

Estimation par régression artificielle du processus
GARCH(1,1).
Influence de la valeur initiale.

BONO Pierre-Henri
D.E.A. économie mathématique et économétrie
GREQAM Marseille

9 septembre 2003

1 Introduction

Les modèles à erreurs $(G)ARCH$ sont connus pour être difficiles à estimer. Malgré cela la plupart des logiciels d'économétrie, de traitement des données ou plus généralement d'aide à la décision proposent des modules d'estimation de modèle $GARCH$. Mais qu'en est-il vraiment de la précision des résultats fournis ? Un premier élément de réponse a été fournis par Brook, Burke, et Persaud [1, 2001]. Pour une même série de données, les principaux progiciels d'économétrie commerciaux donnent des résultats assez différents. Une hypothèse avancée par Brook, Burke, et Persaud pour justifier ces différences dans les valeurs est que les logiciels en questions n'utilisent pas la même initialisation pour la variance conditionnelle. Il était donc tentant de vérifier quelle est vraiment l'influence de la valeur initiale de la variance conditionnelle. En effet si dans la théorie asymptotique Lee et Hansen [5, 1994] ont montré que le choix de la valeur initiale de la variance n'a aucune influence sur l'efficacité et la normalité asymptotique de l'estimateur, qu'en est il en échantillon fini ? Plusieurs méthodes ont été proposées depuis l'apparition des modèles $ARCH$ (Engle 1982) pour le choix des valeurs initiales, six d'entre elles, plus ou moins simples, ont été retenues. On montrera dans ce qui suit que l'influence de la valeur initiale est "infinitésimale" dès que la taille de l'échantillon dépasse 200 éléments.

Le deuxième but de ce mémoire est de présenter une régression artificielle pour l'estimation par maximum de vraisemblance du modèle $GARCH(1,1)$. Les régressions artificielles ont été développées, sur la base de fonctions zéro élémentaires, en exploitant le fait que la distribution normale est complètement caractérisée par ses deux premiers moments. Ces régressions artificielles peuvent être utilisées à la base d'algorithmes numériques pour des estimations par maximum de vraisemblance de modèles avec des éléments distribués selon une loi normale¹. On montre que cette technique correspond à l'estimateur du "score efficace", l'inverse de la matrice d'information est utilisé pour la direction dans l'algorithme du gradient.

2 Le modèle $GARCH(1,1)$

Soit un modèle de régression linéaire simple ou les erreurs suivent un processus $GARCH(1,1)$ introduit par Bollerslev en 1986 [1] :

$$Y = X\beta + u \quad (2.1)$$

avec

$$u_t = \sigma_t v_t \quad \text{où } v_t \sim \text{IID}N(0, 1)$$

et

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_t^2 + \delta \sigma_t^2 \quad (2.2)$$

¹Pour en savoir plus sur les régression artificielle se reporter au travail de R. Davidson

où Y est un vecteur $(n \times 1)$, X est une matrice $(n \times k)$ et β un vecteur $(k \times 1)$. On notera θ le vecteur $(k + 3 \times 1)$ des paramètres du modèle 2.1

$$\theta = (\beta, \alpha_0, \alpha_1, \delta)^T$$

3 Le maximum de vraisemblance

3.1 Fonction de log-vraisemblance

Soit θ_0 le vrais paramètre du PGD, on définit l'espace compact suivant

$$\Theta = \{\theta : \alpha_0 > 0; \alpha_1 > 0; 0 < \delta < 1\}$$

On assume le fait que $\theta_0 \in \Theta$, ce qui implique que u_t est un processus *GARCH* stricte, on supprime la possibilité à u_t d'être un processus *ARCH* pure ou bien un bruit blanc. On définit la variable $z_t = u_t/\sigma_t$. L'estimation des modèles *GARCH* est souvent faite sous l'hypothèse que $z_t \sim i.i.d.N(0, 1)$, la vraisemblance est alors facilement spécifiable. On suivra cette pratique en assumant que c'est la vraisemblance gaussienne qui est utilisée pour former l'estimateur. La contribution de l'observation t à la fonction de log-vraisemblance à la forme :

$$l_t(\theta) = -\frac{1}{2} \ln(2\pi\sigma_t^2) - \frac{1}{2} \frac{u_t^2}{\sigma_t^2}$$

Pour un échantillon de taille T la **fonction de log-vraisemblance** est :

$$L_T = \frac{1}{T} \sum_{t=1}^T l_t(\theta)$$

3.2 Consistence et normalité asymptotique

Lee et Hansen [5, (1994)] ont montré que, sous certaines conditions sur les résidus normalisés, z_t , l'estimateur par quasi-maximum² de vraisemblance d'un processus *GARCH(1,1)* gaussien est asymptotiquement consistant et normale. Ce résultat asymptotique est valable quelque soit l'expression de la variance conditionnelle initiale σ_u . On exprime maintenant le gradient de la vraisemblance.

²On parle d'estimateur par quasi-maximum de vraisemblance si la densité de la vraisemblance n'est pas la bonne ou en d'autres mots si le modèle est mal spécifié.

3.3 Gradient du modèle GARCH(1,1)

A partir de la forme d'une contribution à la fonction de log-vraisemblance on peut calculer les dérivées partielles de l_t par rapport à θ . Pour cela on utilisera les dérivées partielles suivantes

$$\begin{aligned}\frac{\partial l_t}{\partial u_t} &= \frac{u_t}{\sigma_t^2} \\ \frac{\partial l_t}{\partial \sigma_t^2} &= -\frac{1}{2\sigma_t^2} + \frac{u_t^2}{2\sigma_t^4} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4}\end{aligned}$$

de même on utilisera

$$\begin{aligned}\frac{\partial u_t}{\partial \beta_i} &= -1 \\ \frac{\partial u_t}{\partial \alpha_0} &= 0 \\ \frac{\partial u_t}{\partial \alpha_1} &= 0 \\ \frac{\partial u_t}{\partial \delta} &= 0\end{aligned}$$

On a besoin aussi des dérivées de σ_t^2 par rapport à θ . Pour cela on réécrit la récurrence de la variance conditionnelle de la manière suivante

$$\sigma_t^2 - \delta\sigma_{t-1}^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

On remarque que cette expression a la même forme que la récurrence qui définit un processus $AR(1)$, elle peut donc être résolue de la même manière

$$\sigma_t^2 = \sum_{s=0}^{t-2} \delta^s (\alpha_0 + \alpha_1 u_{t-s-1}^2) + \delta^{t-1} \sigma_1^2$$

En calculant la somme du premier terme explicitement et en changeant l'index de somme du second terme on obtient

$$\sigma_t^2 = \frac{\alpha_0(1 - \delta^{t-1})}{1 - \delta} + \alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \sigma_1^2 \quad (3.1)$$

Si l'on avait commencé la récurrence à l'observation 0 on aurait

$$\sigma_t^2 = \frac{\alpha_0(1 - \delta^t)}{1 - \delta} + \alpha_1 \sum_{s=1}^t \delta^{s-1} u_{t-s}^2 + \delta^t \sigma_0^2 \quad (3.2)$$

On va réécrire cette équation de manière à pouvoir la comparer avec la récurrence 3.1 commençant à l'observation 1

$$\begin{aligned}\sigma_t^2 &= \frac{\alpha_0(1-\delta^t)}{1-\delta} + \alpha_1 \sum_{s=1}^t \delta^{s-1} u_{t-s}^2 + \delta^t \sigma_0^2 \\ &= \frac{\alpha_0(1-\delta^{t-1})}{1-\delta} + \frac{\alpha_0(\delta^{t-1}-\delta^t)}{1-\delta} + \alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \\ &\quad + \alpha_1 \delta^{t-1} u_0^2 + \delta^{t-1} \sigma_0^2 + (\delta^t - \delta^{t-1}) \sigma_0^2\end{aligned}$$

On remplace maintenant u_0^2 par son espérance σ_0^2

$$\begin{aligned}\sigma_t^2 &= \frac{\alpha_0(1-\delta^{t-1})}{1-\delta} + \alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \\ &\quad + \delta^{t-1} (\alpha_0 + \alpha_1 \sigma_0^2 + \sigma_0^2 - (1-\delta) \sigma_0^2)\end{aligned}$$

On voit que commencer avec l'indice 0 est équivalent à commencer à l'indice 1 avec

$$\sigma_1^2 = \alpha_0 + \sigma_0^2(\alpha_1 + \delta) \quad (3.3)$$

On notera que pour $\sigma_0^2 = \alpha_0/(1-\alpha_1-\delta)$ ce choix est équivalent. Bien qu'il soit plus simple et aussi plus logique de commencer à l'observation 1, en effet on a qu'une seule inconnue, beaucoup de travaux sur l'estimation des processus *GARCH* utilise l'observation 0 comme point de départ. Le calcul ci-dessus nous permettra lors des simulations de garder le même code et de changer seulement la valeur de l'observation initiale selon que l'on veuille commencer à l'observation 1 où 0 en utilisant la formule 3.3. Dans la suite on notera σ_u la variance initiale.

