Chapitre 3
L'Estimation du Maximum de Vraisemblance
(2020-2021)

Econométrie Financière Avancée

Olivier DARNÉ

L'Estimation du Maximum de Vraisemblance (EMV)

La méthode la plus couramment utilisée pour estimer le vecteur de paramètres inconnus θ de la moyenne et de la variance conditionnelle des modèles GARCH est la méthode d'estimation du maximum de vraisemblance (EMV) (*maximum likelihood estimation*, MLE).

Cette fonction à maximiser est appelée fonction de la log-vraisemblance, notée $L(\theta)$.

Nous présentons la méthode d'EMV sous l'hypothèse de Normalité de la distribution conditionnelle des résidus

Rappels sur l'idée du MV

1. on postule une distribution conditionnelle de ε_t
2. on déduit la log-vraisemblance associée à l'échantillon $L(\varepsilon_t;\theta)$, avec θ le vecteur de paramètres
3. on cherche les estimateurs du MV

$$\widehat{\theta}_{MV} = \arg\max_{\theta\in\Theta} L(\varepsilon_t;\theta)$$

Distribution conditionnelle de ε_t

Hypothèse de Normalité : si $z_t \sim \mathbb{N}(\mu_z, \sigma_z)$ alors les processus ARCH(1) et GARCH(1,1) suivent, conditionnellement à l'ensemble de l'information disponible $\mathcal{F}_{t-1}$, une loi Normale de

- moyenne $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$
- variance $V(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$

Log-vraisemblance $L(\varepsilon_t;\theta)$

La fonction de log-vraisemblance associée à un échantillon de T observations $(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_T)$ de ε_t sous l'hypothèse de Normalité de la loi conditionnelle de ε_t sachant $\mathcal{F}_{t-1}$ s'écrit :

$$\begin{aligned}
L(\varepsilon_t;\theta) &= \prod_{t=1}^{T} f(\varepsilon_t|\mathcal{F}_{t-1}) \\
l(\varepsilon_t;\theta) &= \ln L(\varepsilon_t;\theta) \\
&= \sum_{t=1}^{T} \ln f(\varepsilon_t|\mathcal{F}_{t-1}) = \sum_{t=1}^{T} l_t(\varepsilon_t;\theta) \\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(\ln(2\pi)+\ln(\sigma_t^2)+z_t^2\right) \\
&= -\frac{T}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{T}\ln(\sigma_t^2)-\frac{1}{2}\sum_{t=1}^{T}z_t^2 \\
&= -\frac{T}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{T}\ln(\sigma_t^2)-\frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2}{\sigma_t^2}
\end{aligned}$$

L'EMV $\widehat{\theta}_{MV}$

Si $\partial l(\varepsilon_t;\theta)$ est différentiable, l'EMV de θ, sous l'hypothèse de Normalité, est la solution du système des conditions du premier ordre (CPO) :

$$\mathbf{G}(\theta) = 0 \qquad \text{où} \qquad \mathbf{G}(\theta) = \frac{\partial l(\varepsilon_t;\theta)}{\partial\theta}$$

pour autant que les conditions de second ordre (CSO) sont satisfaites, cad que la matrice des dérivées secondes, évaluée à la solution des CPO, est définie négative.

NB: $\mathbf{G}(\theta)$ est appelé la fonction score ou vecteur gradient

Soit $\widehat{\theta}$ la solution des CPO, alors $\widehat{\theta}$ est l'EMV si

$$\begin{aligned} \mathbf{G}(\widehat{\theta}) = \frac{\partial l(\varepsilon_t;\theta)}{\partial\theta}\Big|_{\theta=\widehat{\theta}} &= 0 \\ \mathbf{H}(\widehat{\theta}) = \frac{\partial^2 l(\widehat{\theta})}{\partial\widehat{\theta}\partial\widehat{\theta}'} &< 0 \end{aligned}$$

avec $\mathbf{H}(\widehat{\theta})$ la matrice Hessienne évaluée à $\widehat{\theta}$

Résoudre analytiquement les CPO, comme pour la régression linéaire, n'est généralement pas possible (pas de formule explicite).

