Ondelettes et calcul variationnel dans une approche non-paramétrique du modèle ARCH

Matthieu Garcin [1] & Clément Goulet [2]

[1] *LabEx REFI et Natixis Asset Management, 21 quai d'Austerlitz, 75013 Paris, prénom.nom [at] polytechnique.edu*
[2] *LabEx REFI et MSE - CES, Université Paris 1 Panthéon-Sorbonne, 106 boulevard de l'hôpital, 75013 Paris, France, prénom.nom [at] univ-paris1.fr*

Résumé. Nous proposons un nouveau modèle pour estimer rendements et volatilité. Notre approche est celle d'un modèle ARCH, dans lequel l'estimation du rendement et de la *news impact curve* sont non-paramétriques. La première est estimée avec des ondelettes et l'estimation de la seconde s'inspire de la littérature relative au bruit multiplicatif en proposant une approche par le calcul variationnel. En faisant cela, nous maximisons la vraisemblance du modèle tout en s'assurant suffisamment de régularité pour la *news impact curve* afin d'éviter un sur-ajustement du modèle. La supériorité de notre approche sur d'autres types de modèles ARCH et quelques modèles GARCH est démontrée empiriquement pour des données de change en haute fréquence et pour des données quotidiennes de l'indice CAC 40, à la fois en précision de l'estimation et en prévision.

Mots-clés. ARCH, estimation non-paramétrique, ondelette, calcul variationnel.

Abstract. We propose a new model to estimate returns and volatility. Our approach is that of an ARCH model, in which the dynamic drift and the news impact curve are non-parametric. The drift is estimated with wavelets and the estimation of the news impact curve, based on a variational approach, is inspired by the literature about multiplicative noise. By doing so, we maximize the likelihood of the model while ensuring sufficient smoothness for the news impact curve so as to avoid overfitting. The superiority of our approach compared to other types of ARCH models as well as some GARCH models is empirically demonstrated for high-frequency exchange rate data and daily data on the CAC 40, both in-sample and out-of-sample.

Keywords. ARCH, non-parametric estimation, wavelet, variational calculus.

1 Introduction

La *news impact curve*, d'Engle et Ng (1993), décrit l'impact d'un rendement en $t-1$ sur la volatilité en t. Selon le modèle ARCH classique, cette courbe est une fonction symétrique décomposée en deux parties linéaires. Cela vient en contradiction avec ce qui est observé dans les séries financières : asymétrie, non-linéarité. Pour y remédier, des évolutions du

Preliminary version – March 25, 2016

modèle ARCH ont été proposées. Outre les approches paramétriques, des approches non-paramétriques consistant en une minimisation d'écarts quadratiques, par exemple entre les deux premiers moments d'un rendement observé et ceux du modèle estimé, ont été testées : avec des noyaux, par Pagan et Schwert (1990), avec une régression polynomiale locale, par Härdle et Tsybakov (1997). La particularité de notre approche, détaillée dans Garcin et Goulet (2015), repose sur les points suivants :

- il n'y a pas d'hypothèse de stationnarité de la série ;
- la tendance est débruitée tout en préservant les particularités[1] grâce à une méthode utilisant des ondelettes, lesquelles sont invoquées pour deux raisons : 1/ la représentation d'un signal régulier dans une base d'ondelettes est parcimonieuse, de sorte qu'un débruitage avec ondelettes revient simplement à éliminer de nombreux coefficients d'ondelette de petite amplitude créés par du bruit, 2/ contrairement à la transformée de Fourier à fenêtre ou, pis, à la transformée de Fourier classique, la contraction et la dilatation de l'ondelette mère pour créer une base d'ondelettes adapte le support de l'ondelette à la résolution, ce qui veut dire qu'une ondelette de plus haute résolution a un support plus petit et capte les phénomènes locaux sans le lissage qu'induirait une transformée de Fourier avec une fenêtre, dont la taille est indépendante de la résolution ;
- l'estimation de la *news impact curve* se fait en maximisant la vraisemblance, grâce à une approche variationnelle.

2 Modèle et estimation

Nous considérons le modèle suivant, pour $t \in \left\{\frac{k}{T}, k \in \{0, ..., T\}\right\}$:

$$\left\{\begin{array}{lcl} y_t & = & x_t + h_t\varepsilon_t \\ h_t & = & g(y_{t-1/T}), \end{array}\right.$$

où y est le rendement observé d'un titre, x est la tendance (non-observée) de ce rendement, $h_t\varepsilon_t$ est la partie bruitée du rendement, avec h_t la volatilité et ε_t une variable aléatoire gaussienne centrée réduite indépendante de x_s et ε_u pour tout u différent de t et pour tout s. Ceci est l'expression générale d'un modèle ARCH. Dans notre approche, nous voulons estimer à la fois x et g avec des méthodes non-paramétriques.