On exprime les dérivées partielles dont on a besoin pour le calcul du gradient

$$\begin{aligned}\frac{\partial \sigma_t^2}{\partial \beta_i} &= -2\alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)i} \\ \frac{\partial \sigma_t^2}{\partial \alpha_0} &= \frac{1-\delta^{t-1}}{1-\delta} + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_0} \\ \frac{\partial \sigma_t^2}{\partial \alpha_1} &= \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \\ \frac{\partial \sigma_t^2}{\partial \delta} &= -\frac{\alpha_0(t-1)\delta^{t-2}}{1-\delta} + \frac{\alpha_0(1-\delta^{t-1})}{(1-\delta)^2} \\ &\quad + \alpha_1 \sum_{s=1}^{t-1} (s-1)\delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \delta}\end{aligned}$$

Avec ces expressions, on peut maintenant exprimer la contribution au **gradient** de la log-vraisemblance.

$$\begin{aligned}
\frac{\partial l_t}{\partial \beta_i} &= \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \beta_i} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \beta_i} = \frac{X_{ti} u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \sum_{s=1}^{t-1} \delta^{s-1} X_{ti} u_{t-s}^2 & (3.4) \\
\frac{\partial l_t}{\partial \alpha_0} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_0} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\frac{1 - \delta^{t-1}}{1 - \delta} + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_0} \right) \\
\frac{\partial l_t}{\partial \alpha} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right) \\
\frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(-\frac{\alpha_0 (t-1) \delta^{t-2}}{1 - \delta} + \frac{\alpha_0 (1 - \delta^{t-1})}{(1 - \delta)^2} \right. \\
&\quad \left. + \alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \delta} \right)
\end{aligned}$$

Il est bien sur impossible d'utiliser ces dernières expressions sans connaître la valeur de σ_u^2 .

4 La maximisation de la vraisemblance

Bien que les algorithmes de maximisation (ou minimisation) de fonctions soient nombreux et variés, des aspects fondamentaux sont partagés. On peut résumer ces propriétés communes comme

- Le processus est itératif où en d'autres termes il existe un point de départ et une règle de récurrence (de mis a jours). A chaque itération, le point de départ pour la prochaine itération est en fait le résultat de l'itération courante.
- Si la fonction dépend de plusieurs arguments, on doit trouver une direction de recherche.
- Une longueur de pas doit être fixé pour la direction choisie.
- Un critère de convergence doit être mis en oeuvre pour permettre de stopper la procedure d'itération et permettre de conclure à la convergence de la procedure après un nombre finis d'étape.

4.1 La méthode du gradient

Les algorithmes de minimisation que nous utiliserons ici sont basés sur la **méthode du gradient**. La k-ième itération de chaque algorithmes prend la forme

$$\theta^k = \theta^{k-1} - \lambda Q^{-1} \frac{\partial L_t(\theta)}{\partial \theta} \quad (4.1)$$

ou θ^{k-1} est l'estimation du vecteur de paramètres obtenue après $k - 1$ itération, Q est une approximation de la hessienne calculé à $k - 1$, qui essentiellement détermine la direction donnée, λ est un réel qui détermine la longueur du pas dans la direction donnée.

L'estimation par maximum de vraisemblance devrait être relativement facile depuis le travail de Fiorentini, Calzolari et Panattoni [5, 1996]. Dans ce papier, sont mis en oeuvre des expressions analytiques, complètement utilisables, du gradient et de la hessienne de la fonction de log-vraisemblance du modèle $GARCH(1,1)$. Il est donc possible d'utiliser la méthode de Newton³ pour la maximisation de la log-vraisemblance avec les dérivées secondes analytiques, cette méthode est connu pour converger plus vite dans un voisinage de l'estimation par maximum de vraisemblance. Mais plutôt que d'utiliser cette méthode qui nous obligerait à calculer les dérivées secondes de la log-vraisemblance qui seraient très consommatrices en temps de calcul, on utilisera un algorithme basé sur l'estimation à partir de la matrice d'information (où "score efficace"). Cette méthode a pour avantage de ne pas être très exigeante en terme de calcul et comme l'on montré FCP, de mettre en moyenne un peu plus d'itération pour converger.⁴ La mise en oeuvre de cette algorithme, le sera sur la base d'une régression artificielle.

4.2 La régression artificielle

On considère la régression artificielle qui va nous permettre d'estimer le modèle de régression linéaire 2.1 à erreurs $GARCH(1,1)$. Pour chaque observation correspondent deux nouvelles observations artificielles (empilées). Soit $R(\theta)$ la matrice $(2n \times k + 3)$ des régresseurs et $r(\theta)$ le vecteur $(2n)$ de la régressande, la régression artificielle est

$$r(\theta) = R(\theta)\Phi + \text{residus} \quad (4.2)$$

ou

$$\begin{pmatrix} r_1(\theta) \\ r_2(\theta) \end{pmatrix} = \begin{pmatrix} R_1(\theta) \\ R_2(\theta) \end{pmatrix} \Phi + \text{residus}$$

Une estimation OLS de la régression ?? nous donne

$$\Phi = (R(\theta)^T R(\theta))^{-1} R(\theta)^T Y(\theta)$$

– Pour l'observation t les deux éléments qui correspondent à la régressande $r(\theta)$ sont

$$r_{1t}(\theta) = \frac{u_t}{\sigma_t} \quad (4.3)$$

$$r_{2t}(\theta) = \frac{(u_t^2 - \sigma_t^2)}{\sigma_t^2 \sqrt{2}} \quad (4.4)$$

³On parle de méthode de Newton (pure) par opposition aux méthode quasi-Newton, lorsqu'on utilise une expression analytique de la hessienne pour la direction.

⁴Le nombre d'itération moyenne pour converger avec la matrice d'information est légèrement supérieur que si l'on utilisait la matrice hessienne. Mais FCP on en fait montrer que si l'on ne prenait que le nombre d'itération comme critère de convergence, le meilleur choix serait un algorithme mixte qui commencerait avec la matrice d'information et changerais après 1 ou 2 itération avec la hessienne.

- Les éléments de $R(\theta)$ de la régression artificielle qui correspondent au paramètre β sont les éléments

$$R_{1t}(\beta) = \frac{X_t}{\sigma_t} \quad (4.5)$$

$$R_{2t}(\beta) = -\frac{\alpha_1 \sqrt{2}}{\sigma_t^2} \sum_{s=1}^{t-1} \delta^{s-1} X_{t-s} u_{t-s} \quad (4.6)$$

- Similairement, les éléments des régresseurs qui correspondent au paramètre α_0 sont

$$R_{1t}(\alpha_0) = 0 \quad (4.7)$$

$$R_{2t}(\alpha_0) = \frac{1}{\sigma_t^2 \sqrt{2}} \left(\frac{(1 - \delta^{t-1})}{1 - \delta} + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right) \quad (4.8)$$

- et les regresseurs correspondant au paramètre α_1 sont

$$R_{1t}(\alpha_1) = 0 \quad (4.9)$$

$$R_{2t}(\alpha_1) = \frac{1}{\sigma_t^2 \sqrt{2}} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_0} \right) \quad (4.10)$$

- Finalement, les éléments des régresseurs qui correspondent à δ sont

$$R_{1t}(\delta) = 0$$

$$R_{2t}(\delta) = \frac{1}{\sigma_t^2 \sqrt{2}} \left(-\frac{\alpha_0 (t-1) \delta^{t-2}}{1 - \delta} + \frac{\alpha_0 (1 - \delta^{t-1})}{(1 - \delta)^2} + \right) \quad (4.11)$$

$$\alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \delta} \quad (4.12)$$

On montre que le gradient de la log-vraisemblance est nul quand il est évalué en des valeurs des paramètres pour lesquels la régressande est orthogonale aux régresseurs. Plus précisément, pour chaque observation et chaque paramètre, le produit des premiers éléments de la régressande avec les premiers éléments correspondant des régresseurs, plus le produit des seconds éléments de la régressande avec les seconds éléments des régresseurs est égal à la dérivée partielle de l_t par rapport au paramètre soit

$$R_{1t}(\beta) r_{1t}(\beta) + R_{2t}(\beta) r_{2t}(\beta) = \frac{\partial l_t}{\partial \beta}$$

$$R_{1t}(\alpha_0)^T r_{1t}(\alpha_0) + R_{2t}(\alpha_0)^T r_{2t}(\alpha_0) = \frac{\partial l_t}{\partial \alpha_0}$$

$$R_{1t}(\alpha_1)^T r_{1t}(\alpha_1) + R_{2t}(\alpha_1)^T r_{2t}(\alpha_1) = \frac{\partial l_t}{\partial \alpha_1}$$

$$R_{1t}(\delta)^T r_{1t}(\delta) + R_{2t}(\delta)^T r_{2t}(\delta) = \frac{\partial l_t}{\partial \delta}$$

comme cela est donnée dans l'expression 3.4.

En premier pour les éléments du vecteur β on forme l'expression

$$\begin{aligned} & \frac{X_{ti} u_t}{\sigma_t \sigma_t} - \frac{u_t^2 - \sigma_t^2}{\sigma_t^2 \sqrt{2}} \frac{\alpha_1 \sqrt{2}}{\sigma_t^2} \sum_{s=1}^{t-1} \delta^{s-1} X_{ti} u_{t-s}^2 \\ &= \frac{X_{ti} u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \sum_{s=1}^{t-1} \delta^{s-1} X_{ti} u_{t-s}^2 \\ &= \frac{\partial l_t}{\partial b} \end{aligned}$$

On retrouve l'expression de 3.1.