L'EMV $\widehat{\theta}$ peut être obtenu en utilisant des méthodes d'optimisation numérique, implémentées dans les logiciels économétriques.

Algorithmes d'optimisation numérique

Les méthodes d'optimisation numérique itératives doivent être utilisées pour obtenir une solution des CPO et donc l'EMV $\widehat{\theta}$

Le principe de ces algorithmes d'optimisation itératifs est de

- prendre une série initiale de valeurs des paramètres, notée $\theta_{(0)}$ (étape d'initialisation)
- calculer à partir de ces valeurs initiales pour obtenir une meilleure série de valeurs de paramètres, notée $\theta_{(1)}$ (étape 1ère itération)
- répéter le processus jusqu'à ce que la fonction de log-vraisemblance ne soit plus améliorée (étape itérative)

McCullough et Renfro (1999) et Brooks et alii (2001) discutent de plusieurs questions pratiques à considérer lors de la maximisation de la fonction de vraisemblance et montrent que les paramètres d'initialisation peuvent influencer les estimations numériques des paramètres des modèles GARCH

Par exemple, les valeurs de départ pour les paramètres du modèle GARCH(1,1) $\theta = (\omega, \alpha, \beta)$ doivent être choisies et une initialisation de ε_t^2 et σ_t^2 doit être fournie.

En règle générale, cela se fait à l'aide d'un ensemble d'initialisations par défaut dans les logiciels, de sorte que l'utilisateur peut ignorer quelles peuvent être les hypothèses concernant les valeurs plausibles qui ont été établies.

La solution par défaut habituelle pour les initialisations de paramètres est de définir

- des valeurs des paramètres dans la moyenne conditionnelle égale à celles estimées à l'aide d'une régression (initiale) des MCO
- des valeurs des paramètres dans l'équation de la variance conditionnelle égale à zéro

Une initialisation souvent utilisée pour la variance conditionnelle (σ_t^2) est

$$\sigma_t^2 = \varepsilon_t^2 = (1/T)\sum_{j=1}^{T} e_j^2$$

avec $t \leq 0$ et où les e_j représentent les résidus d’une régression linéaire de la variable dépendante sur une constante.

Une fois que la log-vraisemblance est initialisée, elle peut être maximisée en utilisant des techniques d’optimisation numérique

Les méthodes d’optimisation les plus populaires utilisent des matrices Hessiennes exactes ou approximatives.

Algorithme de Newton-Raphson

Une approche classique basée sur la matrice Hessienne exacte est l'algorithme de Newton-Raphson. La procédure itérative est donnée par

$$\widehat{\theta}_{(i+1)} = \widehat{\theta}_{(i)} - \lambda_i \mathbf{H}(\widehat{\theta}_{(i)})^{-1} \mathbf{G}(\widehat{\theta}_{(i)})$$

- $\widehat{\theta}_{(i)}$: un vecteur des paramètres estimés du modèle évalué à l'itération i
- $\lambda_{(i)}$: un scalaire (appelé étape) qui peut être fixe ou variable
- $\mathbf{G}(\widehat{\theta}_i)$: le vecteur gradient (ou fonction score) de la fonction de log-vraisemblance

$$\mathbf{G}(\widehat{\theta}_{(i)}) = \frac{\partial l(\widehat{\theta}_{(i)})}{\partial \widehat{\theta}_{(i)}}$$

- $\mathbf{H}(\widehat{\theta}_{(i)})$: la matrice Hessienne (ou une approximation appropriée de celle-ci) de la fonction de log-vraisemblance, toutes évaluées à l'itération i

$$\mathbf{H}(\widehat{\theta}_{(i)}) = \frac{\partial^2 l(\widehat{\theta}_{(i)})}{\partial \widehat{\theta}_{(i)} \partial \widehat{\theta}'_{(i)}}$$

L'algorithme itère jusqu'à $\widehat{\theta}_{(i+1)} \simeq \widehat{\theta}_{(i)}$, avec un seuil de tolérance.

