n,t} \end{pmatrix} \text{ avec } \xi_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

$$\begin{pmatrix} h^2_{1,t} \\ \vdots \\ h^2_{n,t} \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_n \end{pmatrix} + \begin{pmatrix} \alpha_{1,1} \ldots \alpha_{1,n} \\ \\ \alpha_{n,1} \ldots \alpha_{n,n} \end{pmatrix} \begin{pmatrix} y^2_{1,t-1} \\ \vdots \\ y^2_{n,t-1} \end{pmatrix} + \begin{pmatrix} \beta_{1,1} \ldots \beta_{1,n} \\ \\ \beta_{n,1} \ldots \beta_{n,n} \end{pmatrix} \begin{pmatrix} h^2_{1,t-1} \\ \vdots \\ h^2_{n,t-1} \end{pmatrix}$$

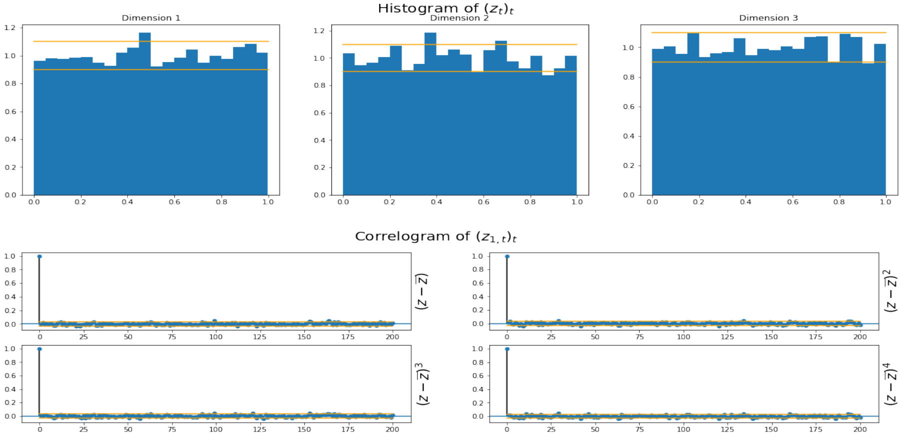
Extensions

Simulation en dimension 3 sur 8000 pas de temps avec

$$\omega = \begin{pmatrix} 0,05 \\ 0,04 \\ 0,06 \end{pmatrix} \ \alpha = \begin{pmatrix} 0,06 & 0,03 & 0,02 \\ 0,02 & 0,10 & 0,04 \\ 0,03 & 0,03 & 0,09 \end{pmatrix} \ \beta = \begin{pmatrix} 0,70 & 0,07 & 0,07 \\ 0,05 & 0,67 & 0,02 \\ 0,06 & 0,07 & 0,60 \end{pmatrix}$$



Extensions



Les 3 suites de $(z_t)_t$ sont bien iid $\mathcal{U}(0,1)$, notre méthode d'évaluation marche aussi en n dimensions.

Conclusion

Conclusion générale

Méthodologie complète pour évaluer des densités de probabilités.

Aucune contrainte sur les modèles donc possibilité d'une immense flexibilité.

Conclusion personelle

Une approche très intéressante.

Ce travail m'a permis de découvrir le domaines des séries temporelles.

Le procédé inverse de la génération des nombres aléatoires.

References

[1] David Ruppert, David S. Matteson. *Statistics and Data Analysis for Financial Engineering*. Seconde édition, Springer 2015.

[2] Bollerslev Tim, *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, volume 31, 1986.

[3] Silverman B.W. *Density Estimation for Statistics and Data Analysis*, 1986.

[4] John Nelder, Roger Mead. *A simplex method for function minimization*. Computer Journal volume 7, 1965.