HEC MONTRÉAL

Mesure de la Valeur à Risque
en utilisant des données transaction par transaction sur
des contrats à terme transigés à la bourse de Chicago

Par

Hassen Mseddi

Sciences de la gestion
(Option finance)

Mémoire présenté en vue de l'obtention
du grade de maître ès sciences
(M.Sc.)

Avril 2008
© Hassen Mseddi, 2008

RÉSUMÉ

Dans le milieu de la finance, l'évolution des technologies d'enregistrement des données financières a permis aux négociateurs d'avoir une nouvelle information qui provient des données à haute fréquence. Ce type d'information permet de prendre des décisions pour transiger d'une façon très active sur le marché. Toutefois, les institutions financières sont confrontées à un risque de marché à l'intérieur de la journée. Pour cette raison, un besoin essentiel d'une mesure de risque intra journalière s'impose. Dans cette étude, nous avons poursuivi le travail de recherche de Dionne, Duchesne et Pacurar (2005). Nous avons appliqué leur nouvelle méthode paramétrique pour calculer une valeur à risque intra journalière sur le marché des contrats à terme des obligations gouvernementales américaines de maturité 10 ans. Tout d'abord, nous avons présenté notre modèle économétrique basé sur une combinaison d'un modèle de durée « ACD » et d'un modèle ARMA-GARCH afin de modéliser les durées entre les transactions et les rendements, respectivement. Ensuite, nous avons montré que les données à haute fréquence contiennent de la saisonnalité intra journalière et inter journalière. Enfin, nous avons estimé notre modèle en utilisant la méthode du maximum de vraisemblance et nous avons appliqué la technique de simulation de Monte Carlo pour calculer la valeur à risque intra journalière.

Mots clés : modèle de durée, GARCH, microstructure de marché , saisonnalité intra journalière et inter journalière, méthode du maximum de vraisemblance, volatilité, la valeur à risque intra journalière.

TABLE DES MATIÈRES

Page

RÉSUMÉ…… ii

TABLE DES MATIÈRES……... iii

LISTE DES TABLEAUX…… v

LISTE DES GRAPHIQUES……..... vi

REMERCIEMENTS……... vii

INTRODUCTION…….. 8

CHAPITRE I : LA REPRÉSENTATION ÉCONOMÉTRIQUE DU MODÈLE « ACD-UHF-GARCH »……... 12

1.1 Présentation de la densité jointe des durées et des rendements…….. 12

1.2 Les modèles de durées « ACD »……..... 14

1.2.1 Présentation …… 14

1.2.2 Quelques simulations sur les modèles de durée «ACD » …. 18

1.3 Le modèle « UHF-GARCH » ……. 20

1.3.1 Présentation du modèle…… ……. 20

1.3.2 Simulation du modèle WACD(1,1)-ARMA(1,1)-UHF-GARCH(1,1)……... 25

CHAPITRE II : PROPRIÉTÉS STATISTIQUES DES DONNÉES À HAUTE FRÉQUENCE…… 27

2.1 Description et statistiques descriptives des données intra journalières…….. 27

2.2 Saisonnalité et ajustement des données à haute fréquence……. 29

2.3 Statistiques descriptives des données ajustées…… 32

Page

CHAPITRE III : ESTIMATION DU MODÈLE ÉCONOMÉTRIQUE, SIMULATIONS ET PRÉVISIONS……..... 33

3.1 Méthodologie d'estimation et interprétation des résultats …….. 33

3.2 Adéquation statistique du modèle……..... 35

3.3 Valeur à Risque Intra journalière (IVaR)…… 37

3.3.1 Définition de la Valeur à Risque Intra journalière ……... 37

3.3.2 Valeur à risque et simulation Monte Carlo intra journalière.. 38

3.3.2 Analyse de la performance du modèle hors échantillon……. 40

CONCLUSION…….. 42

BIBLIOGRAPHIE…… 44

ANNEXE I…… 47

ANNEXE II…….. 52

ANNEXE III…… 54

ANNEXE IV…… 61

ANNEXE V…… 65

ANNEXE VI……. 66

LISTES DES TABLEAUX

Annexe I

Tableau I :

Résultat d'estimation du modèle WACD(1,1) simulé pour 5000 observations……. 51

Annexe II

Tableau II :

Résultats de l'estimation du modèle *UHF-ARMA(1,1)-GARCH(1,1)* en utilisant 5000 observations simulées 1000 fois à partir du modèle *WACD(1,1)-UHF-ARMA(1,1)-GARCH(1,1)*……..52

Annexe III

Tableau III.1 :

Statistiques descriptives des données à haute fréquence brutes……...54

Tableau III.2 :

Statistiques descriptives des données à haute fréquence désaisonnalisées…….....58

Annexe IV

Tableau I V.1:

Résultats de l'estimation du modèle *WACD(1,1)- ARMA(1,1)- UHF-GARCH(1,1)*……..61

Tableau IV.2:

Résultats de la régression des termes d'erreurs provenant du modèle ARMA(1,1)-UHF-GARCH(1,1) sur les termes d'erreurs standardisées provenant du modèle WACD(1,1) et sur les trois retards de la durée……...62

Annexe VI

Tableau IVI :

P-value du test de Kupiec……66

LISTES DES GRAPHIQUES

Annexe I

Graphique I.1: Simulation du modèle *WACD(1,1)*……...47

Graphique I.2: Simulation du modèle *GACD(1,1)*……...48

Graphique I.3: Simulation du modèle *log-WACD(1,1)*……....49

Graphique I.4: Simulation du modèle *log- GACD(1,1)*……...50

Annexe II

Graphique II : Rendements, durées et volatilité conditionnelle transaction par transaction simulés à partir d'un *WACD(1,1)-ARMA(1,1)-UHF-GARCH(1,1)* et fonction d'autocorrélation totale pour les résidus standardisés du processus des rendements. …….53

Annexe III

Graphique III.1 : Histogrammes des rendements intra journalières et histogrammes des durées intra journalières…….55

Graphique III.2 : Facteurs intra journaliers estimés pour les durées……56

Graphique III .3: Facteurs intra journaliers estimés pour les rendements au carré……..57

Graphique III.4 : Séries des données à haute fréquence désaisonnalisées……59

Graphique III.5 : Histogramme des durées ajustées.……....60

Annexe IV

Graphique IV.1 : Densité empirique des erreurs standardisés provenant du modèle WACD(1,1) versus la densité d'une distribution « Weibull »…..63

Graphique IV.2: Densité empirique des erreurs standardisés provenant du modèle ARMA(1,1)-UHF-GARCH(1,1) versus la densité d'une distribution Normale…….64

Annexe V

Graphique V : Rendements Intra journalier Versus IVaR (Intervalle =45 ou 22 minutes, niveau de confiance de 95%)……...65

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche monsieur Georges Dionne qui, tout au long de ce travail, a fait preuve de disponibilité et de patience. Je suis très reconnaissant pour la confiance qu'il m'a accordée.

De plus, je suis très reconnaissant de l'aide financière qui m'a été offerte par le Centre de Recherche en e-Finance (CREF) et la Chaire de recherche du Canada en gestion des risques.

Je saisis cette occasion pour exprimer ma profonde gratitude à mes parents pour tous leurs sacrifices et leur dévouement. Malgré la grande distance qui nous a séparés, leur affection et leurs prières ont éclairé mon chemin.

INTRODUCTION

Récemment, le marché financier a été doté d'un système électronique d'enregistrement des données sur les transactions. D'habitude, ces dernières sont enregistrées avec une fréquence journalière. Toutefois, la présence de ce système a favorisé l'enregistrement instantané des informations sur les transactions comme le prix, le volume, etc. En effet, l'augmentation spectaculaire de la vitesse des transactions et le développement de ce système ont favorisé la naissance d'un intérêt dans la littérature financière sur la modélisation des données à haute fréquence. L'accès à ces données financières est devenu possible à moindre coût pour les analystes financiers, ce qui a crée un nouveau type d'information financière. Toutefois, ces données sont fournies transaction par transaction et la durée entre deux transactions consécutives n'est pas constante. On ne peut pas appliquer les techniques économétriques standards sur les séries temporelles puisque l'application de telles techniques est conçue pour des données à un intervalle de temps fixe.

Afin de résoudre ce problème, Engle et Russell (1997,1998) ont introduit un modèle qui tient compte de l'espacement irrégulier dans le temps des données à haute fréquence. Il s'agit du modèle de durée conditionnelle autorégressive ou « Autoregressive Conditional Duration » (ACD). Pour pouvoir estimer plus précisément la volatilité conditionnelle à haute fréquence, Engle (2000) a combiné ce modèle avec un GARCH « Generalized Autoregressive Conditional Heteroskedasticity » introduisant ainsi dans la littérature un nouveau modèle intitulé « Ultra High Frequency- GARCH model» (UHF-GARCH) . L'avantage de ce dernier est qu'on pourra modéliser les rendements tout en considérant les durées et

l'espacement irrégulier dans le temps des données à haute fréquence. Ce modèle nous permet d'analyser l'impact de la durée sur la volatilité. En effet, une longue durée entre les transactions se traduit par une faible volatilité puisque il n'y a pas assez d'information sur le marché financier pour pouvoir faire varier le prix. Selon les modèles théoriques de microstructure du marché, ce mécanisme demeure ambigu. Selon Easley et O'Hara (1987), si les négociateurs transigent en se basant sur l'apparition d'une nouvelle information, alors pas de transaction signifie qu'il n y a pas un événement sur le marché qui fait influencer le prix. Ainsi, une longue durée est associée à une faible volatilité. Toutefois, la liquidité assez faible sur le marché du titre en question peut aussi causer cette longue durée reflétant ainsi une forte volatilité. Selon O'Hara (1995), l'importance de l'impact de la durée sur la volatilité est une question purement empirique.