La convergence est réalisée lorsque

$$\widehat{\theta}_{(i+1)} - \widehat{\theta}_{(i)} = -\mathbf{H}(\widehat{\theta}_{(i)})^{-1}\mathbf{G}(\widehat{\theta}_{(i)}) \simeq 0$$

Cette condition sera respectée si seulement si

$$\mathbf{G}(\widehat{\theta}_{(i+1)}) \simeq \mathbf{G}(\widehat{\theta}_{(i)}) \simeq 0$$

car les matrices Hessiennes $\mathbf{H}(\widehat{\theta}_{(i+1)})$ et $\mathbf{H}(\widehat{\theta}_{(i)})$ sont définies négatives.

C'est exactement la condition qui définit l'EMV $\widehat{\theta}$ tel que $\widehat{\theta}_{(i+1)} \simeq \widehat{\theta}$ lors de l'itération finale

Algorithme du scoring

Engle (1982) utilise pour estimer les paramètres du modèle ARCH(q) l'algorithme du scoring qui est proche de l'algorithme de Newton-Raphson

La différence est que l'algorithme du scoring dépend de la valeur espérée de la dérivée seconde et non de la dérivée seconde observée

$$\widehat{\theta}_{(i+1)} = \widehat{\theta}_{(i)} + E\left(\mathbf{H}(\widehat{\theta}_{(i)})\right)^{-1}\mathbf{G}(\widehat{\theta}_{(i)})$$

Algorithme de Berndt-Hall-Hall-Hausman (BHHH)

Une autre méthode basée sur la matrice Hessienne approximée est l'algorithme de Berndt-Hall-Hall-Hausman (BHHH) (Berndt et alii, 1974) qui n'utilise que des dérivées premières.

L'algorithme de BHHH est semblable à l'algorithme de Newton-Raphson mais au lieu d'utiliser la matrice Hessienne des dérivées secondes $\mathbf{H}(\widehat{\theta}_i)$, il se base sur une approximation formée par la somme du produit extérieur des vecteurs de gradient (*outer product of the gradient*, OPG) pour la contribution de chaque observation à la fonction objective.

La matrice Hessienne peut être approximée par

$$-\mathbf{H}(\widehat{\theta}_{(i)}) \approx \mathbf{B}(\widehat{\theta}_{(i)}) = \sum_{t=1}^{T} \mathbf{G}_t(\widehat{\theta}_{(i)})\mathbf{G}_t(\widehat{\theta}_{(i)}) = \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta_{(i)}} \frac{\partial l_t}{\partial \theta'_{(i)}}$$

où $\mathbf{G}_t(\widehat{\theta})$ est le gradient de la fonction de log-vraisemblance pour chaque observation.

L'algorithme BHHH peut être modifié en employant la correction de Marquardt qui ajoute une matrice de correction à la somme des vecteurs OPG (approximation Hessienne) de la manière suivante :

$$\mathbf{B}(\widehat{\theta}_i) = \sum_{t=1}^{T} \mathbf{G}_t(\widehat{\theta}_i)\mathbf{G}_t(\widehat{\theta}_i) - aI$$

où I est la matrice identité, et a un nombre positif choisi par l'algorithme.