Pour ce qui est de x, nous utilisons des ondelettes, qui permettent de décomposer un signal en la somme d'une structure approximative (les coefficients d'échelle) et de détails (les coefficients d'ondelette), lesquels sont pollués par du bruit. La décompositions d'un

[1] Par particularités, on désigne des exceptions à la régularité, c'est-à-dire des points pour lesquels un signal est höldérien avec un exposant de Hölder local plus faible qu'en d'autres points. Pour une série de rendements, cela peut correspondre à un brusque changement de régime.

signal régulier sur une base d'ondelettes étant parcimonieuse, les coefficients de grande amplitude sont considérés comme reflétant principalement la partie pure du signal alors que les petits coefficients d'ondelette sont censés refléter du bruit et sont donc réduits à zéro. Le seuil permettant d'opérer ce filtrage est défini en fonction de l'amplitude du bruit, soit de h_t et donc de g.

Concernant l'estimation de g, nous nous inspirons de la littérature sur le bruit multiplicatif et en particulier du travail d'Aubert et Aujol (2008). Nous supposons x estimé[2] et nous pouvons alors écrire que :

$$g(y_{t-1/T})\varepsilon_t = y_t - x_t.$$

Dans un premier temps, nous classons les rendements par ordre croissant en définissant θ tel que $y_{\theta(0/T)} \leq y_{\theta(1/T)} \leq ... \leq y_{\theta(T/T)}$.[3] Nous pouvons alors estimer g en maximisant la somme de la vraisemblance à chaque t. On peut écrire cette somme sous forme continue lorsque T tend vers l'infini :

$$\int_0^1 \tilde{\mathcal{L}}(t, \mathcal{G}(t))dt,$$

où nous avons posé $\mathcal{G}$ la fonction définie par $\mathcal{G} : \{k/T, k \in \{0, ..., T\}\} \to \mathbb{R}^+, t \mapsto g(y_t)$ et où la vraisemblance, pour un bruit gaussien, est :

$$\tilde{\mathcal{L}}(t, \mathcal{G}(t)) = C + \log(\mathcal{G}(t)) + \frac{1}{2}\left(\frac{y_{\theta(t)} - x_{\theta(t)}}{\mathcal{G}(t)}\right)^2,$$

où C est une constante. Pour éviter le sur-apprentissage propre aux modèles non-paramétriques, nous contraignons g à être régulière en ajoutant un terme de variation quadratique à minimiser. Ainsi cherchons-nous un extremum de l'expression suivante[4] :

$$\int_0^T \mathcal{L}\left(t, \mathcal{G}(t), \frac{d}{dt}\mathcal{G}(t)\right)dt, \tag{1}$$

[2]Notre méthode est itérative : nous alternons l'estimation de x pour un g déjà estimé et l'estimation de g pour un x déjà estimé. Pour initialiser l'itération, nous estimons grossièrement g par une fonction constante égale à la volatilité estimée du signal. Cette estimation ne dépend pas d'un estimation de x. Concrètement, nous estimons la volatilité par la médiane des coefficients d'ondelette en valeur absolue à l'échelle la plus fine, divisée par 0.6745, qui est le quantile à 75% d'une loi normale centrée réduite. Nous pouvons ensuite utiliser cette estimation grossière de g pour faire une première estimation de x.

[3] Ce choix est motivé par le fait que nous avons des observations irrégulièrement espacées et que nous voulons calculer une intégrale sur cet espace de rendements ainsi discrétisé. Plusieurs approches sont possibles, détaillées dans Garcin et Guégan (2015) et Garcin (2016), parmi lesquelles la technique que nous avons choisie et qui recourt à une intervalle curviligne.