Engle et Sun (2005) ont traité cette question et ils ont trouvé que la durée courante influence faiblement la volatilité conditionnelle par transaction. Donc, la volatilité conditionnelle par seconde est inversement reliée à la durée. Ce résultat vient de confirmer les résultats trouvés par Easley et O'Hara (1987), Engle (2000).

Engle et Sun (2005) et Dionne, Duchesne et Pacurar (2005) ont appliqué un modèle semblable à celui d'Engle (2000). Toutefois, leurs objectifs étaient différents. Engle et Sun (2005) ont appliqué ce type de modèle pour prévoir la volatilité. Toutefois, Dionne, Duchesne et Pacurar (2005) l'ont appliqué pour pouvoir simuler une valeur à risque intra journalière « IVaR » pour trois titres cotés à la bourse de Toronto (Banque Royale, Placer Dome et Alcan). La valeur à risque (VaR) mesure l'espérance de la perte maximale pour un horizon donné, sous des conditions normales de marché et à un certain niveau de confiance (Jorion, 2000, p.xxii). En plus, elle est maintenant considérée comme une valeur de référence utilisée par les institutions financières. En effet, ces dernières produisent généralement leur VaR du marché à la fin de la journée pour mesurer leur risque de marché pour le prochain jour.

Ce mémoire s'inscrit dans le nouveau domaine de la recherche en finance : la finance à haute fréquence. Il consiste à appliquer la méthodologie de calcul d'une Valeur à Risque intra journalière, récemment introduite dans la littérature par Dionne, Duchesne et Pacurar (2005), pour les actions aux produits dérivés, en particulier les contrats à terme sur les obligations gouvernementales américaines de 10 ans. Holder *et al* (2004) ont analysé les données transaction par transaction sur ces contrats à terme et ils ont utilisé une approche similaire à celle de Dufour et Engle (2000) dans le but d'identifier l'impact de la durée entre les transactions sur les prix.

En appliquant la méthodologie de Dionne *et al* (2005) sur le marché des titres à revenu fixe, cette étude est une extension de leur étude. D'abord, au niveau du processus décrivant la dynamique des durées, nous avons supposé que les termes d'erreurs suivent une distribution « Weibull ». Cette dynamique est représentée par un simple modèle de durée « WACD(1,1) ». Ensuite, au niveau du processus décrivant la dynamique des rendements, nous avons appliqué un simple « ARMA(1,1)-GARCH(1,1) ». Toutefois, en spécifiant qu'une seule fonction de vraisemblance, l'estimation de ces deux processus est faite de façon simultanée.

Nous avons détecté une forme de saisonnalité dans les durées et les rendements que nous avons ensuite ajustée. Puis nous avons estimé notre modèle économétrique et nous avons appliqué des tests d'adéquation statistique. Enfin, nous avons mesuré la IVaR en utilisant les techniques de simulation de Monte Carlo.

L'importance de ce travail découle principalement de l'originalité des données transaction par transaction sur des contrats à terme. Selon notre connaissance, il s'agit de la première application des modèles de type ACD sur le marché des contrats à terme. La plupart des études utilisent des données irrégulièrement espacées portent sur le marché des actions alors qu'il est connu dans la littérature que la structure institutionnelle d'un marché peut affecter le comportement des agents et le processus de formation des prix. De plus, les contrats à terme sur obligations sont souvent

transigés par les gestionnaires de portefeuille à des fins de spéculation grâce à l'effet de levier qu'ils offrent.

Dans le premier chapitre, nous présentons notre modèle économétrique (ACD-UHF-GARCH) et nous pratiquons aussi quelques simulations pour valider ce type de modèle. Dans le deuxième chapitre, nous exposons les propriétés statistiques des données à haute fréquence. Nous formulons aussi la méthodologie appliquée afin de désaisonnaliser ces données. Dans le troisième chapitre, nous procédons à l'estimation de notre modèle en utilisant des données déjà ajustées. Ensuite, nous validons notre modèle en appliquant des tests d'adéquation statistiques, nous procédons à des simulations pour mesurer la IVaR et nous appliquons un test de performance hors échantillon de notre modèle. Enfin, une conclusion est présentée.

CHAPITRE I
LA REPRÉSENTATION ÉCONOMÉTRIQUE DU MODÈLE « ACD-UHF-GARCH »

1.1 Présentation de la densité jointe des durées et des rendements

Nous définissons la durée entre deux transactions consécutives et le rendement continuellement composé correspondant de la façon suivante :

$$x_i = t_i - t_{i-1}$$

$$r_i = \log\left(\frac{p_i}{p_{i-1}}\right)$$

où x_i est la durée et r_i est le rendement entre la transaction i et i-1. Donc, notre objectif consiste à modéliser le processus décrit par la séquence suivante :

$$\{(x_i, r_i), i = 1, \ldots\ldots\ldots, n\}$$

Toutefois, selon Engle (2000), les rendements et les durées sont interdépendantes. Pour cette raison, nous devrions estimer notre modèle en spécifiant la distribution jointe des durées et des rendements.

Pour la $i^{ème}$ observation, la densité jointe de $(x_i, r_i) | F_{i-1}$ est définie de la façon suivante :

$$f\left(x_i, r_i | \tilde{x}_{i-1}, \tilde{r}_{i-1}; \theta\right) \approx g\left(x_i | \tilde{x}_{i-1}, \tilde{r}_{i-1}; \theta_1\right) q\left(r_i | x_i, \tilde{x}_{i-1}, \tilde{r}_{i-1}; \theta_2\right)$$

où F_{i-1} est l'ensemble d'information disponible à la $(i\text{-}1)^{ème}$ observation, $\tilde{r}_{i-1}$ est une séquence de rendements passés jusqu'à la $(i\text{-}1)^{ème}$ observation , $\tilde{x}_{i-1}$ correspond à l'historique des durées , $g\left(x_i | \tilde{x}_{i-1}, \tilde{r}_{i-1}; \theta_1\right)$ est la densité marginale de la durée x_i conditionnelle aux durées et aux rendements passés, où θ_1 est défini comme étant le vecteur de paramètres à estimer pour décrire le processus des durées .

$q\left(r_i | x_i, \tilde{x}_{i-1}, \tilde{r}_{i-1}; \theta_2\right)$ représente la densité conditionnelle du rendement r_i, conditionnelle aux durées passés, aux rendements passés et à la durée courante x_i, où θ_2 est défini comme étant le vecteur de paramètres à estimer pour décrire le processus des rendements.

Supposant que les paramètres θ_1 et θ_2 sont finis et invariants à travers les événements, nous pouvons estimer notre modèle général par la méthode de maximum de vraisemblance. En effet, nous pouvons maximiser la fonction logarithmique de vraisemblance qui est définie de la façon suivante :

$$l\left(\theta_1, \theta_2\right) = \sum_{i=1}^{n} \left[\log\left(g\left(x_i | \tilde{x}_{i-1}, \tilde{r}_{i-1}; \theta_1\right)\right) + \log\left(q\left(r_i | x_i, \tilde{x}_{i-1}, \tilde{r}_{i-1}; \theta_2\right)\right)\right]$$

Si nous considérons que les durées sont des variables exogènes, nous pouvons alors maximiser séparément les deux fonctions de vraisemblance. Toutefois, en

analysant les données à hautes fréquence pour cinq titres transigés à la bourse de New York, Dolado, Rodriguez-Poo et Veredas (2004) ont rejeté l'hypothèse nulle d'exogèniétè des durées. Dans notre étude, pour contourner ce problème, nous maximisons simultanément les deux fonctions logarithmiques de vraisemblance.

Dans les prochaines sections, nous allons tout d'abord présenter les différents types de modèles décrivant la dynamique des durées. Nous allons ensuite présenter le modèle introduit par Engle(2000), intitulé « UHF-GARCH », décrivant la dynamique des rendements transaction par transaction et enfin nous allons présenter une extension de ce modèle introduite par Dionne, Duchesne et Pacurar (2005).

1.2 Les modèles de durées « ACD »

1.2.1 Présentation

Ces modèles sont introduits par Engle et Russell (1997). Ces derniers ont supposé que la dépendance des durées peut être modélisée en appliquant le même principe qu'un GARCH décrivant la dépendance de la volatilité. En effet, le modèle ACD est basé sur l'hypothèse que la dépendance entre les durées est capturée par leur moyenne conditionnelle.

Ils spécifient la durée observée comme étant un processus mixte qui est défini de la façon suivante :

$$x_i = \psi_i \varepsilon_i$$

où ε_i sont des variables aléatoires positives, indépendamment et identiquement distribuées *(iid)* avec une densité $p(\varepsilon)$:

$$E(\varepsilon_i) = 1 \ , \ Var(\varepsilon_i) = \sigma^2$$

$$E(x_i \mid H_{i-1}) = \psi_i$$

En choisissant différentes spécifications pour la durée conditionnelle ψ_i et différentes densités $p(\varepsilon)$, nous pouvons obtenir plusieurs formes de modèles décrivant la dynamique des durées.

Le modèle le plus populaire est le « *ACD(m,q)* » basé sur la paramétrisation linéaire suivante :

$$\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j},$$

où ω, α_j *et* β_j doivent être non-négatifs. Selon Engle et Russell (1997), ces conditions sont suffisantes mais pas nécessaires pour assurer la positivité des durées conditionnelles.