Avantage / inconvénient

- La vitesse de calcul est augmentée en ne calculant pas la matrice Hessienne réelle à chaque itération pour chaque pas de temps, mais l'approximation peut être faible lorsque la fonction de log-vraisemblance est loin de son maximum, ce qui nécessite plus d'itérations pour atteindre l'optimum
- La correction de Marquardt a pour effet de "pousser" les estimations des paramètres en direction du vecteur de gradient $\Rightarrow$ meilleure performance lorsque l'estimation est loin de l'optimum, et permet les calculs lorsque la matrice Hessienne est presque singulière

Algorithme de Quasi-Newton

Une approche alternative basée sur la matrice Hessienne approximée est la famille des méthodes de Quasi-Newton qui consiste à construire une estimation de la matrice Hessienne à chaque itération, à partir d'une estimation initiale.

L'inverse de la matrice Hessienne est donné par

$$\mathbf{H}(\widehat{\theta}_i) = \mathbf{H}(\widehat{\theta}_{i-1}) + \mathbf{C}_{i-1}$$

où **C** est une matrice de correction.

Les algorithmes de Quasi-Newton ne diffèrent que par leur choix de cette matrice de correction, et les plus largement utilisés sont

- l'algorithme DFP proposé par Davidson (1959) et Fletcher et Powell (1963)
- l'algorithme BFGS développé par Broyden (1970), Fletcher (1970), Goldfarb (1979) et Shanno (1970)

Méthodes de programmation non linéaire contrainte

Généralement, les modèles GARCH sont représentés comme des problèmes d'optimisation non contraints.

Cependant, si la maximisation de la fonction de log-vraisemblance doit être réalisée sous des contraintes non-linéaires dans le vecteur des paramètres inconnus alors les méthodes de programmation non linéaire contrainte (NLP, *constrained nonlinear programming*) peuvent être utilisées, comme les algorithmes de programmation quadratique séquentielle (SQP, *sequential quadratic programming*).

Un exemple de contraintes non linéaires est l'imposition des limites inférieure et supérieure sur les paramètres, tel que

- la contrainte de stationnarité $0 < \alpha + \beta < 1$ dans le modèle GARCH(1,1)
- la contrainte $-1 < \gamma < 1$ dans le modèle GJR-GARCH(1,1)

Propriétés de l'EMV

Si les hypothèses sont correctes, l'EMV est l'estimateur le plus efficient asymptotiquement, au sens qu'il est non-biaisé asymptotiquement et possède la plus petite variance.
Intuitivement, le MV utilise au mieux l'information des données pour estimer les paramètres

- l'EMV est convergent

$$\widehat{\theta} \underset{p}{\rightarrow} \theta_0$$

- l'EMV est asymptotiquement efficace

$$V(\widehat{\theta}) = I_T^{-1}(\theta_0)$$

- l'EMV est asymptotiquement Normalement distribué

$$\sqrt{T}(\widehat{\theta} - \theta_0) \underset{d}{\rightarrow} \mathbb{N}(0, I_T^{-1}(\theta_0))$$

Estimateurs de la matrice de variance-covariance

Il est important de connaître la manière dont sont calculés les écart-types des estimateurs des paramètres.

Ces différentes méthodes de calcul des écart-types des coefficients des modèles GARCH peuvent conduire à différentes estimations de ces écart-types, et auront donc des implications importantes sur la significativité de ces paramètres.

La matrice de variance-covariance (VCV) des estimateurs de paramètres, lorsque les variables sont supposées être Normalement distribuées, est calculée comme l'inverse de la matrice d'information de Fisher, $\mathbf{I}(\widehat{\theta})$, définie par

$$\widehat{V} = \mathbf{Q}(\widehat{\theta}) = \mathbf{I}(\widehat{\theta})^{-1}$$

C'est une matrice diagonale dans le cas d'une densité Normale, mais dans le cas contraire, il peut s'agir d'une matrice non diagonale compliquée.

Dès lors, lorsque l'EMV est appliquée à des distributions non Normales, les calculs de la matrice de variance-covariance des estimateurs peuvent être difficiles.