De même, pour α_0 , on a

$$\begin{aligned} & \frac{u_t^2 - \sigma_t^2}{\sigma_t^2 \sqrt{2}} \frac{1}{\sigma_t^2 \sqrt{2}} \left(\frac{1 - \delta^{t-1}}{1 - \delta} + \delta^{t-1} \frac{\partial \sigma_1^2(\alpha_0, \alpha_1, \delta)}{\partial \alpha_0} \right) \\ &= \frac{\partial l_t}{\partial \alpha_0} \end{aligned}$$

puis pour α_1

$$\begin{aligned} & \frac{u_t^2 - \sigma_t^2}{\sigma_t^2 \sqrt{2}} \frac{1}{\sigma_t^2 \sqrt{2}} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_1^2(\alpha_0, \alpha_1, \delta)}{\partial \alpha_1} \right) \\ &= \frac{\partial l_t}{\partial \alpha_1} \end{aligned}$$

Et enfin pour δ

$$\begin{aligned} & \frac{u_t^2 - \sigma_t^2}{\sigma_t^2 \sqrt{2}} \frac{1}{\sigma_t^2 \sqrt{2}} \left(-\frac{\alpha_0 (t-1) \delta^{t-2}}{1 - \delta} + \frac{\alpha_0 (1 - \delta^{t-1})}{(1 - \delta)^2} \right. \\ & \quad \left. + \alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_1^2(\alpha_0, \alpha_1, \delta)}{\partial \delta} \right) \\ &= \frac{\partial l_t}{\partial \delta} \end{aligned}$$

On va maintenant montrer que $R(\theta)^T R(\theta)$ est la matrice d'information de la vraisemblance I

$$R(\theta)^T R(\theta) = I$$

Où en terme de contribution, si ϕ_1 et ϕ_2 sont des paramètre du modèle, c'est à dire des élément de β ou de la variance conditionnelle

$$E \left(\frac{\partial l_t}{\partial \phi_1} \frac{\partial l_t}{\partial \phi_2} \middle| \Omega_t \right) = R_{1t}(\phi_1)^T R_{1t}(\phi_2) + R_{2t}(\phi_1)^T R_{2t}(\phi_2)$$

Pour cela on remarque que σ_t^2 fait parti de l'ensemble d'information Ω_t . on voit que une contribution à la matrice d'information peut s'exprimer comme

$$\begin{aligned} E \left(\frac{\partial l_t}{\partial \phi_1} \frac{\partial l_t}{\partial \phi_2} \middle| \Omega_t \right) &= E \left(\left(\frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \phi_1} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \phi_1} \right) \left(\frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \phi_2} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \phi_2} \right) \middle| \Omega_t \right) \\ &= \frac{1}{\sigma_t^4} E(u_t^2) \frac{\partial u_t}{\partial \phi_1} \frac{\partial u_t}{\partial \phi_2} + \frac{1}{4\sigma_t^8} E((u_t^2 - \sigma_t^2)^2) \frac{\partial \sigma_t^2}{\partial \phi_1} \frac{\partial \sigma_t^2}{\partial \phi_2} \\ &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \phi_1} \frac{\partial u_t}{\partial \phi_2} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \phi_1} \frac{\partial \sigma_t^2}{\partial \phi_2} \end{aligned}$$

Commençons par regarder pour $\phi_1 = \beta_i, \phi_2 = \beta_j$,

$$\begin{aligned} E \left(\frac{\partial l_t}{\partial \beta_i} \frac{\partial l_t}{\partial \beta_j} \middle| \Omega_t \right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \beta_i} \frac{\partial u_t}{\partial \beta_j} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \beta_i} \frac{\partial \sigma_t^2}{\partial \beta_j} \\ &= \frac{1}{\sigma_t^2} X_{ti} X_{tj} + \frac{2\alpha_1}{\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)i} \right) \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)j} \right) \\ &= R_{1t}(\beta_i)^T R_{1t}(\beta_j) + R_{2t}(\beta_i)^T R_{2t}(\beta_j) \end{aligned}$$

Ce qui est bien la contribution de l'observation t de l'élément qui correspond à β_i et β_j du produit matricielle $R(\beta, \theta)^T R(\beta, \theta)$. Pour les paramètres $\phi_1 = \beta_i, \phi_2 = \alpha_0$

$$\begin{aligned} E \left(\frac{\partial l_t}{\partial \beta_i} \frac{\partial l_t}{\partial \alpha_0} \middle| \Omega_t \right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \beta_i} \frac{\partial u_t}{\partial \alpha_0} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \beta_i} \frac{\partial \sigma_t^2}{\partial \alpha_0} \\ &= 0 + \frac{1}{2\sigma_t^4} \left(-2\alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)i} \right) \left(\frac{1 - \delta^{t-1}}{1 - \delta} + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_0} \right) \\ &= \frac{-\alpha_1}{\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)i} \right) \left(\frac{1 - \delta^{t-1}}{1 - \delta} + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_0} \right) \\ &= R_{1t}(\beta_i)^T R_{1t}(\alpha_0) + R_{2t}(\beta_i)^T R_{2t}(\alpha_0) \end{aligned}$$

Pour les paramètres $\phi_1 = \beta_i, \phi_2 = \alpha_1$

$$\begin{aligned} E \left(\frac{\partial l_t}{\partial \beta_i} \frac{\partial l_t}{\partial \alpha_1} \middle| \Omega_t \right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \beta_i} \frac{\partial u_t}{\partial \alpha_1} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \beta_i} \frac{\partial \sigma_t^2}{\partial \alpha_1} \\ &= 0 + \frac{1}{2\sigma_t^4} \left(-2\alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)i} \right) \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right) \\ &= \frac{-\alpha_1}{\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)i} \right) \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right) \\ &= R_{1t}(\beta_i)^T R_{1t}(\alpha_1) + R_{2t}(\beta_i)^T R_{2t}(\alpha_1) \end{aligned}$$

Pour les paramètres $\phi_1 = \beta_i$, $\phi_2 = \delta$

$$\begin{aligned}
E\left(\frac{\partial l_t}{\partial \beta_i} \frac{\partial l_t}{\partial \delta} \middle| \Omega_t\right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \beta_i} \frac{\partial u_t}{\partial \delta} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \beta_i} \frac{\partial \sigma_t^2}{\partial \delta} \\
&= 0 - \frac{\alpha_1}{\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 X_{(t-s)i} \right) \left(-\frac{\alpha_0(t-1)\delta^{t-2}}{1-\delta} + \frac{\alpha_0(1-\delta^{t-1})}{(1-\delta)^2} \right. \\
&\quad \left. + \alpha_1 \sum_{s=1}^{t-1} (s-1)\delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \delta} \right) \\
&= R_{1t}(\beta_i)^T R_{1t}(\delta) + R_{2t}(\beta_i)^T R_{2t}(\delta)
\end{aligned}$$

Pour les paramètres $\phi_1 = \alpha_0$, $\phi_2 = \alpha_0$

$$\begin{aligned}
E\left(\frac{\partial l_t}{\partial \alpha_0} \frac{\partial l_t}{\partial \alpha_0} \middle| \Omega_t\right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \alpha_0} \frac{\partial u_t}{\partial \alpha_0} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \alpha_0} \frac{\partial \sigma_t^2}{\partial \alpha_0} \\
&= \frac{1}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right)^2 \\
&= R_{1t}(\alpha_0)^T R_{1t}(\alpha_0) + R_{2t}(\alpha_0)^T R_{2t}(\alpha_0)
\end{aligned}$$

Pour les paramètres $\phi_1 = \alpha_0$, $\phi_2 = \alpha_1$

$$\begin{aligned}
E\left(\frac{\partial l_t}{\partial \alpha_0} \frac{\partial l_t}{\partial \alpha_1} \middle| \Omega_t\right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \alpha_0} \frac{\partial u_t}{\partial \alpha_1} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \alpha_0} \frac{\partial \sigma_t^2}{\partial \alpha_1} \\
&= 0 + \frac{1}{2\sigma_t^4} \left(\frac{1-\delta^{t-1}}{1-\delta} + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_0} \right) \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right) \\
&= R_{1t}(\alpha_0)^T R_{1t}(\alpha_1) + R_{2t}(\alpha_0)^T R_{2t}(\alpha_1)
\end{aligned}$$

Pour les paramètres $\phi_1 = \alpha_0$, $\phi_2 = \delta$

$$\begin{aligned}
E\left(\frac{\partial l_t}{\partial \alpha_0} \frac{\partial l_t}{\partial \delta} \middle| \Omega_t\right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \alpha_0} \frac{\partial u_t}{\partial \delta} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \alpha_0} \frac{\partial \sigma_t^2}{\partial \delta} \\
&= \frac{1}{2\sigma_t^4} \left(\frac{1-\delta^{t-1}}{1-\delta} + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_0} \right) \left(-\frac{\alpha_0(t-1)\delta^{t-2}}{1-\delta} + \frac{\alpha_0(1-\delta^{t-1})}{(1-\delta)^2} \right. \\
&\quad \left. + \alpha_1 \sum_{s=1}^{t-1} (s-1)\delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \delta} \right) \\
&= R_{1t}(\alpha_0)^T R_{1t}(\delta) + R_{2t}(\alpha_0)^T R_{2t}(\delta)
\end{aligned}$$

Pour les paramètres $\phi_1 = \alpha_1$, $\phi_2 = \alpha_1$

$$\begin{aligned} E \left(\frac{\partial l_t}{\partial \alpha_1} \frac{\partial l_t}{\partial \alpha_1} \middle| \Omega_t \right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \alpha_1} \frac{\partial u_t}{\partial \alpha_1} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \alpha_1} \frac{\partial \sigma_t^2}{\partial \alpha_1} \\ &= 0 + \frac{1}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right)^2 \\ &= R_{1t}(\alpha_1)^T R_{1t}(\alpha_1) + R_{2t}(\alpha_1)^T R_{2t}(\alpha_1) \end{aligned}$$