[4] Dans cette expression, nous écrivons une intégrale de $\mathcal{L}$ et une dérivée de $\mathcal{G}$, ce qui est possible dans le cadre asymptotique. Dans le cas discret, cela reviendra simplement à une somme et à un accroissement entre deux instants consécutifs

où la forme fonctionnelle $\mathcal{L}$ est définie par :

$$\mathcal{L} = \mu\tilde{\mathcal{L}} + \frac{1}{2}\left(\frac{d}{dt}\mathcal{G}\right)^2.$$

Les T problèmes d'optimisation deviennent interdépendants du fait de l'ajout du terme de variation quadratique. C'est alors qu'intervient la théorie variationnelle, qui affirme que $\mathcal{G}(t)$ maximise l'intégrale (1) si et seulement si $\mathcal{G}(t)$ est solution de l'équation d'Euler-Lagrange :

$$\forall t \in (0, T),\ 0 = \frac{\partial}{\partial \mathcal{G}}\mathcal{L}\left(t, \mathcal{G}(t), \frac{d}{dt}\mathcal{G}(t)\right) - \frac{d}{dt}\frac{\partial}{\partial \frac{d}{dt}\mathcal{G}}\mathcal{L}\left(t, \mathcal{G}(t), \frac{d}{dt}\mathcal{G}(t)\right),$$

où $\frac{\partial}{\partial \mathcal{G}}\mathcal{L}$ et $\frac{\partial}{\partial \frac{d}{dt}\mathcal{G}}\mathcal{L}$ sont les dérivées partielles de $\mathcal{L}$ selon respectivement le deuxième et le troisième argument.

Nous avons vu que l'estimation de x dépendait de celle de g. Réciproquement, l'estimation de g nécessite l'estimation préalable de x. Nous proposons donc un algorithme itératif en pseudo-code,[5] alternant l'estimation de x et de g :

```
WaveletCoefficients = GetWaveletCoefficients(Y)
Theta = GetOrderingIndexes(Y)
G = Median(WaveletCoefficients)/0.6745
for(i = 0; i < NumberIteration1; i++)
        FilteredWaveletCoefficients =
                GetFilteredCoefficients(WaveletCoefficients, NoiseAmplitude = G)
        X = GetWaveletReconstruction(FilteredWaveletCoefficients)
        for(n = 0; n < NumberIteration2; n++)
                G = Median(WaveletCoefficients)/0.6745
                for(t = 0; t <= T; t++)
                        G(t) = G(t) + delta * (G(t+1) - 2 * G(t) + G(t-1)
                                -mu * (G(t)^2 - (Y(Theta(t)) - X(Theta(t)))^2)/G(t)^3)
```

3 Application

Nous avons appliqué notre modèle sur plusieurs échantillons. Nous nous concentrons ici sur le GBP/USD, en données haute-fréquence espacées de 15 minutes, pour le mois de février 2015. Nous comparons notre modèle (nommé WV-ARCH pour *wavelet variational ARCH* et dont l'estimation apparaît dans la Figure 1) avec les modèles ARCH,

[5] Ces lignes de code ne présentent que l'architecture globale de l'itération. Elles font appel à des fonctions au nom explicite, qui existent dans de nombreux langages, sous un autre nom. Ainsi, *Median* calcule une médiane, *GetWaveletCoefficients* crée un tableau contenant des coefficients d'ondelettes, *GetOrderingIndexes* donne la permutation permettant de trier les coordonnées d'un vecteur par ordre croissant, *GetFilteredCoefficients* applique un filtre à seuil à des coefficients d'ondelette, *GetWaveletReconstruction* fait une transformée inverse par ondelettes.

ARCH-GJR, ARCH-NP, GARCH, GARCH-GJR avec résidus gaussiens (en utilisant le même x_t pour la tendance de tous les modèles). ARCH-GJR et GARCH-GJR modélisent l'asymétrie de manière paramétrique. ARCH-NP est la méthode non-paramétrique par noyaux. La Table 1 indique que seuls les résidus du WV-ARCH vérifient de manière acceptable l'hypothèse de gaussienneté.

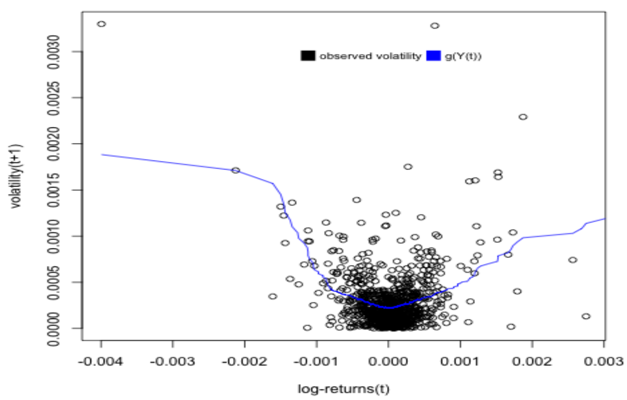

Figure 1: Estimation de g pour les log-rendement de GBP/USD. Les points correspondent à la volatilité instantanée, la ligne à la fonction g estimée après seulement une itération de i dans notre algorithme.