Généralement, quatre densités $p(\varepsilon)$ ont été documentées et utilisées afin de modéliser la moyenne conditionnelle des durées. Il s'agit des distributions « Gamma généralisée », « Weibull », « Exponentielle » et « Burr ». Nous allons présenter les

trois premières densités de la forme la plus générale à la forme la plus particulière. Pour une distribution « Gamma généralisée », la fonction de densité est la suivante :

$$f\left(\varepsilon \mid \gamma_1, \gamma_2\right) = \frac{\gamma_1 \varepsilon^{\gamma_1 \gamma_2 - 1}}{\gamma_3^{\gamma_1 \gamma_2} \Gamma\left(\gamma_2\right)} EXP\left[-\left(\frac{\varepsilon}{\gamma_3}\right)^{\gamma_1}\right], \quad \varepsilon \succ 0$$
$$= 0 \qquad , si\, non$$

$$\gamma_3 = \frac{\Gamma\left(\gamma_2\right)}{\Gamma\left(\gamma_2 + \frac{1}{\gamma_1}\right)}, \gamma_1 \succ 0, \gamma_2 \succ 0.$$

$\Gamma(.)$ est la fonction « Gamma » définie de la façon suivante pour $k>0$:

$$\Gamma(k) = \int_0^\infty x^{k-1} EXP\left(-x\right) dx \cdot$$

Pour avoir la densité de la distribution « Weibull », il faut que γ_2 soit égal à un. De même, pour avoir la densité de la distribution « Exponentielle », il faut que les paramètres γ_2 et γ_1 soient égaux à un.

Pour décrire le processus de ε , Engle et Russell (1998) ont utilisé les densités des distributions « Exponentielle » et « Weibull » résultant des modèles de durées intitulés ACD exponentielle (EACD) et Weibull ACD (WACD), respectivement. Toutefois, en analysant des données empiriques, Engle et Russell (1998) ont rejeté la spécification exponentielle car la fonction hasard d'une distribution exponentielle est constante dans le temps alors que le modèle WACD a une fonction plutôt monotone. En effet, le modèle WACD est préférable puisque la fonction d'intensité observée sur les transactions n'est pas constante mais plutôt monotone dans le temps. Pour avoir

plus de flexibilité dans la fonction hasard, Lunde (1999) propose la distribution « Gamma généralisée » et Grammig et Maurer (2000) propose la distribution « Burr ».

Dans notre étude, nous allons appliquer le modèle *WACD(1,1)* afin de modéliser la dynamique des durées. Ainsi, notre modèle est représenté de la façon suivante :

$$\psi_i = \omega + \alpha\ x_{i-1} + \beta\ \psi_{i-1}, \quad \omega > 0, \alpha > 0, \ \beta > 0.$$

Si ε_i suit une distribution « Weibull », sa fonction densité est donc définie de la façon suivante :

$$f\left(\varepsilon \mid \gamma\right) = \frac{\gamma\ \varepsilon^{\gamma-1}}{\lambda^{\gamma}} EXP\left[-\left(\frac{\varepsilon}{\lambda}\right)^{\gamma}\right], \quad \varepsilon > 0$$
$$= 0 \qquad ,si\ non$$

$$\lambda = \frac{1}{\Gamma\left(1+\frac{1}{\gamma}\right)}, \gamma > 0,.$$

Bauwens et Giot (2000) ont introduit une version logarithmique du modèle ACD « log-ACD » représenté de la façon suivante :

$$\psi_i = EXP\left(\omega + \alpha\ \varepsilon_{i-1} + \beta\ \psi_{i-1}\right).$$

En effet, l'avantage de cette version est qu'elle nous assure la positivité de la durée conditionnelle sans avoir de restrictions sur les paramètres du modèle ACD. Toutefois, pour s'assurer de la stationnarité du processus, il faut que $|\beta|<1$.

1.2.2 Quelques simulations sur les modèles de durée «ACD »

Pour illustrer le processus ACD, nous avons généré 5000 observations à partir de modèle *ACD(1,1)* en utilisant deux différentes distributions des ε_i « Weibull » et « Gamma généralisée ».

Notre modèle calibré se présente de la façon suivante :

$$x_i = \psi_i \varepsilon_i$$

$$\psi_i = 0.3 + 0.2\ x_{i-1} + 0.7\ \psi_{i-1}$$

Le graphique I.1 dans l'annexe I nous montre les durées simulées à partir d'un WACD(1,1)[1] ainsi que les séries standardisées (ϵ_i), l'histogramme des séries simulées et enfin la fonction d'autocorrélation pour les séries standardisées. En analysant la fonction d'autocorrélation, nous remarquons qu'il existe une dépendance temporelle des durées.

Le graphique I.2 dans l'annexe I nous expose les séries simulées à partir d'un GACD(1,1)[2] ainsi que les séries standardisées, l'histogramme des séries simulées et enfin la fonction d'autocorrélation pour les séries standardisées.

---

[1] Le paramètre choisi pour la distribution « Weibull » est égal à 1.5 (γ=1,5).
[2] Les paramètres choisis pour la distribution « Gamma généralisée » sont égaux à 0.5 et 1.5 (γ_1=0.5, γ_2=1,5).

Les graphiques I.3 et I.4 montrent la même analyse que précédemment mais appliquée sur des séries simulées à partir d'une version logarithmique du ACD.

Après avoir simulé des données à partir des modèles de durée, nous allons maintenant montrer comment nous pouvons les estimer en utilisant la méthode du maximum de vraisemblance. Pour le modèle WACD, nous utilisons la fonction de densité d'une distribution « Weibull » et nous obtenons ensuite la fonction logarithmique de vraisemblance conditionnelle représentée par :

$$L(x \mid \theta_1) = \sum_{i=i_0+1}^{T} \gamma \log\left[\Gamma\left(1+\frac{1}{\gamma}\right)\right] + \log\left(\frac{\gamma}{x_i}\right) + \gamma \log\left(\frac{x_i}{\psi_i}\right) - \left[\frac{\Gamma\left(1+\frac{1}{\gamma}\right)x_i}{\psi_i}\right]^{\gamma}$$

Pour le modèle GACD, nous utilisons plutôt la fonction de densité d'une distribution « Gamma généralisée » et la fonction de vraisemblance conditionnelle est définie de la façon suivante :

$$L(x \mid \theta_1) = \sum_{i=i_0+1}^{T} \log\left(\frac{\gamma_1}{\Gamma(\gamma_2)}\right) + (\gamma_1\gamma_2 - 1)\log(x_i) - \gamma_1\gamma_2 \log(\gamma_3\psi_i) - \left(\frac{x_i}{\gamma_3\psi_i}\right)^{\gamma_1}$$

Sous certaines conditions, les paramètres, estimés par maximisation de la fonction de vraisemblance conditionnelle, suivent asymptotiquement une loi normale (voir Engle et Russell (1998) pour plus de détails).

Pour valider notre fonction de vraisemblance conditionnelle, nous avons estimé 1000 fois le modèle *WACD(1,1)* en utilisant 5000 observations simulées. Nous avons trouvé que la moyenne des 1000 paramètres estimés est presque égale à la vraie

valeur (voir le tableau I de l'annexe I). Ainsi, nous avons juste validé notre fonction de vraisemblance que nous allons maximiser plus tard en utilisant des données réelles.

Après avoir présenté le modèle *WACD(1,1)* décrivant la dynamique des durées, nous allons présenter dans la prochaine section le modèle « UHF-GARCH » décrivant la dynamique des rendements transaction par transaction.

1.3 Le modèle « UHF-GARCH »

1.3.1 Présentation du modèle

Dans la section précédente, nous avons analysé les durées conditionnelles aux informations passées. Dans cette section, nous allons modéliser les rendements et leur volatilité correspondante pour une durée courante donnée et pour des informations passées données.

Puisque les données à haute fréquence sont irrégulièrement espacées dans le temps, nous ne pouvons pas appliquer l'approche économétrique standard appliquée sur les séries temporelles. Pour cette raison, Engle (2000) propose un modèle GARCH qui tient compte de cette irrégularité. La variance conditionnelle est définie de la façon suivante :

$$h_i = V_{i-1}(r_i \mid x_i)$$

où les informations conditionnelles contiennent la durée courante et les rendements ainsi que les durées passées.

Dans les travaux de recherche effectués par Engle (2000), Meddahi, Renault et Werker (2006), la mesure de volatilité considérée n'est pas la volatilité traditionnelle mais plutôt une volatilité par unité de temps définie de la façon suivante :

$$V_{i-1}\left(\frac{r_i}{\sqrt{x_i}} \mid x_i\right) = \sigma_i^2 \Leftrightarrow h_i = x_i \sigma_i^2$$

La volatilité par unité de temps est modélisée comme étant un simple processus GARCH(1,1) défini de la façon suivante :

$$\sigma_i^2 = \tilde{w} + \tilde{\alpha}\left(\frac{e_{i-1}}{\sqrt{x_{i-1}}}\right)^2 + \tilde{\beta}\,\sigma_{i-1}^2.$$

Pour s'assurer que ce processus soit stationnaire et que la variance conditionnelle soit positive, il faut que ces conditions soient respectées :

$$\tilde{\alpha} + \tilde{\beta} < 1, \tilde{\omega} > 0, \tilde{\alpha} > 0 \; et \; \tilde{\beta} > 0.$$

Les termes d'erreur standardisés sont représentés par :

$$z_i = \frac{e_i}{\sqrt{h_i}} = \frac{\left(r_i - \mathrm{E}\left(r_i \mid F_{i-1}\right)\right)}{\sqrt{h_i}}$$

Ces derniers sont supposés suivre une distribution normale. De plus, ils ne sont pas corrélés avec les résidus passés.

$$\left(z_i \mid F_{i-1} \overset{iid}{\approx} N(0,1)\right) \text{ et } \left(E_t\left[z_i \cdot e_{i-s}\right] = 0, \forall\, s > 0\right).$$

Une version plus générale de ce modèle est introduite par Engle et Sun (2005) et elle est représentée de la façon suivante :

$$V_{i-1}\left(\frac{r_i}{\sqrt{f(x_i)}} \mid x_i\right) = \sigma_i^2 \Leftrightarrow h_i = f(x_i)\sigma_i^2$$

$$\sigma_i^2 = \tilde{w} + \tilde{\alpha}\left(\frac{e_{i-1}}{\sqrt{f(x_{i-1})}}\right)^2 + \tilde{\beta}\,\sigma_{i-1}^2,\ \tilde{w} > 0,\ \tilde{\alpha} > 0,\ \tilde{\beta} > 0.$$

Puisque la forme de la fonction $f(x_i)$ est inconnue, nous devons trouver une façon de l'estimer. Comme nous pouvons le constater, Engle (2000), Meddahi, Renault et Werker (2006) ont supposé que $f(x_i) = x_i$.