Pour les paramètres $\phi_1 = \alpha_1$, $\phi_2 = \delta$

$$\begin{aligned} E \left(\frac{\partial l_t}{\partial \alpha_1} \frac{\partial l_t}{\partial \delta} \middle| \Omega_t \right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \alpha_1} \frac{\partial u_t}{\partial \delta} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \alpha_1} \frac{\partial \sigma_t^2}{\partial \delta} \\ &= \frac{1}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \alpha_1} \right) \left(-\frac{\alpha_0(t-1)\delta^{t-2}}{1-\delta} + \frac{\alpha_0(1-\delta^{t-1})}{(1-\delta)^2} \right) \\ &\quad + \alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \delta} \\ &= R_{1t}(\alpha_1)^T R_{1t}(\delta) + R_{2t}(\alpha_1)^T R_{2t}(\delta) \end{aligned}$$

Pour les paramètres $\phi_1 = \delta$, $\phi_2 = \delta$

$$\begin{aligned} E \left(\frac{\partial l_t}{\partial \delta} \frac{\partial l_t}{\partial \delta} \middle| \Omega_t \right) &= \frac{1}{\sigma_t^2} \frac{\partial u_t}{\partial \delta} \frac{\partial u_t}{\partial \delta} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \delta} \frac{\partial \sigma_t^2}{\partial \delta} \\ &= \frac{1}{2\sigma_t^4} \left(-\frac{\alpha_0(t-1)\delta^{t-2}}{1-\delta} + \frac{\alpha_0(1-\delta^{t-1})}{(1-\delta)^2} \right) \\ &\quad + \alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{t-2} u_{t-s}^2 + \delta^{t-1} \frac{\partial \sigma_u^2}{\partial \delta} \Big)^2 \\ &= R_{1t}(\delta)^T R_{1t}(\delta) + R_{2t}(\delta)^T R_{2t}(\delta) \end{aligned}$$

On vient donc de montrer que la régression artificielle que nous avons construite correspond à une maximisation par quasi-newton avec comme approximation numérique de la **Hessienne** la matrice d'information du maximum de vraisemblance, on parle d'estimateur du score efficace, le processus itératif est le suivant

$$\Phi_{i+1} = \Phi_i + k \left(R(\theta)^T R(\theta) \right)^{-1} R(\theta)^T r(\theta)$$

équivalent à

$$\theta_{i+1} = \theta_i + k I^{-1} \frac{\partial L}{\partial \theta}$$

en terme d'algorithme du gradient.

5 Les différents choix initiaux

Si la distribution stationnaire de la série σ_t^2 était connue comme une fonction des paramètres *GARCH* alors sous l'hypothèse que le processus *GARCH*(1,1) est bien stationnaire la fonction de log-vraisemblance non conditionnelle serait donnée par la somme des contributions $l_t(\theta)$ avec $\sigma_t^2(\theta)$ comme dans 3.1 et $u_t = u_t(\beta)$, multiplié par la densité stationnaire de σ_1^2 et intégré selon σ_1^2 de 0 à ∞ . Tant que la distribution stationnaire d'un processus *GARCH*(1,1) n'est pas connue de manière analytique, l'estimation par maximum de vraisemblance non conditionnelle n'est pas faisable. Cependant diverses procédures ont été mises en oeuvre pour contourner ce problème.

5.1 Le choix initial compatible avec la stationnarité

Avant de regarder les choix particuliers, nous allons montrer que $\sigma_1^2 = \alpha_0/(1 - \alpha_1 - \delta)$ est la seule valeur de σ_1^2 compatible avec la stationnarité. Si nous partons de l'observation 0 la récurrence est

$$\sigma_t^2 = \frac{\alpha_0(1 - \delta^{t-1})}{1 - \delta} + \alpha_1 \sum_{s=1}^t \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \sigma_0^2 \quad (5.1)$$

On s'aperçoit que dans cette expression u_0^2 et σ_0^2 sont tous deux inconnus. Notons σ_u^2 leur esperance non conditionnelle commune. On remplace les deux quantités inconnues par leur esperance non conditionnelle

$$\sigma_t^2 = \frac{\alpha_0(1 - \delta^{t-1})}{1 - \delta} + \alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \sigma_u^2 + \alpha_1 \delta^t \sigma_u^2 \quad (5.2)$$

D'un autre coté on peut réécrire la recurrence de base 3.1 avec $\sigma_1^2 = \sigma_u^2$

$$\sigma_t^2 = \frac{\alpha_0(1 - \delta^{t-1})}{1 - \delta} + \alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \sigma_u^2 \quad (5.3)$$

En soustrayant les équations 5.2 et 5.3 on obtient

$$\frac{\alpha_0(1 - \delta^{t-1})}{1 - \delta} - \delta^{t-1}(1 - \alpha_1 - \delta)\sigma_u^2 = \delta^{t-1}(\alpha_0 - (1 - \alpha_1 - \delta)\sigma_u^2)$$

expression qui n'est valide si et seulement si $\sigma_u^2 = \alpha_0/(1 - \alpha_1 - \delta)$.

5.2 Les différentes méthodes

La méthode simpliste $\sigma_1^2 = 0$

C'est la méthode la plus simple à mettre en oeuvre mais sûrement la plus mauvaise du point de vue de l'estimation. Les dérivées partielles de σ_t^2 par rapport au paramètre sont :

$$\begin{aligned}\frac{\partial \sigma_t^2}{\partial b} &= -2\alpha_1 \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \\ \frac{\partial \sigma_t^2}{\partial \alpha_0} &= \frac{1 - \delta^{t-1}}{1 - \delta} \\ \frac{\partial \sigma_t^2}{\partial \alpha_1} &= \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \\ \frac{\partial \sigma_t^2}{\partial \delta} &= -\frac{\alpha_0(t-1)\delta^{t-2}}{1 - \delta} + \frac{\alpha_0(1 - \delta^{t-1})}{(1 - \delta)^2} \\ &\quad + \alpha_1 \sum_{s=1}^{t-1} (s-1)\delta^{t-2} u_{t-s}^2\end{aligned}$$

Avec cela on peut écrire la contribution au **gradient** de la fonction de log-vraisemblance

$$\begin{aligned}\frac{\partial l_t}{\partial b} &= \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \delta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{X_t u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \sum_{s=1}^{t-1} \delta^{s-1} X_t u_{t-s}^2 \\ \frac{\partial l_t}{\partial \alpha_0} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_0} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \frac{(1 - \delta^{t-1})}{1 - \delta} \\ \frac{\partial l_t}{\partial \alpha_1} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_1} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \right) \\ \frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(-\frac{\alpha_0(t-1)\delta^{t-2}}{1 - \delta} + \frac{\alpha_0(1 - \delta^{t-1})}{(1 - \delta)^2} + \right. \\ &\quad \left. \alpha_1 \sum_{s=1}^{t-1} (s-1)\delta^{t-2} u_{t-s}^2 \right)\end{aligned}$$

Les méthodes alternatives $\sigma_1^2 = \frac{1}{T} \sum_{s=1}^T \mathbf{u}_s^2$

C'est la méthode préconisée par Bollerslev dans son article introductif aux modèles *GARCH*. Il suggère que σ_u^2 soit égal à la somme des carrés des résidus, $\|y - X\hat{b}\|^2$, divisé par la taille de l'échantillon. Ceci peut être obtenu par une simple régression de y sur X par les *OLS*. Alternativement on pourrait considérer que σ_u^2 est une fonction de β et de le traiter ainsi dans la fonction de vraisemblance et dans la définition de σ_t^2 . C'est en quelques sorte une mise à jours en continue. Dans tous ces cas, on a le choix de commencer soit à l'observation 0 soit à l'observation 1

- La première sous méthode à prendre u_t^2 comme les résidus d'une estimation *OLS*.

$$\sigma_1^2 = \frac{1}{T} \sum_{s=1}^T \widehat{u}_s^2 = \widehat{\sigma}_1^2 \quad (5.4)$$

En commençant à l'observation 1, on peut écrire la contribution au **gradient** de la fonction de log-vraisemblance

$$\begin{aligned} \frac{\partial l_t}{\partial b} &= \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \delta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{X_t u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \sum_{s=1}^{t-1} \delta^{s-1} X_t u_{t-s}^2 \\ \frac{\partial l_t}{\partial \alpha_0} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_0} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \frac{(1 - \delta^{t-1})}{1 - \delta} \\ \frac{\partial l_t}{\partial \alpha_1} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \right) \\ \frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(-\frac{\alpha_0 (t-1) \delta^{t-2}}{1 - \delta} + \frac{\alpha_0 (1 - \delta^{t-1})}{(1 - \delta)^2} + \right. \\ &\quad \left. \alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{t-2} u_{t-s}^2 + (t-1) \delta^{t-2} \widehat{\sigma}_1^2 \right) \end{aligned}$$

Si l'on avait choisi l'observation 0 comme point de départ on obtiendrait comme contribution au gradient

$$\begin{aligned} \frac{\partial l_t}{\partial b} &= \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \delta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{X_t u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \sum_{s=1}^{t-1} \delta^{s-1} X_t u_{t-s}^2 \\ \frac{\partial l_t}{\partial \alpha_0} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_0} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \frac{(1 - \delta^t)}{1 - \delta} \\ \frac{\partial l_t}{\partial \alpha_1} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \widehat{\sigma}_1^2 \right) \\ \frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(-\frac{\alpha_0 t \delta^{t-1}}{1 - \delta} + \frac{\alpha_0 (1 - \delta^t)}{(1 - \delta)^2} + \right. \\ &\quad \left. \alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{s-2} u_{t-s}^2 + \delta^{t-2} (\alpha_1 (t-1) + t\delta) \widehat{\sigma}_1^2 \right) \end{aligned}$$