	Moyenne	Skewness	Kurtosis en excès	P-value	Log-vrais.
WV-ARCH	0.0001	-0.0000	-0.2063	0.2962	9814
ARCH	0.0020	0.7552	9.3998	0.0000	9700
ARCH-GJR	-0.0331	1.2233	12.8414	0.0000	9687
ARCH-NP	0.0146	0.3505	8.5028	0.0000	9700
GARCH	0.0000	0.0511	10.7096	0.0000	9747
GARCH-GJR	0.0000	0.0511	10.7096	0.0000	9782

Table 1: Moyenne, skewness et kurtosis en excès des résidus pour GBP/USD. La *p-value* est celle d'un test de normalité de Shapiro. Les modèles sont estimés sur les 1500 premiers log-rendements observés. ARCH, ARCH-GJR, GARCH et GARCH-GJR sont estimés par maximum de vraisemblance.

Les résultats de prévision hors échantillon sont aussi en faveur du WV-ARCH. Nous comparons les modèles avec une fonction QLIKE associée à un test de Diebold et Mariano (2002). Les résultats sont consignés dans la Table 2.

	WV-ARCH	ARCH	ARCH_GJR	ARCH_NP	GARCH	GJR-GARCH
WV-ARCH	-	1.0000	1.0000	0.9984	1.0000	1.0000
ARCH	0.0000	-	0.0204	0.0000	0.9786	0.8852
ARCH-GJR	0.0000	0.9796	-	0.0016	0.9916	0.9572
ARCH-NP	0.0016	1.0000	0.9984	-	0.9998	0.9998
GARCH	0.0000	0.0214	0.0084	0.0002	-	0.0752
GJR-GARCH	0.0000	0.1148	0.0428	0.0002	0.9248	-

Table 2: Risque de première espèce du test de Diebold et Mariano pour les log-rendement de l'USD/GBP : une valeur de 1 indique la suprématie du modèle indiqué à gauche sur le modèle indiqué en haut.

4 Conclusion

Notre approche permet, avec un algorithme relativement simple, de modéliser la volatilité et plus précisément la *news impact curve* en décrivant son asymétrie. Les résultats d'estimation et de prévision de notre approche non-paramétrique sont concluants et ne révèlent pas de sur-estimation. Par ailleurs, cette approche permet d'avoir des résidus gaussiens : le phénomène de queue épaisse, par exemple, ou aussi d'asymétrie, sera décrit par la fonction g non-paramétrique plutôt que par la loi du résidu. Cela permet de contourner la pratique en vogue consistant à améliorer un ARCH en introduisant des distributions plus riches, en termes de possibilités de modélisation, qu'une simple loi gaussienne. Enfin, par parvenir à cela, nous avons eu recours au calcul variationnel, dont l'utilisation en économétrie est nouvelle, à notre connaissance.

Bibliographie

[1] Aubert, G. et Aujol, J.-F. (2008), *A variational approach to removing multiplicative noise*, SIAM Journal on Applied Mathematics, 68, 4 : 925-946
[2] Diebold, F.X. et Mariano, R.S. (2002), *Comparing predictive accuracy*, Journal of Business & Economic Statistics, 20, 1 : 134-144
[3] Engle, R. et Ng, V. (1993), *Measuring and testing the impact of news on volatility*, The Journal of Finance, 48, 5 : 1749-1778
[4] Garcin, M. (2016), *Empirical wavelet coefficients and denoising of chaotic data in the phase space*, à paraître dans Skiadas, C.H. and Skiadas, C. (editors), *Handbook of applications of chaos theory*, CRC/Taylor & Francis
[5] Garcin, M. et Goulet, C. (2015), *A fully non-parametric heteroskedastic model*, document de travail, disponible à *https://halshs.archives-ouvertes.fr/halshs-01244292/*
[6] Garcin, M. et Guégan, D. (2016), *Wavelet shrinkage of a noisy dynamical system with non-linear noise impact*, à paraître dans Physica D: Nonlinear Phenomena, version

préliminaire disponible à *https://halshs.archives-ouvertes.fr/halshs-01244239/*
[7] Härdle, W. et Tsybakov, A. (1997), *Local polynomial estimators of the volatility function in nonparametric autoregression*, Journal of Econometrics, 81, 1 : 223-242
[8] Pagan, A.R. et Schwert, G.W. (1990), *Alternative models for conditional stock volatility*, Journal of Econometrics, 45, 1 : 267-290