Toutefois, Dionne, Duchesne et Pacurar (2005) ont estimé cette fonction en utilisant une approche paramétrique plus générale et ils ont supposé que :

$$f(x_i) = x_i^{\upsilon}$$

$$h_i = x_i^{\upsilon} \sigma_i^2$$

où v est un paramètre qui détermine la pondération de la durée dans la volatilité.

Engle et Sun(2005) ont utilisé une approche semi-nonparamétrique pour estimer la fonction $f(x_i)$. En effet, ils ont utilisé la méthode de « sieves » (voir Grenander (1981) et Ai et Chen (2003)).

Dans notre étude, nous allons suivre l'approche paramétrique appliquée par Dionne, Duchesne et Pacurar (2005) et nous allons utiliser un simple processus GARCH(1,1) pour modéliser la volatilité par unité de temps. Dans les études de Dionne, Duchesne et Pacurar (2005) et Engle et Sun (2005), un processus EGARCH(1,1) introduit par Nelson (1991) a été utilisé. Le premier avantage de ce dernier est qu'il tient compte de l'effet d'asymétrie où la variance conditionnelle réagit différemment aux chocs positifs, négatifs et à la taille des chocs. Le deuxième avantage est l'absence de contrainte de positivité des paramètres, ce qui simplifie l'optimisation.

Ce modèle se présente de la façon suivante :

$$\log\left(\sigma_i^2\right) = \widetilde{w} + \widetilde{\beta}\log\left(\sigma_{i-1}^2\right) + a\left[\frac{|e_{i-1}|}{\sqrt{f(x_{i-1})}\sigma_{i-1}} - E\left(\frac{|e_{i-1}|}{\sqrt{f(x_{i-1})}\sigma_{i-1}}\right)\right] + \widetilde{\alpha}\left(\frac{e_{i-1}}{\sqrt{f(x_{i-1})}\sigma_{i-1}}\right),.$$

Pour s'assurer que ce processus soit stationnaire, il faut que cette condition soit respectée : $\left|\widetilde{\beta}\right| < 1$.

Dans la littérature de la microstructure du marché, il est très connu que les rendements sont fortement autocorrélés à cause de la présence des effets de la microstructure de marché comme par exemple l'effet des transactions non synchronisées et l'effet « Bid-Ask bounce » (voir Tsay (2002)). Ce dernier effet s'explique par le fait que le prix d'un titre n'est pas unique mais dépend plutôt de la direction des transactions qu'il s'agisse d'achat ou de vente. En effet, selon Roll(1984), lorsqu'on analyse des données à haute fréquence, cet effet induit une autocorrélation négative de premier ordre sur les rendements. Ainsi, pour éliminer les effets de la microstructure du marché, nous allons appliquer un processus ARMA(1,1)[3] sur les rendements transaction par transaction présenté de la façon suivante :

$$r_i = c + \varphi\, r_{i-1} + e_i + \theta\, e_{i-1}$$

[3] Processus autorégressif à moyenne mobile.

Pour estimer notre modèle ARMA(1,1)-UHF-GARCH(1,1) par la méthode du maximum de vraisemblance, nous allons maximiser la fonction logarithmique de vraisemblance conditionnelle suivante:

$$L(h \mid \theta_2) = -\frac{1}{2}\left[(T-1)\log(2\pi) + \sum_{i=i_0+1}^{T}\log(h_i) + \sum_{i=i_0+1}^{T}\left(\frac{e_i}{\sqrt{h_i}}\right)^2\right]$$

$$= -\frac{1}{2}\left[(T-1)\log(2\pi) + \sum_{i=i_0+1}^{T}\log(x_i) + \sum_{i=i_0+1}^{T}\log(\sigma_i^2) + \sum_{i=i_0+1}^{T}\left(\frac{e_i}{\sqrt{x_i}\,\sigma_i}\right)^2\right]$$

1.3.2 Simulation du modèle WACD(1,1)-ARMA(1,1)-UHF-GARCH(1,1)

L'objectif de cette simulation consiste à valider notre fonction de vraisemblance pour que nous puissions ensuite la maximiser en utilisant des données réelles. Tout d'abord, nous avons simulé 500 observations de durées à partir d'un processus WACD(1,1). Ensuite, nous avons simulé 500 observations de rendements transaction par transaction à partir d'un processus ARMA(1,1)-UHF-GARCH(1,1). Par la suite, nous avons estimé ce processus en maximisant la fonction logarithmique de vraisemblance décrite ci haut. Enfin, nous avons répété cette procédure 100 fois et nous avons calculé la moyenne de chaque paramètre pour examiner s'ils correspondent bien à leur vraie valeur utilisée pour simuler les données des durées et des rendements (voir les résultats des simulations dans l'annexe II). En observant les résultats des simulations, nous pouvons valider notre fonction de vraisemblance, puisque les valeurs des paramètres estimés sont très proches à leur vraie valeur.

Dans le graphique II de l'annexe II, nous remarquons qu'il existe un lien étroit entre la variance conditionnelle transaction par transaction et la durée. En effet, notre objectif est de pouvoir intégrer l'effet de la durée sur la volatilité. En plus, en analysant la fonction d'autocorrélation, nous avons constaté que les résidus

standardisés ne sont pas autocorrélés. Nous pouvons donc valider statistiquement notre modèle.

Dans le prochain chapitre, nous allons présenter les propriétés statistiques des données réelles à haute fréquence ainsi que la méthodologie appliquée afin de les désaisonnaliser pour pouvoir enfin estimer notre modèle.

CHAPITRE II
PROPRIÉTÉS STATISTIQUES DES DONNÉES À HAUTE FRÉQUENCE

2.1 Description et statistiques descriptives des données intra journalières

Dans cette étude, nous analysons des données à haute fréquence fournies par la compagnie TickData[4]. Il s'agit des données sur des transactions portant sur des contrats à terme sur des obligations gouvernementales américaines dont l'échéance est de 10 ans. Nous avons choisi la période du 1 septembre jusqu'au 30 novembre 2006 contenant 91 jours d'activité boursière. Ce marché est sélectionné principalement pour sa liquidité, de plus, il permet d'étendre l'analyse des données à haute fréquence pour les produits dérivés. En effet, les contrats à terme sur les obligations gouvernementales américaines sont négociés à la bourse de Chicago. Les contrats à terme sur les produits des taux d'intérêt sont introduits par CBOT en 1975 reflétant la croissance énorme du marché au comptant. Ce type de produit dérivé a permis aux investisseurs de se couvrir contre le risque de fluctuation des taux d'intérêt et de spéculer sur leur direction future. À chaque moment, quatre dates d'expiration sont à la disposition de l'investisseur : mars, juin, septembre et décembre. Généralement, le contrat le plus près est le plus liquide. Pour cette raison, nous avons choisi le contrat de décembre pour estimer la valeur à risque intra journalière.

Avant de commencer à analyser ces données, il est crucial de comprendre au niveau intra journalier le système des opérations à la bourse de Chicago : pour le marché des contrats à terme sur le produits des taux d'intérêt, les activités de

[4] Cette compagnie commercialise des données transaction par transaction de plusieurs produits financiers tels que des contrats à terme, des options, des indices, des actions, des devises et plusieurs indicateurs.

transactions régulières commencent à 7:20 am et se terminent à 2 :00 pm. Toutefois, les intervenants sur le marché peuvent transiger au cours de la nuit.

La base de données intra journalière de la bourse de Chicago contient ainsi la date, le temps d'enregistrement, le prix de la transaction et le volume des contrats échangés.

Pour filtrer nos données, tout d'abord, nous avons éliminé les transactions non valides et les durées inter journalières. Ensuite, nous n'avons sélectionné que les transactions qui sont à l'intérieur de la session régulière. Enfin, comme nous avons parfois plusieurs transactions simultanées, nous avons calculé la moyenne pondérée de leur prix de la façon suivante :

$$moyenne_pondéré = \frac{\sum_{i=1}^{n} p_i V_i}{\sum_{i=1}^{n} V_i}$$

où n est le nombre de transactions simultanées, p_i et v_i sont le prix et le volume de la transaction i, respectivement. Nous avons également éliminé toutes les transactions simultanées sauf la dernière. Nous avons calculé par la suite les durées et les rendements transaction par transaction. Enfin, nous avons éliminé les valeurs extrêmes des durées et des rendements de telle sorte que les durées qui sont plus élevées à 25 écart-types ainsi que les rendements en valeur absolue qui sont plus élevées à 10 écart-types, sont éliminées. Nous avons obtenu en tout 269801 observations.

Selon le tableau III.1 de l'annexe III, nous remarquons qu'en moyenne la durée entre deux transactions est de 5 secondes. Nous constatons également que l'écart type

des durées est supérieur à la moyenne résultant d'un indice de dispersion[5] supérieur à un. En analysant la statistique du multiplicateur de LaGrange (LM1) d'ordre 15 sur les durées, nous remarquons que sa valeur est très élevée. Nous pouvons ainsi extraire les deux propriétés statistiques des durées qui sont démontrées empiriquement par plusieurs chercheurs. En effet, les durées sont fortement autocorrélées et elles sont, en plus, sur-dispersées. Donc, ces résultats suggèrent qu'une distribution exponentielle ne peut pas décrire les durées observées.