En général, deux estimateurs de la matrice d'information de Fisher sont calculés

- Le premier estimateur est la valeur espérée des dérivées premières de la fonction de log-vraisemblance.
  On l'appelle l'estimateur BHHH ou estimateur OPG :

$$\mathbf{I}(\widehat{\theta}) \approx E\left[\frac{\partial lnl}{\partial \theta}\frac{\partial lnl}{\partial \theta'}\right] \qquad \Rightarrow \qquad \widehat{V}_{OPG} = \left(E\left[\frac{\partial lnl}{\partial \theta}\frac{\partial lnl}{\partial \theta'}\right]\right)^{-1}$$

- Le deuxième estimateur, appelé estimateur basé sur la matrice Hessienne, est la valeur espérée négative des dérivées secondes de la fonction de log-vraisemblance :

$$\mathbf{I}(\widehat{\theta}) = -E\left[\frac{\partial^2 lnl}{\partial \theta \partial \theta'}\right] = -E\left[\mathbf{H}(\widehat{\theta})\right] \qquad \Rightarrow \qquad \widehat{V}_H = \left(-E\left[\mathbf{H}(\widehat{\theta})\right]\right)^{-1}$$

Ces deux estimateurs de la matrice d'information de Fisher sont asymptotiquement équivalents, mais ils pourraient donner des résultats différents en échantillons finis.

Des études suggèrent que dans les échantillons de petite taille ou de taille moyenne, l'estimateur basé sur la matrice Hessienne $\widehat{V}_H$ est préférable (Greene, 2007).

Cependant, dans la plupart des cas, l'estimateur OPG $\widehat{V}_{OPG}$ sera le plus facile à calculer.

Si la fonction de log-vraisemblance est maximisée en utilisant des algorithmes basés sur la matrice Hessienne, la matrice d'information de Fisher peut être obtenue comme un sous-produit de la procédure d'optimisation, comme pour les méthodes de Newton-Raphson et BHHH.

D'un autre côté, si l'on emploie des méthodes Quasi-Newton, telle que la méthode BFGS, il n'est pas garanti que la matrice Hessienne converge vers sa valeur exacte à l'approche de l'optimum.

L'utilisation de ces méthodes pour obtenir la matrice d'information de Fisher estimée peut donc être erronée.

L'hypothèse des innovations standardisées Normalement distribuées est rarement vérifiée par les données, spécialement en finance avec des distributions fortement leptokurtiques (*fat tail*) $\Rightarrow$ la fonction de vraisemblance est formellement mal spécifiée

Ce constat a motivé l'utilisation d'autres hypothèses de distribution, comme, par exemple, les distributions de Student-*t* ou de Skewed-Student-*t* (Bollerslev, 1987 ; Lambert et Laurent, 2001).

Le MV nécessite de choisir la distribution de probabilité. Quelles sont donc les conséquences d'une hypothèse erronée sur la distribution de probabilité des ε_t

- dans certains cas, l'EMV n'est même pas convergent
- dans d'autres cas, il est toujours convergent et asymptotiquement Normal mais la matrice VCV des paramètres est différente
  L'EMV garde ses propriétés intéressantes même si la distribution supposée est fausse, pour autant que $E[\mathbf{G}_t(\widehat{\theta})] = 0$

Estimateurs du pseudo maximum de vraisemblance (QML)

Alternativement, les EMV basées sur la densité Normale s'interprétent comme des estimateurs du pseudo maximum de vraisemblance (PMV) (*quasi-maximum likelihood*, QML).

L'idée générale des estimateurs du PMV (EPMV) consiste à démontrer que si l'on commet une erreur sur la distribution conditionnelle des résidus en utilisant à tort une log-vraisemblance fondée sur une loi Normale, l'EMV ainsi obtenu peut tout de même être convergent si la vraie loi des résidus appartient à la même classe de loi que la loi Normale (Gourieroux et Montfort, 1989).