- La deuxième sous méthode consiste à prendre u_s^2 comme un paramètre du modèle

$$\sigma_1^2 = \frac{1}{T} \sum_{s=1}^T u_s^2 = \frac{1}{T} \sum_{s=1}^T (y_s - X_s \delta)$$

Avec cela on peut écrire la contribution au **gradient** de la fonction de log-vraisemblance en commençant a l'observation 1

$$\begin{aligned}
\frac{\partial l_t}{\partial b} &= \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \delta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} \\
&= \frac{X_t u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} X_t u_{t-s}^2 + \frac{\delta^{t-1}}{\alpha_1 T} \sum_{s=1}^T X_s u_s \right) \\
\frac{\partial l_t}{\partial \alpha_0} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_0} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \frac{(1 - \delta^{t-1})}{1 - \delta} \\
\frac{\partial l_t}{\partial \alpha} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \right) \\
\frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(-\frac{\alpha_0 (t-1) \delta^{t-2}}{1 - \delta} + \frac{\alpha_0 (1 - \delta^{t-1})}{(1 - \delta)^2} + \right. \\
&\quad \left. \alpha \sum_{s=1}^{t-1} (s-1) \delta^{t-2} u_{t-s}^2 + (t-1) \frac{\delta^{t-2}}{T} \sum_{s=1}^T u_s^2 \right)
\end{aligned}$$

Si l'on avait choisit de commencer à l'observation 0 on aurait

$$\begin{aligned}
\frac{\partial l_t}{\partial b} &= \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \delta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{X_t u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \\
&\quad \left(\sum_{s=1}^{t-1} \delta^{s-1} X_t u_{t-s}^2 + \frac{(\alpha_1 + \delta) \delta^{t-1}}{\alpha_1 T} \sum_{s=1}^T X_s u_s \right) \\
\frac{\partial l_t}{\partial \alpha_0} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_0} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \frac{(1 - \delta^t)}{1 - \delta} \\
\frac{\partial l_t}{\partial \alpha_1} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \delta^{t-1} \widehat{\sigma}_1^2 \right) \\
\frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(-\frac{\alpha_0 t \delta^{t-1}}{1 - \delta} + \frac{\alpha_0 (1 - \delta^t)}{(1 - \delta)^2} + \right. \\
&\quad \left. \alpha_1 \sum_{s=1}^{t-1} (s-1) \delta^{s-2} u_{t-s}^2 + \delta^{t-2} (\alpha_1 (t-1) + t\delta) \widehat{\sigma}_1^2 \right)
\end{aligned}$$

La méthode stationnaire $\sigma_1^2 = \alpha_0 / (1 - \alpha - \delta)$

L'une des méthodes consiste à remplacer σ_u^2 par son espérance non conditionnelle $\alpha_0 / (1 - \alpha - \delta)$. C'est le seul choix qui ne dépend pas de l'observation de départ 1 ou 0. On remplace σ_u^2

par son expression dans σ_t^2

$$\sigma_t^2 = \frac{\alpha_0(1 - \delta^{t-1})}{1 - \delta} + \alpha \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \frac{\alpha_0 \delta^{t-1}}{1 - \alpha - \delta}$$

On peut écrire la contribution au **gradient** de la fonction de log-vraisemblance

$$\begin{aligned} \frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \delta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{X_t u_t}{\sigma_t^2} - \frac{\alpha_1 (u_t^2 - \sigma_t^2)}{\sigma_t^4} \sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 \\ \frac{\partial l_t}{\partial \alpha_0} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_0} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\frac{(1 - \delta^{t-1})}{1 - \delta} + \frac{\delta^{t-1}}{1 - \alpha - \delta} \right) \\ \frac{\partial l_t}{\partial \alpha_1} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha_1} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\sum_{s=1}^{t-1} \delta^{s-1} u_{t-s}^2 + \frac{\alpha_0 \delta^{t-1}}{(1 - \alpha - \delta)^2} \right) \\ \frac{\partial l_t}{\partial \delta} &= \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(-\frac{\alpha_0(t-1)\delta^{t-2}}{1 - \delta} + \frac{\alpha_0(1 - \delta^{t-1})}{(1 - \delta)^2} + \right. \\ &\quad \left. + \alpha_1 \sum_{s=1}^{t-1} (s-1)\delta^{t-2} u_{t-s}^2 + \frac{\alpha_0(t-1)\delta^{t-2}}{1 - \alpha - \delta} + \frac{\alpha_0 \delta^{t-1}}{(1 - \alpha - \delta)^2} \right) \end{aligned}$$

On a donc retenu 6 méthodes pour remplacer la variance initiale.

6 Les simulations

Pour vérifier les avantages ou les inconvénients de chaque méthode d'initialisation du processus *GARCH* on va simuler un processus *GARCH(1,1)* gaussien puis l'estimer par maximum de vraisemblance conditionnelle avec une maximisation par quasi-newton. L'ensemble des simulation est mis en oeuvre en *C++* et est décrite dans l'[annexe A](#). On simule trois groupes de séries dont les modèles sont

$$\begin{aligned} y_t &= 0.5 + u_t && \text{(modele 1)} \\ \sigma_t^2 &= 0.5 + 0.1u_{t-1}^2 + 0.7\sigma_{t-1}^2 \end{aligned}$$

et

$$\begin{aligned} y_t &= 0.5 + u_t && \text{(modele 2)} \\ \sigma_t^2 &= 0.05 + 0.05u_{t-1}^2 + 0.9\sigma_{t-1}^2 \end{aligned}$$

et enfin

$$\begin{aligned} y_t &= 0.5 + u_t \\ \sigma_t^2 &= 0.25 + 0.3u_{t-1}^2 + 0.4\sigma_{t-1}^2 \end{aligned} \quad (\text{modele 3})$$

le modèle 1 sera simulé 10 000 fois, le modèle 2, 2 000 fois. Les valeurs des paramètres pour l'initialisation du processus itératif sont respectivement (0.5, 0.5, 0.1, 0.7) et (0.5, 0.05, 0.07, 0.87). Les tailles d'échantillons sont 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 4000 et enfin 5000 pour le modèle 1 et 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000 et 5000. Soit θ_0^1 et θ_0^2 les vecteurs des paramètres des modèles simulés, on définit l'espérance de l'erreur quadratique en fonction de chaque initialisation et de chaque taille d'échantillon T comme

$$E(I_l) = \frac{1}{T} \sum_{i=1}^T (\hat{\theta}_i - \theta_0^j)^2 \quad j = \{1, 2, 3\} \quad l = \{1, 2, 3, 4, 5, 6\}$$

où $\hat{\theta}_i$ est le vecteur des paramètres estimés pour la i -ème simulation. Cette quantité nous permettra d'apprécier la précision général de l'estimation. De même on définit l'erreur quadratique pour chaque paramètre du vecteur estimer

$$\begin{aligned} eq_1(\text{init}) &= \frac{1}{T} \sum_{i=1}^T (\hat{\alpha}_{0i} - \alpha_{00}^j)^2 \quad j = \{1, 2, 3\} \\ eq_2(\text{init}) &= \frac{1}{T} \sum_{i=1}^T (\hat{\alpha}_{1i} - \alpha_{01}^j)^2 \quad j = \{1, 2, 3\} \\ eq_3(\text{init}) &= \frac{1}{T} \sum_{i=1}^T (\hat{\delta}_i - \delta_0^j)^2 \quad j = \{1, 2, 3\} \end{aligned}$$

On regardera aussi l'espérance du biais de chaque paramètre

$$\begin{aligned} E(b_1) &= \frac{1}{T} \sum_{i=1}^T (\hat{\alpha}_{0i} - \alpha_{00}^j) \quad j = \{1, 2, 3\} \\ E(b_2) &= \frac{1}{T} \sum_{i=1}^T (\hat{\alpha}_{1i} - \alpha_{01}^j) \quad j = \{1, 2, 3\} \\ E(b_3) &= \frac{1}{T} \sum_{i=1}^T (\hat{\delta}_i - \delta_0^j) \quad j = \{1, 2, 3\} \end{aligned}$$

Le nombre maximale d'itération est fixé à 30 pour les modèle 1 et 2 et 50 pour le troisième modèle.

6.1 Les résultats des simulations

Les résultats des simulations seront présentés sous forme de graphique, l'ensemble des valeurs numériques sont disponibles dans l' [annexe B](#). L'initialisation avec $\sigma_u^2 = 0$ ne sera pas présenté. En effet cette méthode qui présente une forte instabilité numérique et des résultats vraiment mauvais avait tendance à nuire à la lecture des autres résultats. Le [graphique 1](#) entre la taille de l'échantillon et l'espérance de l'erreur quadratique de chaque séries nous montre, sans équivoque, l'influence de la valeur initiale sur la précision de l'estimation : il n'y a aucune différence entre les méthodes d'initialisations, à tel point que les courbes des différentes méthodes sont carrément confondues et cela même pour des tailles d'échantillon faibles. Dès 200 éléments les courbes se rejoignent pour ne plus se quitter. De plus ce graphique nous renseigne sur la précision de l'estimation en fonction de la taille de l'échantillon.