Nous constatons une moyenne des rendements égale à zéro et un écart-type égal à 0.0000725. Nous observons que le coefficient d'asymétrie est négatif alors que le coefficient d'aplatissement est supérieur à celui d'une loi de distribution normale (voir graphique III.1 dans l'annexe III pour les histogrammes des rendements et des durées).

Dans la prochaine section, nous allons montrer que les données à haute fréquence sont caractérisées par une saisonnalité intra journalière et inter journalière. Nous allons aussi démontrer la méthodologie appliquée afin d'ajuster les durées et les rendements pour ce type de problème.

2.2 Saisonnalité et ajustement des données à haute fréquence

Plusieurs études ont démontré que les données intra journalières sont caractérisées par une forme de saisonnalité. En effet, Engle et Russell (1998) ont montré que la courbe des durées sur le marché des actions a une forme de « U inversée» qui reflète une fréquence de transactions plus élevée au début et à la fin qu'au milieu de la journée. Selon Bauwens et Giot (2000), la saisonnalité intra journalière est définie en mesurant l'espérance des durées entre les transactions pour

[5] Dans la version de base des processus de Poisson ou exponentielle, par définition, la moyenne des durées est égale à son écart-type générant un indice de dispersion égal à un.

une journée particulière. Cette espérance est définie en calculant la moyenne des durées chaque 30 minutes pour chaque jour de la semaine. Ceci est équivalent à dire que dans chaque intervalle, la durée est constante. Toutefois, ceci n'est pas vrai puisque la durée change. Pour cette raison, nous avons supposé qu'il existe un changement graduel représenté par une fonction cubique. Le graphique III.2 dans l'annexe III montre que sur le marché des contrats à terme la saisonnalité est, plutôt, sous forme d'une tendance à la hausse au cours de la journée. Ce phénomène est expliqué par le fait qu'au début de la journée, les transactions sont très fréquentes puisque les nouvelles informations, tel que les annonces macroéconomiques, incitent les échangistes à prendre des positions. Toutefois, à la fin de la journée, les échangistes sont moins actifs sur le marché.

D'autres études ont montré que la saisonnalité intra journalière est observée également au niveau de la volatilité. Ainsi, pour examiner ce phénomène, nous avons appliqué la même méthodologie qu'à celle des durées, mais cette fois sur les rendements au carré (voir graphique III.3 dans l'annexe III). Ainsi, nous observons plutôt la forme d'une tendance décroissante de la volatilité durant la journée. Cette courbe explique que la volatilité est plus élevée au début qu'à la fin de la journée.

Il faut noter que l'analyse précédente a été appliquée sur des données ajustées à la saisonnalité inter journalière. Anatolyev et Shakin (2004) ont trouvé que les durées et les volatilités des rendements des titres, cotés à la bourse de la Russie, varient considérablement d'une journée à une autre. Nous pouvons constater que les durées sont plus élevées les vendredis que durant le reste de la semaine.

Pour ajuster les données des durées et des rendements à la saisonnalité inter et intra journalière, nous avons appliqué l'approche de Anatolyev et Shakin (2004)[6] définie de la façon suivante :

$$x_{t,\text{inter}} = \frac{x_t}{\overline{x}_s},$$
$$r_{t,\text{inter}} = \frac{r_t}{\sqrt{\overline{r}^2_s}},$$

où $\overline{x}_s$ correspond à la moyenne des durées pour le jour « *s* » si x_t est observé à l'intérieur de la journée « *s* ».

$$x_{t,\text{intra}} = \frac{x_{t,\text{inter}}}{E\left(x_{t,\text{inter}} \mid F_{t-1}\right)},$$
$$r_{t,\text{intra}} = \frac{r_{t,\text{inter}}}{\sqrt{E\left(r^2_{t,\text{inter}} \mid F_{t-1}\right)}},$$

et les espérances sont définies en calculant la moyenne des variables chaque 30 minutes pour chaque jour de la semaine et en estimant le changement graduel par une fonction cubique.

[6] D'autres méthodes d'ajustement des données à haute fréquence ont été proposées par d'autres chercheurs. Par exemple Bauwens et Giot (2000) et Veredas, Rodriguez-Poo end Espasa (2001) ont utilisé les méthodes non paramétriques basées sur des estimations « Kernel ».

2.3 Statistiques descriptives des données ajustées

Dans la section précédente, nous avons tenu compte de la saisonnalité inter et intra journalière. En effet, cet ajustement influence significativement les propriétés statistiques des données non ajustées. Puisque nous allons estimer notre modèle économétrique en utilisant des données désaisonnalisées. Dorénavant, nous allons utiliser plutôt les expressions courtes (les durées et les rendements) plutôt que les expressions suivantes : les durées et les rendements désaisonnalisées.

Selon le tableau III.2 de l'annexe III, les durées ainsi que les rendements possèdent des statistiques LM d'ordre 15 très élevées indiquant que ces séries sont fortement auto corrélées et possèdent de l'hétéroscédasticité. Comparées aux données non désaisonnalisées, ces statistiques ne sont pas très élevées ce qui suggère que l'ajustement des données diminue l'autocorrélation des séries initiales. Maintenant, si nous examinons les indices de dispersion (ratio écart type sur la moyenne) avant et après l'ajustement des durées, nous constatons que l'indice, après ajustement, a diminué de 1.71 à 1.59 donc nous nous approchons plus d'une distribution de Poisson où l'indice de dispersion est égal à un. Les graphiques III.4 nous montrent les séries temporelles des durées ajustés ainsi que des rendements ajustés et le graphique III.5 nous montrent l'histogramme des durées.

Dans le prochain chapitre, nous allons présenter la méthodologie appliquée pour estimer conjointement la durée et le rendement en utilisant la spécification présentée dans le premier chapitre. Nous allons aussi présenter les résultats de l'estimation et l'adéquation statistique de notre modèle. Enfin, quelques simulations, seront présentées pour mesurer la VaR intrajournalière.

CHAPITRE III

ESTIMATION DU MODÈLE ÉCONOMÉTRIQUE, SIMULATIONS ET PRÉVISIONS

3.1 Méthodologie d'estimation et interprétation des résultats

Nous estimons simultanément les paramètres du modèle WACD et du modèle ARMA(1,1)-UHF-GARCH(1,1) par la méthode de maximum de vraisemblance. Ainsi, nous utilisons les données intra journalière désaisonnalisées du premier mois pour l'estimation (du 1 septembre 2006 jusqu'au 29 septembre 2006 contenant 88950 observations) et nous utilisons les observations des deux autres mois (du 2 octobre 2006 jusqu'au 30 novembre 2006 contenant 180851 observations) pour la validation, la prévision et l'estimation de la valeur à risque. Ainsi, nous allons maximiser la fonction logarithmique de vraisemblance suivante :

$$L(x,h \mid \theta) = -\frac{1}{2}\left[(T-1)\log(2\pi) + \sum_{i=i_0+1}^{T}\log(x_i) + \sum_{i=i_0+1}^{T}\log(\sigma_i^2) + \sum_{i=i_0+1}^{T}\left(\frac{e_i}{\sqrt{x_i}\,\sigma_i}\right)^2\right] + \ldots$$

$$\sum_{i=i_0+1}^{T}\gamma\log\left[\Gamma\left(1+\frac{1}{\gamma}\right)\right] + \log\left(\frac{\gamma}{x_i}\right) + \gamma\log\left(\frac{x_i}{\psi_i}\right) - \left[\frac{\Gamma\left(1+\frac{1}{\gamma}\right)x_i}{\psi_i}\right]^{\gamma}$$

Pour trouver les valeurs initiales des paramètres, nous avons estimé séparément les deux modèles (WACD et ARMA-UHF-GARCH) et nous avons ensuite utilisé les paramètres trouvés pour estimer le modèle en entier. Enfin, nous avons appliqué les

tests d'hypothèse en utilisant une matrice variance covariance asymptotiquement efficiente et convergente.

Cette matrice est définie de la façon suivante :

$$V\left(\theta^{0}\right) = A^{-1} B \; A^{-1} \; ,$$

où

$$A = -\sum_{i=1}^{n} \frac{\partial^{2} L\left(x,h \mid \theta^{0}\right)}{\partial \theta^{0} \partial \theta^{0'}}$$

$$B = \sum_{i=1}^{n} \left[\frac{\partial L\left(x,h \mid \theta^{0}\right)}{\partial \theta^{0'}} \right] \left[\frac{\partial L\left(x,h \mid \theta^{0}\right)}{\partial \theta^{0'}} \right]'$$

Donc asymptotiquement :

$$\hat{\theta} \approx N\left(\theta^{0}, V\left(\theta^{0}\right)\right)$$

L'avantage de cette procédure est qu'elle nous permet d'appliquer les tests d'hypothèse standard et elle est souvent associée à la méthode de quasi maximum de vraisemblance. Les résultats de l'estimation sont présentés dans le tableau IV.1 de l'annexe IV. Nous constatons que le terme moyenne mobile MA(1) représenté par θ est négatif mais non significatif et que le terme autorégressif AR(1) représenté par φ est négatif et, également, non significatif. Nous remarquons aussi que la volatilité est très persistante avec un terme autorégressif représenté par le paramètre

statistiquement significatif $\widetilde{\beta}$. Les résultats montrent aussi que le paramètre v est statistiquement significatif et égal à 0.35.