L'EPMV est l'EMV calculé pour une fonction de vraisemblance construite pour une distribution de probabilité (éventuellement fausse) où $E[\mathbf{G}_t(\widehat{\theta})] = 0$, notée $\widehat{\theta}_Q$

La fonction de vraisemblance est alors appelée fonction de pseudo-vraisemblance

Algorithmes d'optimisation numérique

Bollerslev et Wooldridge (1992) ont montré que la maximisation de la fonction de log-vraisemblance Normale donne des EPMV du vecteur de paramètre θ qui sont convergentes et asymptotiquement Normalement distribuées, même lorsque la distribution des résidus standardisés z_t n'est pas Normale.

Sous des conditions de régularité appropriées, les propriétés de l'EPMV sont :

- l'EPMV est asymptotiquement convergent

$$\widehat{\theta}_Q \theta_0$$

- l'EPMV est asymptotiquement Normalement distribuées

$$\sqrt{T}(\widehat{\theta}_Q - \theta_0) \underset{d}{\rightarrow} \mathbb{N}(0, V_Q)$$

Néanmoins, les conditions de régularité appropriées ont seulement été établies pour certains modèles GARCH

- modèles GARCH (Lee et Hansen, 1994 ; Berkes et alii, 2003 ; Francq et Zakoian, 2004)
- modèles GJR-GARCH (Hamadeh et Zakoian, 2011)

Bollerslev et Wooldridge (1992) ont établi une matrice de variance-covariance asymptotique pour les EPMV qui est robuste à la non-Normalité conditionnelle.

Cette matrice VCV V_Q est estimée en utilisant à la fois les matrices de dérivées du second ordre et des dérivées premières des produits externes (OPG) :

$$\widehat{V}_Q = \mathbf{Q}(\widehat{\theta}) = \mathbf{H}(\widehat{\theta}_Q)^{-1}\mathbf{B}(\widehat{\theta}_Q)\mathbf{H}(\widehat{\theta}_Q)^{-1}$$

- $\widehat{\theta}_Q$: l'EPMV de θ, également appelé estimateur "sandwich"
- les écart-types des coefficients sont nommés les écart-types robustes de Bollerslev-Wooldridge

Sous une spécification correcte d'une densité Normale, cad la vraie loi sous-jacente est Normale, l'EPMV est équivalent aux autres estimateurs de la matrice VCV de l'EMV, car il donne asymptotiquement l'inverse de la matrice d'information de Fisher (donc, bloc diagonale) :

$$V(\widehat{\theta}_Q - \theta_0) = \mathbf{H}(\widehat{\theta}_Q)^{-1}$$

car $\mathbf{H}(\widehat{\theta}_Q) = \mathbf{B}(\widehat{\theta}_Q)$

L'EPMV a l'avantage de fournir une inférence statistique correcte même si le choix de la distribution est fausse et donne des estimations convergentes

L'EMPV est généralement proche de l'EMV exacte pour les déviations symétriques par rapport à la Normalité.
Cependant, Engle et Gonzales-Rivera (1991) ont montré que la perte d'efficacité de l'EPMV peut être assez élevée pour les distributions non symétriques.

Figure: Estimation des modèles GARCH sur la série ajustée des outliers avec la VCV Hessienne ($\widehat{V}_H$)

	Coefficient		|t-value| > 1.64	persistance	half-life	log-likehood	Akaike	HQ
GARCH	omega > 0	0.0039	1.83	0.9894	65.04	-1024.50	1.721	1.728
	alpha ≥ 0	0.0615	4.13					
	beta ≥ 0	0.9279	50.0					
	alpha + beta < 1	0.9894						
EGARCH	omega	-0.0127	-1.64	0.9878	56.47	-1023.1	1.721	1.729
	beta < 1	0.9878	154.7					
	theta1 (gamma)	-0.0354	-2.31					
	theta2 (alpha)	0.1337	4.79					
GJR-GARCH	omega > 0	0.0035	1.72	0.9901	69.67	-1022.6	1.720	1.728
	alpha	0.039	2.54					
	beta ≥ 0	0.9329	51.2					
	gamma	0.0364	1.94					
	alpha + gamma ≥ 0	0.0754						
	(gamma/2) < 1	0.9901						
TGARCH	omega	0.0092	2.03	1.1469	5.06	-1030.3	1.733	1.741
	alpha	0.0838	5.22					
	beta ≥ 0	0.9188	51.0					
	gamma (eta)	0.1895	1.82					
	constraint	1.1469						
IGARCH	omega > 0	0.0016	1.77			-1026.00	1.722	1.727
	alpha ≥ 0	0.0625	4.34					
	beta	0.9375						
Riskmetrics	alpha	0.06				-1029.9	1.725	1.727
	beta	0.94						