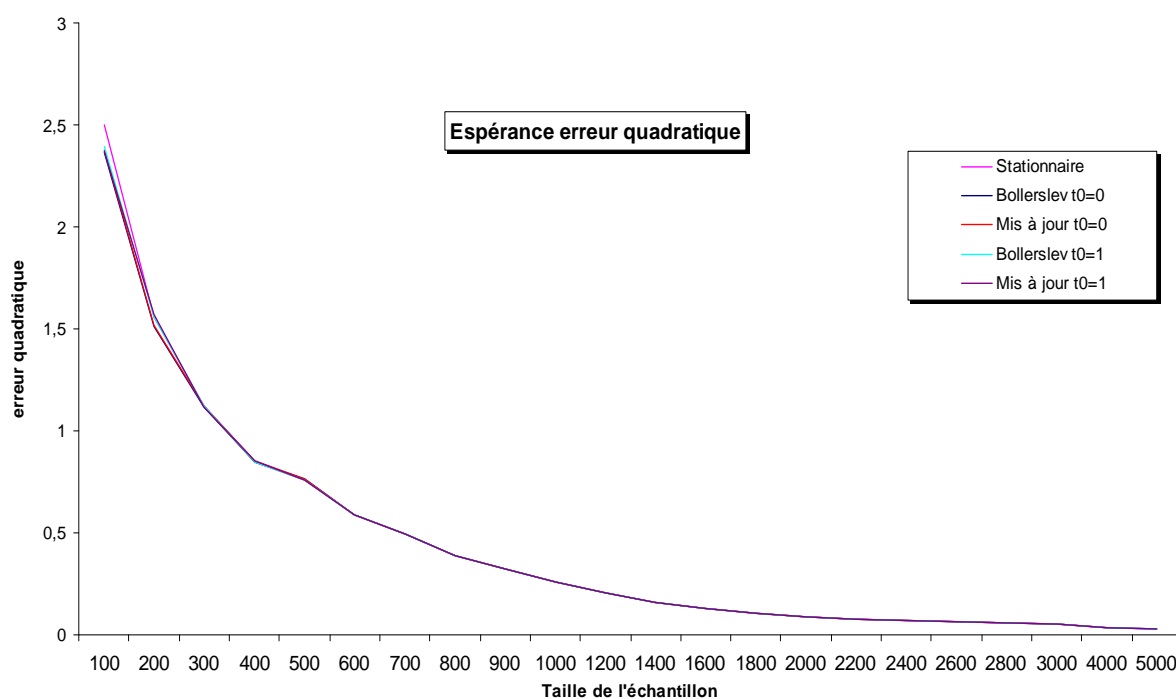
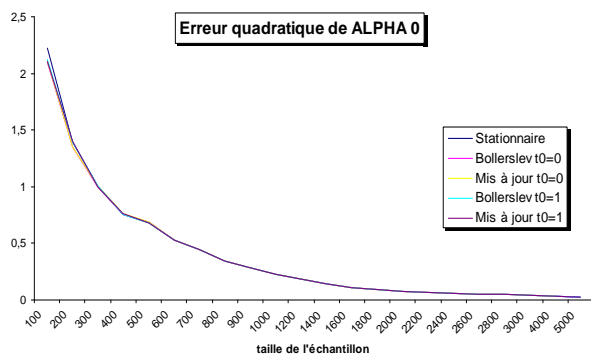
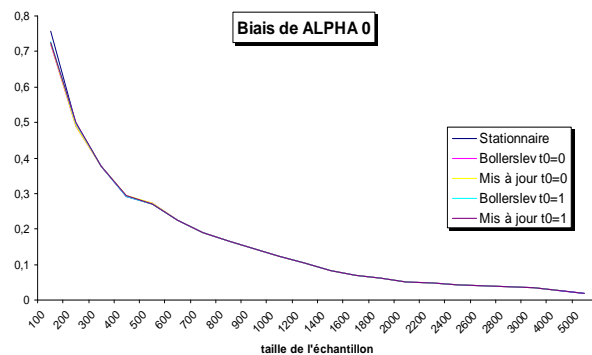


Figure 1: .graphique 1 : modèle 1

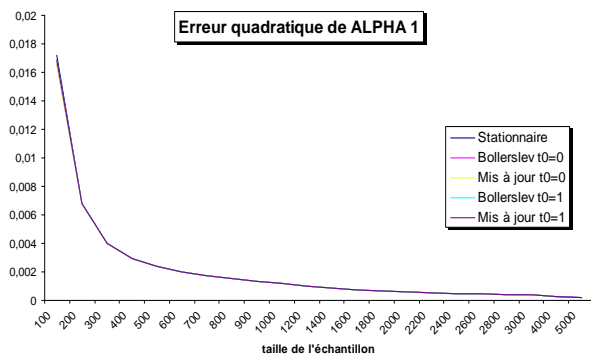
Les graphiques suivant, 2 et 3, 4 et 5, 6 et 7 nous confirme les résultats globaux. Les courbes du biais et de l'espérance quadratique de chaque paramètres pour les différentes méthode d'initialisation sont la aussi confondues.



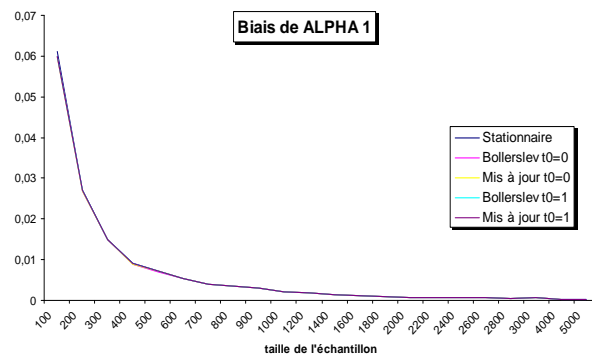
Graphique 2 : modèle 1



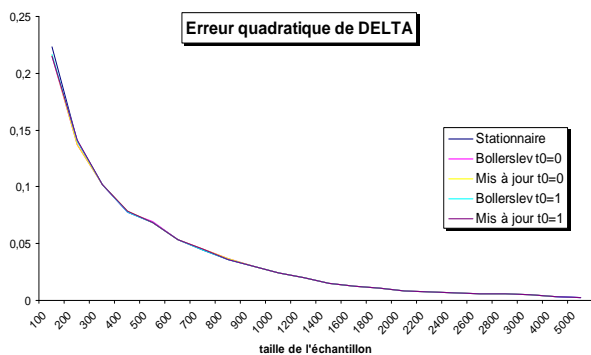
Graphique 3 : modèle 1



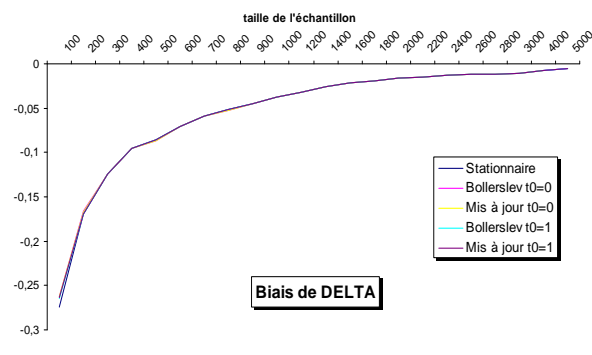
Graphique 4 : modèle 1



Graphique 5 : modèle 1



Graphique 6 : modèle 1



Graphique 7 : modèle 1

On notera que le biais pour δ est négatif.

En ce qui concerne le modèle 2 dont les paramètres sont proches de la non stationnarité, les

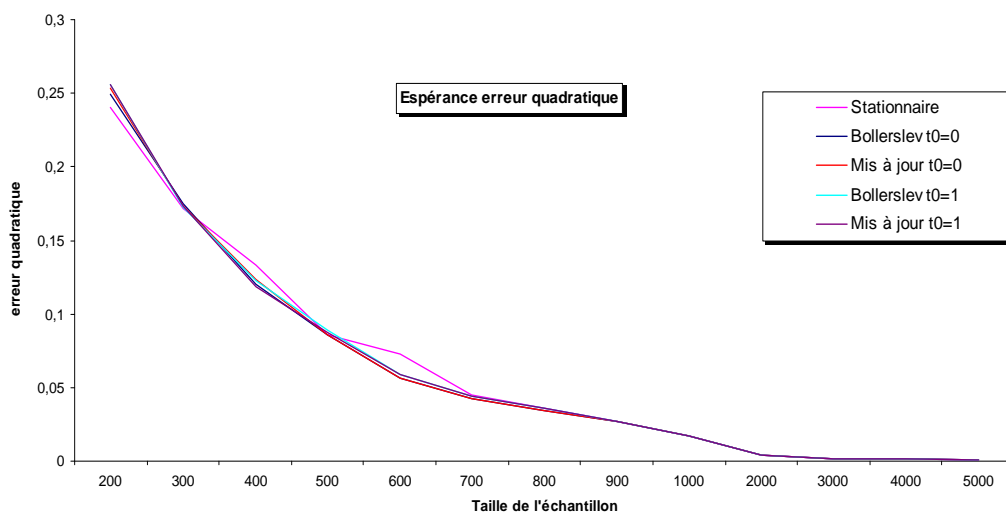
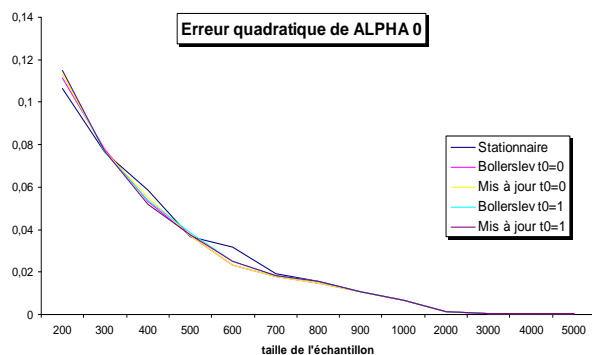
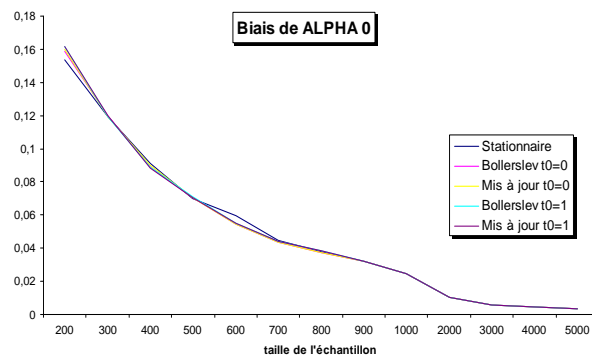


Figure 2: Graphique 8 : modèle 2

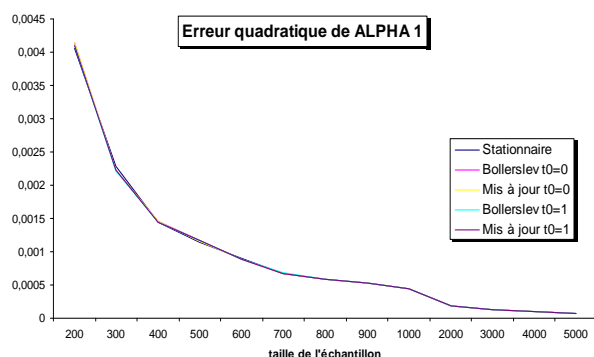
résultats sont sensiblement les mêmes.