3.2 Adéquation statistique du modèle

Pour tester la validité statistique du modèle *WACD(1,1)*, nous avons besoin de tester si les termes d'erreurs standardisés « ε_i » sont indépendamment et identiquement distribués ou *iid* avec une moyenne égale à 1. Selon le tableau I de l'annexe IV, le test d'auto corrélation du multiplicateur de Lagrange (LM1) nous permet de rejeter l'hypothèse nulle de l'indépendance temporelle dans les termes d'erreurs standardisés avec un niveau de significativité de 1%. De plus, la moyenne des « ε_i » est égale à 1.0082. Nous pouvons ainsi conclure que le modèle *WACD(1,1)* présente d'une façon adéquate la dynamique des durées.

Ensuite, nous avons testé la validité statistiques du modèle ARMA(1,1)-UHF-GARCH(1,1). Nous cherchons à voir si le résidu « e_i » n'est pas corrélé avec les rendements et les durées passés ainsi qu'avec la durée courante. Tout d'abord, nous avons examiné les statistiques des multiplicateurs de Lagrange LM1 et LM2, appliqué sur la série des termes d'erreurs standardisés et sur la série des termes d'erreurs standardisées au carré, respectivement (voir tableau IV.1 dans l'annexe IV). En effet, nous rejetons l'hypothèse nulle d'indépendance temporelle de ces termes d'erreurs ce qui signifie que les rendements demeurent autocorrélés. La présence d'un effet de la microstructure du marché peut être à l'origine de ce résultat. Ensuite, nous avons testé la dépendance temporelle des termes d'erreurs « e_i » avec la durée courante et les durées passées. Pour cela, nous avons régressé « e_i » sur « ε_i » et sur les durées retardées d'une jusqu'à trois périodes (voir tableau IV.2 dans l'annexe IV). Nous remarquons que tous les coefficients de la régression ne sont pas significatifs avec un niveau de confiance de 95%. Nous pouvons alors conclure que notre modèle

WACD(1,1)-ARMA(1,1)-UHF-GARCH(1,1) est statistiquement valide. Toutefois, l'élimination totale de la dépendance temporelle des termes d'erreurs *« e_i »* serait préférable pour une meilleure prévision hors échantillon.

Après avoir testé l'autocorrélation des termes d'erreurs, nous avons testé la présence de l'hétéroscédasticité. En effet, le test LM2 appliqué sur les résidus standardisés au carré (z_i au carré), montre que nous ne rejetons pas l'hypothèse nulle d'absence d'hétéroscédasticité des termes d'erreurs avec un niveau de significativité 5%. Donc, l'effet GARCH n'a pas été éliminé.

Enfin, reste à déterminer si les termes d'erreurs standardisés *« ε_i »* provenant du modèle WACD(1,1) suivent une distribution « Weibull » et si les termes d'erreurs standardisés *« z_i »,* provenant du modèle ARMA(1,1)-UHF-GARCH(1,1), suivent une distribution normale. Pour cela, nous avons illustré la densité empirique de *« ε_i »* et la densité d'une distribution « Weibull » en utilisant le paramètre estimé de notre modèle (voir graphique I dans l'annexe IV). Nous remarquons que la densité empirique n'est pas très proche à la densité d'une distribution « Weibull ». Selon le graphique II dans l'annexe IV, nous remarquons également que la densité empirique des *« z_i »* s'éloigne de la densité d'une loi normale à cause de la présence énorme des rendements zéros dans les données à haute fréquence. Pour contourner ce problème, Engle et Sun (2005) se débarrassent des transactions où il n'y pas de changement de prix, mais ils gaspillent beaucoup d'information.

Dans les prochaines sections, nous allons calibrer notre modèle et nous allons procéder à la simulation des rendements et des durées futurs afin d'estimer une valeur à risque intra journalière à partir de ces données.

3.3 Valeur à risque Intra journalière (IVaR)

Dans le milieu financier, une des questions cruciales consiste à déterminer, dans quelle mesure un actif financier, détenu par un négociateur très actif sur le marché, est exposé au risque. Pour cette raison, un grand besoin est apparu pour mesurer la valeur à risque à l'intérieur de la journée. Par exemple, supposons que vers 8h du matin, un gestionnaire de portefeuille très actif sur le marché boursier, veuille mesurer la valeur à risque de son portefeuille pour le reste de la journée. Donc une valeur à risque intra journalière (« IVaR ») sera très utile pour cette fin.

3.3.1 Définition de la Valeur à risque Intra journalière (Dionne, Duchesne et Pacurar, 2005)

Supposant que $Y = \{y_k, k \in Z\}$ est un ensemble de variables aléatoires qui correspond aux rendements réajustés pour des intervalles de temps réguliers dont le nombre est égal à T unité de temps. Nous considérons *n* réalisations de Y où $\{y_k;, k = 1,.....n\}$ tel que y_k est obtenu au temps $t_k^{'}$ et $t_k^{'} - t_{k-1}^{'} = T$. Ainsi, y_k est une mesure de rendement de *T* périodes de temps. Elle est simplement la somme des rendements transaction par transaction. Elle est calculée de la façon suivante :

$$y_k = \sum_{i=1}^{\tau(k)-1} r_i$$

où $\tau(k)$ est le cumul des durées qui excèdent *T* pour la première fois c'est-à-dire

$$\sum_{i=1}^{\tau-1} x_i \leq T \quad et \quad \sum_{i=1}^{\tau} x_i > T \quad .$$

En modélisant les durées, nous pourrons déterminer tout d'abord $\tau(k)$ et ensuite calculer la IVaR qui est définie de la façon suivante :

$$\Pr\left(y_k < -IVaR_k(\alpha) \mid \varsigma_k\right) = \alpha$$

où ς_k correspond à l'ensemble d'information qui inclut toutes les données transaction par transaction (durées et rendements) jusqu'au temps $t_k^{'}$. Dans notre étude empirique, nous allons utiliser des intervalles de confiances de 95% pour estimer l'IVaR calculée de la façon suivante :

$$-IVaR_k(\alpha) = Q_k\left(\alpha \mid \varsigma_k\right)$$

où $Q_k\left(\alpha \mid \varsigma_k\right)$ est un quantile de la distribution conditionnelle de y_k.

Dans notre étude empirique, nous allons calculer la IVaR en utilisant le modèle ARMA(1,1)-UHF-GARCH(1,1) présenté dans le chapitre I. En effet, en estimant ce modèle et en utilisant la technique de simulation Monte Carlo, nous allons générer des rendements futurs y_k régulièrement espacés dans le temps selon lesquels la IVaR est calculée.

3.3.2 Valeur à risque et simulation Monte Carlo intra journalière

DANS CETTE SECTION, NOUS ALLONS DÉCRIRE LA MÉTHODOLOGIE UTILISÉE POUR CALCULER LA IVAR EN UTILISANT DES DONNÉES TRANSACTION PAR TRANSACTION. PUISQUE CES DONNÉES SONT IRRÉGULIÈREMENT ESPACÉE DANS LE TEMPS, NOUS AURONS BESOIN PLUTÔT DES DONNÉES À INTERVALLE FIXE POUR CALCULER LA IVAR. POUR

CETTE RAISON, NOUS AVONS CALCULÉ LA SOMME DES DURÉES ET DES RENDEMENTS POUR AVOIR UN INTERVALLE DE TEMPS FIXE DE 45. CET INTERVALLE EST ÉQUIVALENT EN MOYENNE À UN INTERVALLE DE 22 MINUTES QUI NOUS PERMET D'AVOIR 721 INTERVALLES POUR UN TOTAL DE 44 JOURS (DU 2 OCTOBRE 2006 JUSQU'AU 30 NOVEMBRE 2006). POUR ESTIMER LA IVAR ET SELON L'APPROCHE DE DIONNE, DUCHESNE ET PACURAR (2005), NOUS PROPOSONS LA MÉTHODOLOGIE SUIVANTE:

- TOUT D'ABORD, NOUS AVONS GÉNÉRÉ DES VARIABLES ALÉATOIRES D'UNE LOI NORMALE STANDARD ET D'UNE LOI « WEIBULL » POUR CALCULER LES TERMES D'ERREURS DE NOTRE MODÈLE WACD(1,1)-ARMA(1,1)-UHF-GARCH(1,1) ET GÉNÉRER, AINSI, DES SCENARIOS DES DURÉES ET DES RENDEMENTS.
- EN UTILISANT LES PARAMÈTRES ESTIMÉS DE NOTRE MODÈLE POUR LA PÉRIODE DU 1 SEPTEMBRE 2006 JUSQU'AU 29 SEPTEMBRE 2006, NOUS AVONS OBTENU LES DURÉES, LES RENDEMENTS SIMULÉES AINSI QUE LES VOLATILITÉS[7].
- PAR LA SUITE, NOUS AVONS CALCULÉ LES DONNÉES À INTERVALLE DE TEMPS FIXE D'UNE FAÇON ITÉRATIVE ET À CONDITION QUE LA SOMME, DES DURÉES SIMULÉES, NE DÉPASSE PAS L'INTERVALLE DE TEMPS CHOISI (45 DANS NOTRE CAS).
- NOUS AVONS RÉPÉTÉ CETTE PROCÉDURE 5000 FOIS.

[7] Pour simuler les durées et les rendements futurs, nous avons supposé que les valeurs initiales sont les dernières observations des durées et des rendements pour la période du 1 septembre 2006 jusqu'au 29 septembre 2006.

- ENFIN, NOUS AVONS CALCULÉ, POUR CHAQUE INTERVALLE DE TEMPS FIXE, LA IVAR QUI CORRESPOND AU QUANTILE DE LA DISTRIBUTION DES RENDEMENTS FUTURS POUR CET INTERVALLE.