Figure: Estimation des modèles GARCH sur la série ajustée des outliers avec la VCV robuste $(\widehat{V}_Q)$

	Coefficient		|t-value| > 1.64	persistance	half-life	log-likehood	Akaike	HQ
GARCH	omega > 0	0.0039	1.52	0.9894	65.04	-1024.50	1.721	1.728
	alpha ≥ 0	0.0615	3.17					
	beta ≥ 0	0.9279	38.3					
	alpha + beta < 1	0.9894						
EGARCH	omega	-0.0127	-1.51	0.9878	56.47	-1023.1	1.721	1.729
	beta < 1	0.9878	131.9					
	theta1 (gamma)	-0.0354	-1.75					
	theta2 (alpha)	0.1337	3.73					
GJR-GARCH	omega > 0	0.0035	1.39	0.9901	69.67	-1022.6	1.720	1.728
	alpha	0.039	1.90					
	beta ≥ 0	0.9329	38.0					
	gamma	0.0364	1.51					
	alpha + gamma ≥ 0	0.0754						
	(gamma/2) < 1	0.9901						
TGARCH	omega	0.0092	1.62	1.1469	5.06	-1030.3	1.733	1.741
	alpha	0.0838	3.50					
	beta ≥ 0	0.9188	34.5					
	gamma (eta)	0.1895	1.18					
	constraint	1.1469						
IGARCH	omega > 0	0.0016	1.51			-1026.00	1.722	1.727
	alpha ≥ 0	0.0625	3.71					
	beta	0.9375						
Riskmetrics	alpha	0.06				-1029.9	1.725	1.727
	beta	0.94						

Table: Les EMV des logiciels économétriques sous leur option de défault.

Softwares Packages	Ox 7.2 G@RCH 7	Eviews 9	Matlab 13a MFE	Matlab 13a Ek	Gauss 17 Fanpac 3.0	R 3.4 rugarch 1.4-0	Stata 14	Gretl 2018a GIG 2.21
Initialization for cond. variance	sample uncond. variance	backcast exponential smoothing	local average	sample uncond. variance	sample uncond. variance	sample uncond. variance	expected uncond. variance	sample uncond. variance
Optimization algorithm	BFGS	BFGS with Marcquart correction	BFGS	SQP	BFGS	Augmented Lagrange based SQP	BHHH & BFGS	BFGS
Derivatives	numeric	numeric & analytic	numeric	numeric	numeric	numeric	numeric	numeric
Variance estimator	$\widehat{V}_Q$	$\widehat{V}_{OPG}$	$\widehat{V}_H$ & $\widehat{V}_Q$	$\widehat{V}_{OPG}$	$\widehat{V}_H$	$\widehat{V}_H$ & $\widehat{V}_Q$	$\widehat{V}_H$ ou $\widehat{V}_{OPG}{}^*$	$\widehat{V}_Q$
Convergence criteria	0.0001	0.0001	0.0001	0.000001	0.0001	0.000001	0.0001	depends on computer power
Max. iterations	1000	500	500	400	1000	400	16000	500

Les différentes options lors de l'estimation des modèles GARCH peuvent donner des résultats différents, surtout en petits échantillons (cf. Annexe Chapitre 3).