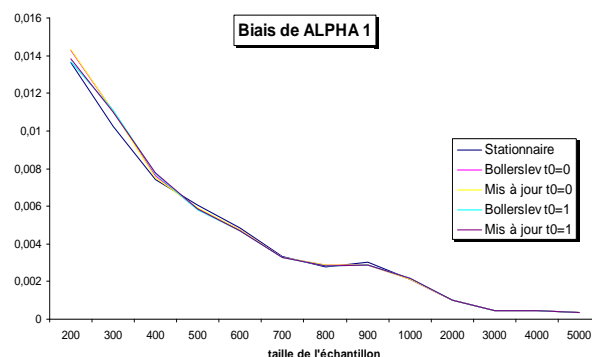
Graphique 9 : modèle 2



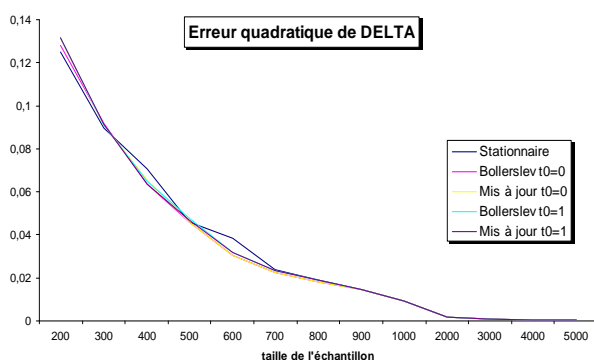
Graphique 10 : modèle 2



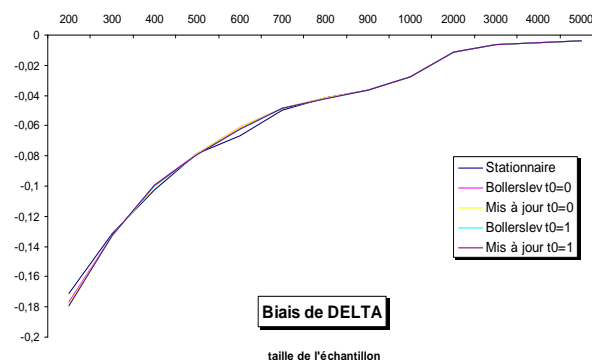
Graphique 11 : modèle 2



Graphique 12 : modèle 2



Graphique 13 : modèle 2



Graphique 14 : modèle 2

Cette première série de test nous montre que pour une estimation par maximum de vraisemblance l'influence du choix de la valeur initiale est vraiment faible.

7 La matrice de covariance et l'inférence

8 Le benchmark et les autres progiciels

L'une des idées majeurs de ce travail était de vérifier si la spécification de la variance initiale était, comme le soupçonnait Brook, Burke et Persaud[2], à l'origine des différences entre les résultats de leur enquête. Leur papier, datant de 2001, passait en revue 9 progiciels d'économétrie avec pour but principale de comparer leurs estimations des modèles *GARCH*. Ils ont adopté, comme méthode de test, la comparaison de l'estimation d'une série de 1974 éléments face à une estimation " " proposé FCP[4] qui utilise un algorithme de Newton pure, avec comme

direction la hessienne analytique. la série en question est le taux de change entre la livre et le mark. Ils constatèrent que, malgré le bon comportement connue de la série estimer, les résultats donnés par défaut étaient considérablement différent d'un logiciel à l'autre. Une lecture plus approfondit de leur papier relativise ce résultat, si l'on modifie certaine option de l'estimation dans les logiciels qui ont les plus mauvais résultats avec les options par défaut on se rapproche considérablement du *benchmark* avec une précision des résultats largement acceptable. Après tout il n'est pas dit que le *benchmark* donne les bonnes valeurs à 6 décimale près! Mais la ou les différences restent encore considérables sont dans l'estimation de la matrice de variance. Une conclusion de leur travail préconise que les progiciels d'économétrie doivent, par défaut, utiliser la meilleur méthode d'estimation connu et que l'utilisateur n'ait pas à naviguer dans l'aide pour avoir de bons résultats. En particulier ils preconise l'emploi de la matrice Hessienne analytique pour l'estimation du modèle $GARCH(1,1)$.

Dans un premier temps il était tentant de regarder le comportement de notre algorithme et des différentes initialisation face au *benchmark*.

Depuis la date de la publication de Brook, Burke et Persand (publiée en 2001) certains des logiciels ont été mis à jour vers de nouvelles versions. On va maintenant regarder si les nouvelles versions donnent des résultats différents, de même on va tester, toujours contre le *benchmark* d'autre progiciels d'économétrie ou modules d'estimation des processus $GARCH$. On commencera par une présentation des logiciels et contrairement à Brook, Burke et Persand, une note sur l'ergonomie et les différentes options sera faite. Il est évident que de telles critiques (positives ou négatives) est totalement subjective. Une autres remarques concerne deux expressions communément employées par les chercheurs lorsqu'ils parlent des progiciels d'économétrie, "les presses bouton" ou "boite noire". Je pense que ce genre d'expression ne correspond pas du tout à la réalité des faits. Tous les progiciels d'économétrie sont des presses boutons et des boites noires. En effet, en ce qui concerne le terme de presse bouton, que l'on coche une option dans une boite de dialogue ou que l'on rajoute une lettre dans une ligne de scripte, le résultats est finalement le même. Pour ce qui est du terme de boite noire, du moments que l'on utilise une fonction quelconque pour une estimation ou une inférence et que l'on ne possède pas le code source (ce qui, à part dans le monde libre, est toujours le cas!) de la mise en oeuvre de cette fonction, tous les logiciels sont des boites noires. Il me semblerait que la notion de "logiciels de confiance" soit plus en accord avec l'utilisation de ces progiciels. J'entend par notion de confiance, que les résultats qui sont fournis soit, naturellement, corrects et précis masi aussi que, si le code source des différentes fonctions n'est pas fourni, les algorithmes utilisés soit mentionner de la manière la plus précise possible. Ce genre de comportement permétrait, primo un meilleur choix des logiciels selon ses besoins mais aussi une duplication des résultats des publications beaucoup plus facile.

Eviews 4.1 build 18 avril 2002

Eviews est un logiciel qui propose une mise en oeuvre graphique des principaux outils nécessaire à une étude économétrique. Un langage scripte est aussi présent. Les caractéristiques de l'estimation et de l'inférence des modèle $GARCH$ est décrite dans le tableau suivant

EViews 4.1

	Algorithme	Modèle	Variance	Variance initiale
Defaut	Marquardt	arch, garch	numérique	na
Optionnel	BHHH	arch-m, egarch, tgarch	$(I^{-1})OP(I^{-1})$	défini par l'utilisateur

L'utilisation des procédure d'estimation peut se faire soit graphiquement soit par l'appel d'une fonction dans un script. Les deux méthode sont aisé, c'est de loing le logiciel le plus facile à utilisé. L'aide en ligne est bien faite mais pas assez détaillée. On regrettera le choix des option par défaut, qui ne sont vraiment pas adaptées. Les lignes de code pour l'estimation par défaut et pour une estimation plus juste sont :

```
arch(1,1) g c
arch(1,1,z,showopts,h,b,c=0,000001) g c
Matlab 6.5.01,13a release 13, 18 juin 2002.
```

Matlab est avant tout un langage de programmation matriciel qui s'est enrichi au fil des versions de modules, les "toolbox, pour permettre l'analyse numérique de toutes sortes de processus. En ce qui concerne les modèles hétéroscédastiques le module testé est dans sa version 1.0.2.⁵. On peut trouver sur l'intenet des modules developpés par des particuliers. les caractéristique du toolbox *GARCH 1.0.2* sont

MATLAB 6.5 & GARCH 1.0.2

	Algorithme	Modèle	Variance	Variance initiale
Defaut	na	arch, garch	numérique	na
Optionnel	na	na	na	défini par l'utilisateur

Il est claire que pour l'instant ce module ne suffit pas a une bonne étude des processus *GARCH*. L'aide en ligne est complète et très orienté sur l'analyse économétrique d'une série temporelle. La prise en main demandera une certaine habitude, sans être vraiment difficile. On notera bon nombre de fonction non encore mis en oeuvre mais dont le nom laisse prévoir une nette amélioration de ce module. LE code est

```
load dmbp;
sp = garchset('P',1,'Q',1)
[coeff, errors, LLF, innovations, sigma, summary] = garchfit(sp,dmbp);
garchdisp(coeff, errors)
```

Spplus 6,0 academic edition Release 2 - 14 septembre 2001

Spplus est un logiciel d'économétrie pure. Certaines procédures économétrique de base sont accessibles au travers d'une interface graphique. Mais c'est surtout avec le langage de programmation que ce logiciel est le plus performant. La langage, orienté objet, nécessite une certaine adaptation, du fait de la programmation objet. On notera le *projet R* qui est un clone de

⁵Un version 2.0 de ce module devrait être prochainement disponible.