LE GRAPHIQUE V DANS L'ANNEXE V NOUS MONTRE LES RÉSULTATS DE LA PROCÉDURE DÉCRITE PLUS HAUT POUR CALCULER LA IVAR DE 22 MINUTES AVEC UN NIVEAU DE CONFIANCE DE 95% ET APPLIQUÉE SUR LE MARCHÉ DES CONTRATS À TERME SUR DES OBLIGATIONS GOUVERNEMENTALES. AINSI, NOUS REMARQUONS QUE SUR 700 INTERVALLES DE TEMPS DE 22 MINUTES, NOUS AVONS DÉPASSÉ LA IVAR SEULEMENT CINQ FOIS. TOUTEFOIS, POUR UN NIVEAU DE CONFIANCE DE 90%, NOUS AVONS DÉPASSÉ LA IVAR 12 FOIS. EN PROCÉDANT DE CETTE FAÇON, UN NÉGOCIATEUR SUR CE MARCHÉ POURRA MIEUX GÉRER LE RISQUE DE SA POSITION À L'INTÉRIEUR DE LA JOURNÉE.

3.3.3 Analyse de la performance du modèle hors échantillon

Dans cette section, nous testons la performance hors échantillon de notre estimation de la IVaR. Nous appliquons le test de Kupiec (1995) utilisé généralement dans la littérature pour évaluer la performance des modèles d'estimation de la VaR. Tout d'abord, nous calculons le taux d'échec empirique $\hat{\alpha}$ comme le pourcentage du temps que le taux de rendement actuel dépasse la IVaR estimée. Ainsi, si ce taux d'échec est statistiquement différent du niveau de significativité α alors notre estimation de la IVaR n'est pas efficiente. Pour appliquer un tel test, il faut calculer la statistique de ratio de vraisemblance de la façon suivante :

$$LR = 2\left[\ln\left(\hat{\alpha}^{m}\left(1-\hat{\alpha}\right)^{n-m}\right) - \ln\left(\alpha^{m}\left(1-\alpha\right)^{n-m}\right)\right]$$

où *m* est le nombre de fois que la IVaR a été dépassé et *n* est la taille de notre échantillon. Ce ratio de vraisemblance suit une distribution asymptotique de Khi-deux de degré de liberté 1. Ainsi, si on ne rejette pas l'hypothèse nulle alors notre estimation de la IVaR est robuste.

Dans l'annexe VI, Tableau VI, nous avons calculé la Pvalue de ce test pour chaque niveau de IVaR (5%, 2.5%, 1% et 0.5%) et pour différents intervalles de temps (15, 25, 35, 45 et 90). Nous remarquons pour un niveau de IVaR de 5%, nous rejetons l'hypothèse nulle avec un niveau de significativité de 1% et ceci pour tout les intervalles de temps. Toutefois, nous ne rejetons pas l'hypothèse nulle à partir d'un niveau de confiance de 97.5%. En plus, nous constatons que pour un niveau de IVaR de 0.5%, plus nous diminuons l'intervalle, plus la Pvalue diminue jusqu'à ce que nous rejetons l'hypothèse nulle pour un niveau de significativité de 5% et pour un intervalle de 15 (équivalent en moyenne à un intervalle de 7 minutes). Ce résultat ne peut être expliqué que par la présence énorme des zéros rendements qui empêchent le rendement de 7mn d'atteindre la valeur théorique de la IVaR.

CONCLUSION

Dans cette étude, nous avons proposé un modèle économétrique pour mesurer une valeur à risque intra journalière. En appliquant l'approche ACD-UHF-GARCH introduite par Engle(2000), nous avons modélisé la distribution jointe des rendements transaction par transaction et des durées et afin d'éliminer l'effet de la microstructure de marché, nous avons modélisé la dynamique des rendements en utilisant un processus ARMA. En effet, nous avons estimé notre modèle pour le marché des contrats à terme sur les obligations gouvernementales américaines négociées à la bourse de Chicago et nous avons par la suite testé la validé de notre modèle en appliquant différents tests d'adéquation statistique. Ensuite, nous avons mesuré une valeur à risque intra journalière en utilisant des techniques de simulation de Monte Carlo introduites par Dionne, Duchesne et Pacurar (2005). Enfin, nous avons appliqué un test de la performance hors échantillon de notre modèle.

Le modèle estimé est rejeté par des tests d'adéquation standard mais ceci est peut être relié au nombre impressionnant d'observations ainsi qu'à la complexité des données. Toutefois, l'ordre de grandeur des statistiques de tests indique cependant que le modèle capte une partie de la dépendance temporelle des données. D'autres

distributions des termes d'erreur et/ou plusieurs retards dans les termes de dépendance pourraient améliorer la performance.

En analysons la performance hors échantillon de notre modèle, nous avons constaté qu'il n'a pas été rejeté pour des niveaux de confiance supérieurs à 97.5% et pour des intervalles de temps supérieurs à 35. Ce résultat est peut être à l'origine de la présence énorme des zéros rendements. Ainsi, nous suggestions une extension de ce travail qui consiste à trouver une façon de résoudre ce problème et de tenir compte de la présence des valeurs extrêmes dans les données transaction par transaction. Ne pas tenir compte du risque de liquidité peut également être à l'origine de ce résultat. En effet, la modélisation de la dynamique du volume des transactions peut améliorer la performance de notre modèle. Finalement, une valeur à risque d'un portefeuille de titres sera de très grande utilité pour les gestionnaires de portefeuille.

BIBLIOGRAPHIE

Ai, C. et X., Chen, 2003, "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions ", *Econometrica*,71,1795-1844.

Anatolyev S. et D. Shakin , 2004, "Trade Intensity in an Emerging Stock Market: New Data and a New Model," *working paper*, New Economic School.

Bauwens, L. et P. Giot , 2000, "The Logarithmic ACD Model: an Application to the Bid-Ask Quote Process of Three NYSE Stocks," *Annales d'Économie et de Statistique*, 60, 117-149.

Dionne, G , Duchesne, P et Pacurar, M, 2005, " Intraday Value at Risk (IVaR) Using Tick-by-Tick Data with Application to the Toronto Stock Exchange ", J*ournal of Empirical Finance*, submitted.

Dolado, Rodriguez-Poo et Veredas (2004), " Testing Weak Exogeneity in the Exponential Family : an Application to Financial Point Processes", Core Discussion Paper, 49.

Dufour, A et R.F. Engle, 2000, "Time and The Impact of a Trade ", Journal of Finance, 55, 2467-2498.

Easley, D. et M. O'Hara, 1987, "Price, Trade, Size, and Information in Securities Markets" *Journal of Financial Economics*, 19, 69-90.

Engle, R.F. et J.R. Russell, 1998, "Autoregressive Conditional Duration: a New Model for Irregularly Spaced Transaction Data" *Econometrica*, 66, 1127-1162.

Engle, R.F. et J.R. Russell, 1997, "Forecasting the frequency of changes in quoted foreign exchange prices with autoregressive conditional duration model" *Journal of Empirical Finance,*,4, 187-212.

Engle, R.F. et Z.Sun , 2005, "Forecasting Volatility Using Tick By Tick Data", *Working paper*, Stern School of Business, New York University.

Engle, R.F. ,2000, "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68, 1-22.

Grenander, U, 1981, "Abstract Inference", Wiley, New York.

Grammig. J. et K,O. Maurer, 2000, " Non-monotonic Hazard Functions and the Autoregressive conditional Duration Model", *The Econometrics Journal,*3,16-38.

Holder, M.E, M.Qi et A.K Sinha, 2004, " The Impact of Time Duration between Trades on the Price of Treasury Note Futures Contracts", Journal of Futures Markets, 24, 965-980.

Jorion, P., 2000, " Value at Risk: The New Benchmark for Managing Financial Risk", *McGraw-Hill*, 544 pages.

Kupiec, P, 1995, "Techniques for Verifying the Accuracy of Risk Measurement Models", *Journal of Derivatives*, 2, 73-84.

Lunde, A. ,1999, "A Generalized Gamma Autoregressive Conditional Duration

Model," *discussion paper*, Aalborg University.

Meddahi, N., E. Renault et B. Werker, 2006, "GARCH and Irregularly Spaced Data," *Economics Letters*, 2 ,200-204.

Nelson, D.B. ,1991, "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

O'Hara, M., 1995, Market Microstructure Theory, Blackwell Publishers: Malden, 304 pages.

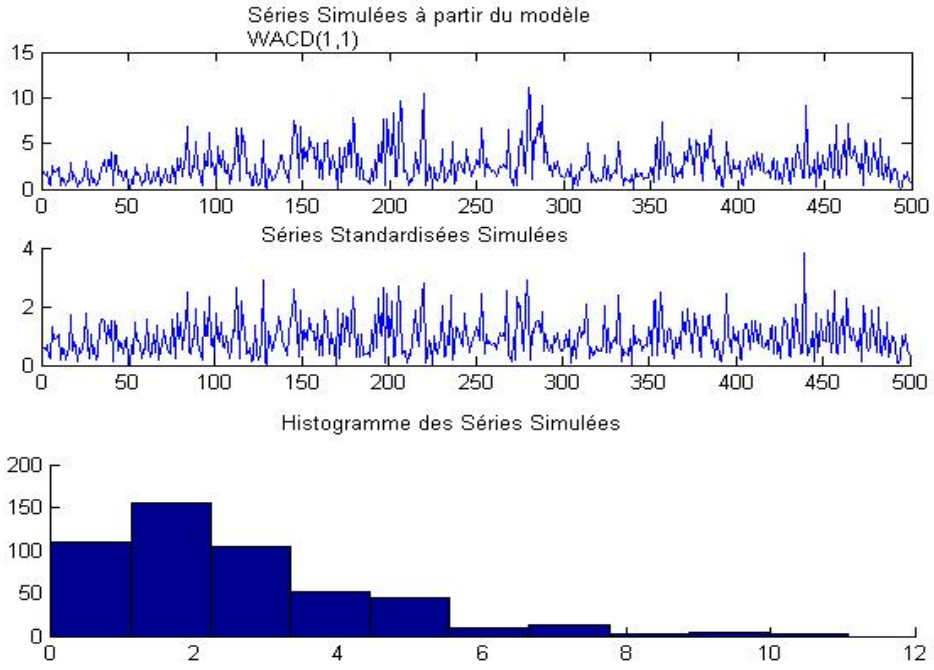
Tsay, R.S. ,2002, Analysis of Financial Time Series, Wiley: New York, 472 pages.