SPLUS mais issus du monde libre. Les possibilité *GARCH* de *Splus* sont fournit par le modula *SGARCH*, tester ici dans sa version 1.3R2

SPLUS 6 & SGARCH 1.3R2

	Algorithme	Modèle	Variance	Variance initiale
Default	BHHH	arch, garch	numérique	na
Optionnel	na	pgarch, egarch, tgarch, arch-m, fgarch	na	défini par l'utilisateur

Il semble que se module est été developpé à la hate pour surement être présent sur le marché. L'aide en ligne est indigente et sans aucune structure. On notera quand même la possibilité de faire du *GARCH* multivarié et du *FIGARCH*. Le code est

```
ga <- dmbp[["col1"]]
ga.mod <- garch.model(ga~1, ~garch(1,1))
print(ga.mod)
ga.fit <- garch(ga~1, ~garch(1,1))
summary.mgarch(ga.fit, max.lag=12)
```

Pc give 3.31 Ox pack 3.0 for givewin2.0

Pc give est un une interface graphique pour l'étude de série chronologique basé sur les routine de *ox*. Les options disponibles sont

PCGIVE 3.31 Ox pack 3.0

	Algorithme	Modèle	Variance	Variance initiale
Default	BFGS	arch, garch	numérique	na
Optionnel	newton quasi newton	egarch, arch-m, tgach	hessienne Information	défini par l'utilisateur

Pc give est sans consteste le logiciel le plus agréable a utilisé. C'est le plus complet en terme d'algorithme et de choix des matrices de variance. Les algorithmes sont fiables et précis. L'aide en ligne est assez courte mais l'ergonomie du logiciel et le nom des fonctions font que cela ne pose pas de problème. On regretera l'absence de script qui empêche le traitement par lot.

Regardons les résultats des logiciels avec leurs options par defaults et les options pour se rapprocher du *benchmark*

Paramètre et t de student								
	b	$t(b)$	α_0	$t(\alpha_0)$	α_1	$t(\alpha_1)$	δ	$t(\delta)$
EViews 4.1	-0.005398	-0.64	0.009571	8.00	0.142216	11.10	0.821442	53.99
	-0.006165	-0.71	0.010782	3.45	0.153254	5.62	0.805775	26.85
G@RCH 3.0	-0.006144	-0.73	0.010770	3.77	0.153378	5.77	0.805848	24.01
MATLAB r13	-0.006190	-0.73	0.010761	8.13	0.15313	10.96	0.80598	48.67
PC-GIVE 3.0	-0.006190	-0.71	0.0107614	3.45	0.153134	5.60	0.805974	26.70
	-0.006190	-0.67	0.0107614	1.66	0.153134	2.86	0.805974	11.10
R-PROJECT								
SAS 8.2	-0.006191	-0.73	0.0108	8.15	0.1524	10.97	0.8058	48.61
S-PLUS 6.0	-0.009187	-1.075	0.011696	8.09	0.154295	10.77	0.800276	45.90

Une première constatation est que malgré le fait que les versions de Matlab et Eviews est évoluer leur estimation par défaut sont toujours les mêmes. PC-give, qui n'avait pas été testé par Brook, Burke et Persand donne de bons résultats, même par défaut. Si l'on coche l'option matrice hessienne, pc-give atteint le benchmark.

9 Conclusion

References

- [1] Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometric*, **31**, 307-27.
- [2] Brook, C., S. P. Burke, G. Persaud (2001). "Benchmarks and accuracy of GARCH model estimation," *International Journal of Forecasting*, **17**, 45-56.
- [3] Davidson, R. (2003). "Maximum Likelihood Estimation by Artificial Regression," *Working paper*.
- [4] Fiorentini, G., G. Calzolari, and L. Panattoni (1996). "Analytic derivatives and the computation of GARCH estimate," *Journal of Applied Econometrics*, **11**, 399-47.
- [5] Lee, S-W., and B. E. Hansen (1994). "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator," *Econometric Theory*, **10**, 29-52.

A Le code C++

Le C++ est un langage de programmation qui se prête bien à la simulation. Bien que de nombreux chercheurs utilisent encore le *Fortran* pour tout ce qui est calcul numérique, pour surement des questions d'habitudes (dans le monde informatique il est souvent difficile de changer sa manière de programmer acquise après de longues heures de programmation), les avantages de performance de ce dernier sont maintenant atteints par certains compilateurs C++. Les chercheurs peuvent maintenant bénéficier à la fois de la souplesse de programmation du C++ et d'une rapidité d'exécution comparable à celle obtenue en *Fortran* qui a été pendant longtemps la norme sur des systèmes monoprocesseurs dont sont équipés la plupart des laboratoires de recherche en économétrie.

Le code mis en oeuvre pour les simulations ne sera pas présenté en détail, car il fait partie d'un projet plus vaste : *Ects4* développé par R. Davidson en C++ qui devrait être disponible dans les prochaines années. Le projet *Ects4* fait suite à la version 3.3, il reprend la même syntaxe et le même but à savoir d'être un progiciel d'économétrie orienté pour l'apprentissage de l'économétrie. La version 4 apporte son lot de nouveautés et en particulier la possibilité de programmer des modules. En effet R. Davidson a complètement décrit les différentes bibliothèques de base et a mis en oeuvre un mécanisme qui permet à chacun de programmer ses propres fonctions ou commandes et de pouvoir les appeler dans un script ect. C'est dans ce cadre que ont été développées les simulations de ce mémoire. La principale bibliothèque utilisée est la bibliothèque matricielle qui à l'heure actuelle est pleinement utilisable. Une autre bibliothèque employée est la bibliothèque d'optimisation qui est toujours en développement. Cette bibliothèque apporte une structure de base pour mettre en oeuvre n'importe quel processus itératif de maximisation ou minimisation.

B Les tableaux de résultats

C Brook, Burke et Persaud

Le tableau suivant reprend les résultats de Brook, Burke et Persaud pour l'estimation de la série *benchmark* par le logiciel testé avec les options par défaut.

	Paramètre et t de student			$t(\alpha_0)$	α_1	$t(\alpha_1)$	δ	$t(\delta)$
	b	$t(b)$	α_0					
EVIIEWS 3.1	-0.00540	-0.64	0.0096	8.01	0.143	11.09	0.821	53.83
GAUSS* 3.2.29	-0.00600	-0.75	0.0110	3.67	0.153	5.67	0.806	23.71
LIMDEP 8.0**	-0.00619	-0.71	0.0108	3.44	0.153	5.61	0.806	26.73
MATLAB 11	-0.00619	-0.73	0.0108	8.13	0.153	10.96	0.806	48.67
MICROFIT 4.0	-0.00621	-0.73	0.0108	3.78	0.153	5.78	0.806	24.02
RATS 4.3	-0.00619	-0.74	0.0108	8.15	0.153	10.97	0.806	48.60
SAS 6.12	-0.00613	-0.73	0.0107	5.58	0.154	7.91	0.806	36.96
SHAZAM 8.0	-0.00625	-0.71	0.0108	3.76	0.153	5.79	0.806	23.93
TSP	-0.00619	-0.67	0.0108	1.66	0.153	2.86	0.806	11.11

*gauss avec FANPAC 1.1.11/2

** version d'essais à l'époque

	Paramètre et t de student			$t(\alpha_0)$	α_1	$t(\alpha_1)$	δ	$t(\delta)$
	b	$t(b)$	α_0					
EVIIEWS 4.1	-0.005398	-0.64	0.009571	8.00	0.142216	11.10	0.821442	53.99
	-0.006165	-0.71	0.010782	3.45	0.153254	5.62	0.805775	26.85
G@RCH 3.0 ox	-0.006144	-0.73	0.010770	3.77	0.153378	5.77	0.805848	24.01
MATLAB 6.5r13	-0.0061902	-0.73	0.010761	8.13	0.15313	10.96	0.80598	48.67
PC-GIVE 3.0	-0.00619040	-0.71	0.0107614	3.45	0.153134	5.60	0.805974	26.70
	-0.00619040	-0.67	0.0107614	1.66	0.153134	2.86	0.805974	11.10
R-PROJECT								
SAS 8.2								
S-PLUS 6.0	-0.009187	-1.075	0.011696	8.09	0.154295	10.77	0.800276	45.90

Contents

1	Introduction	2
2	Le modèle GARCH(1,1)	2
3	Le maximum de vraisemblance	3
3.1	Fonction de log-vraisemblance	3
3.2	Consistence et normalité asymptotique	3
3.3	Gradient du modèle GARCH(1,1)	4
4	La maximisation de la vraisemblance	6
4.1	La méthode du gradient	6
4.2	La régression artificielle	7
5	Les différents choix initiaux	13
5.1	Le choix initial compatible avec la stationnarité	13
5.2	Les différentes méthodes	13
6	Les simulations	17
6.1	Les résultats des simulations	19
7	La matrice de covariance et l'inférence	22
8	Le benchmark et les autres progiciels	22
9	Conclusion	26
	References	27
A	Le code C++	28
B	Les tableaux de résultat	28
C	Brook, Burke et Persand	28
	Contents	30