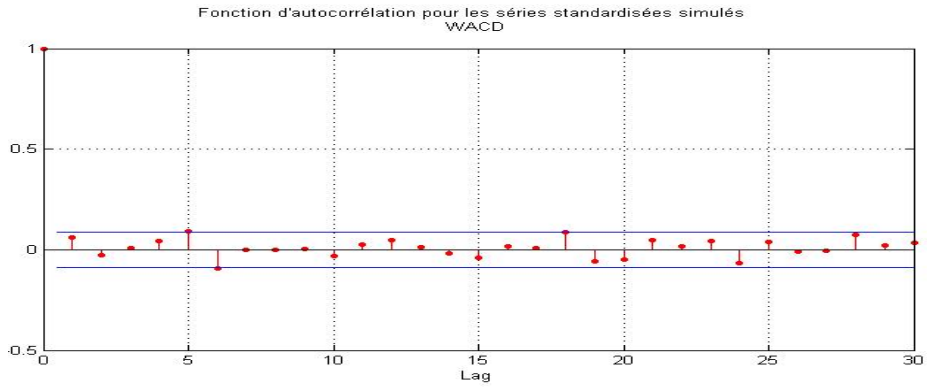
Roll, R. ,1984, "A simple implicit measure of the bid-ask spread in an efficient market", *Journal of Finance*, 39, 1127-1139.

Veredas, D.. Rodriguez-Poo, et A. Espasa (2001), "On the (Intradaily) Seasonality and Dynamics of Financial Point Process: A Semi parametric Approach", *Working Paper*, 2001-19, CREST.

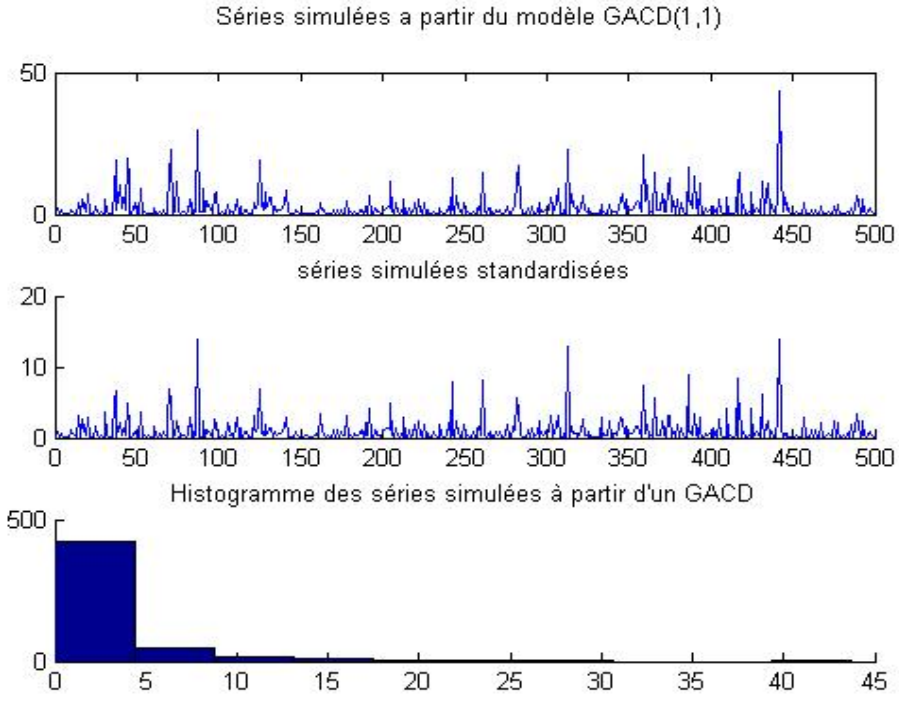
Annexe I

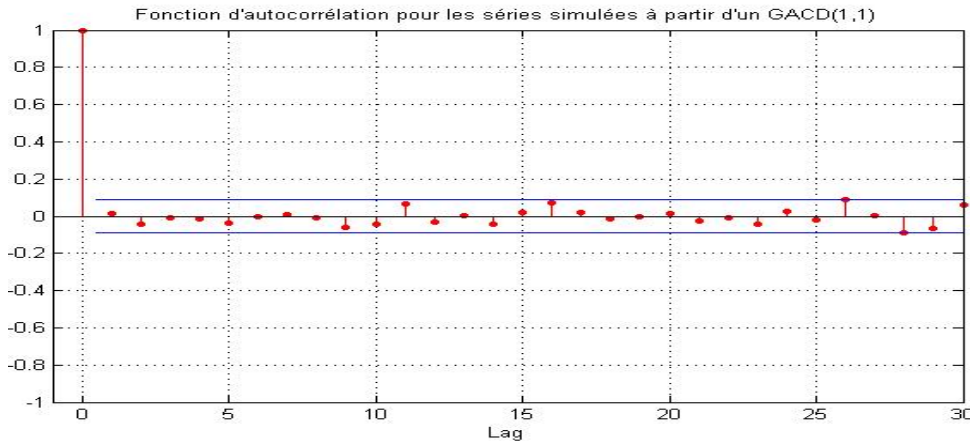
Graphique I.1: Simulation du modèle *WACD(1,1)*



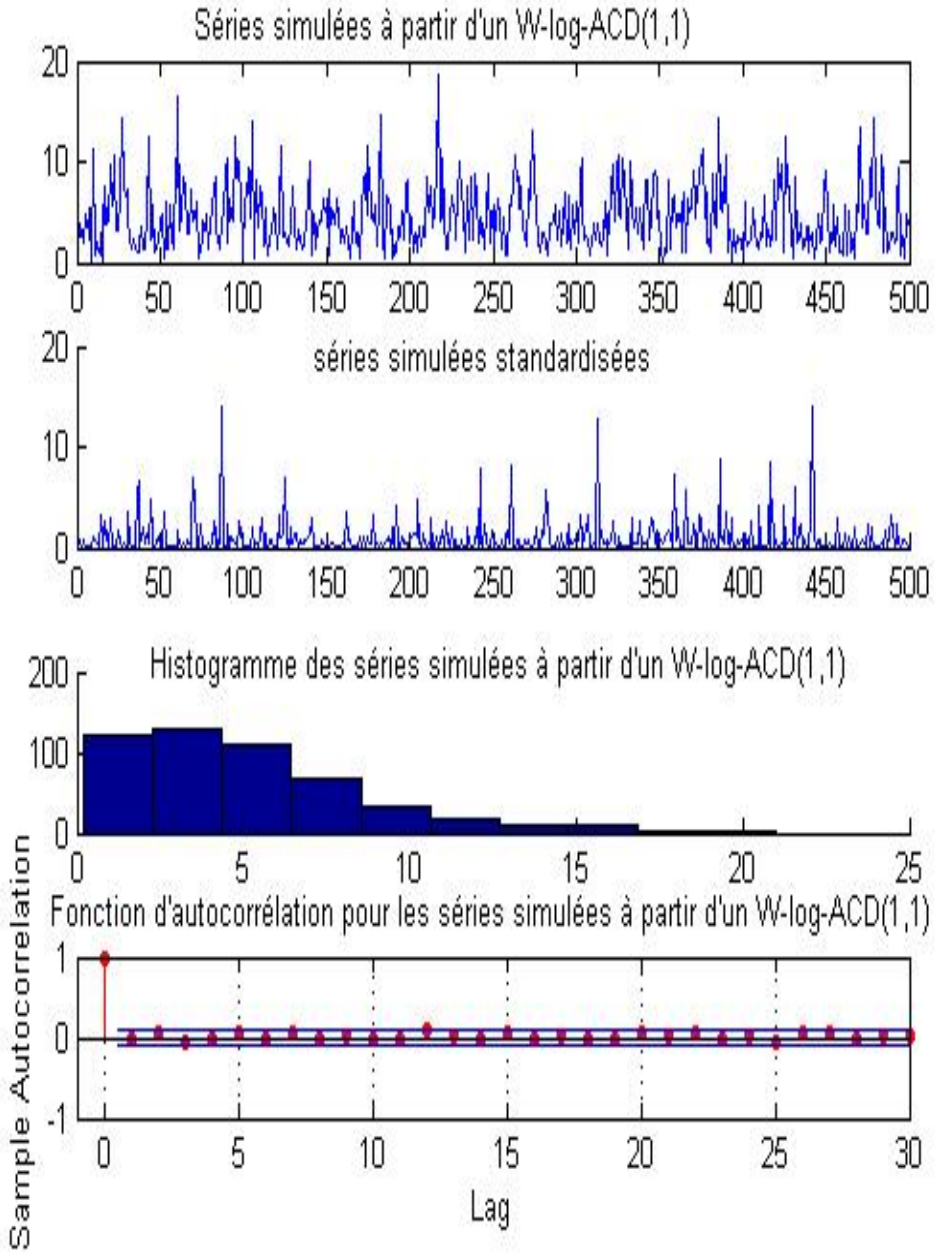


Graphique I.2 : Simulation du modèle *GACD(1,1)*





Graphique I.3 : Simulation du modèle *log-WACD(1,1)*



Graphique I.4 : Simulation du modèle *log-GACD(1,1)*

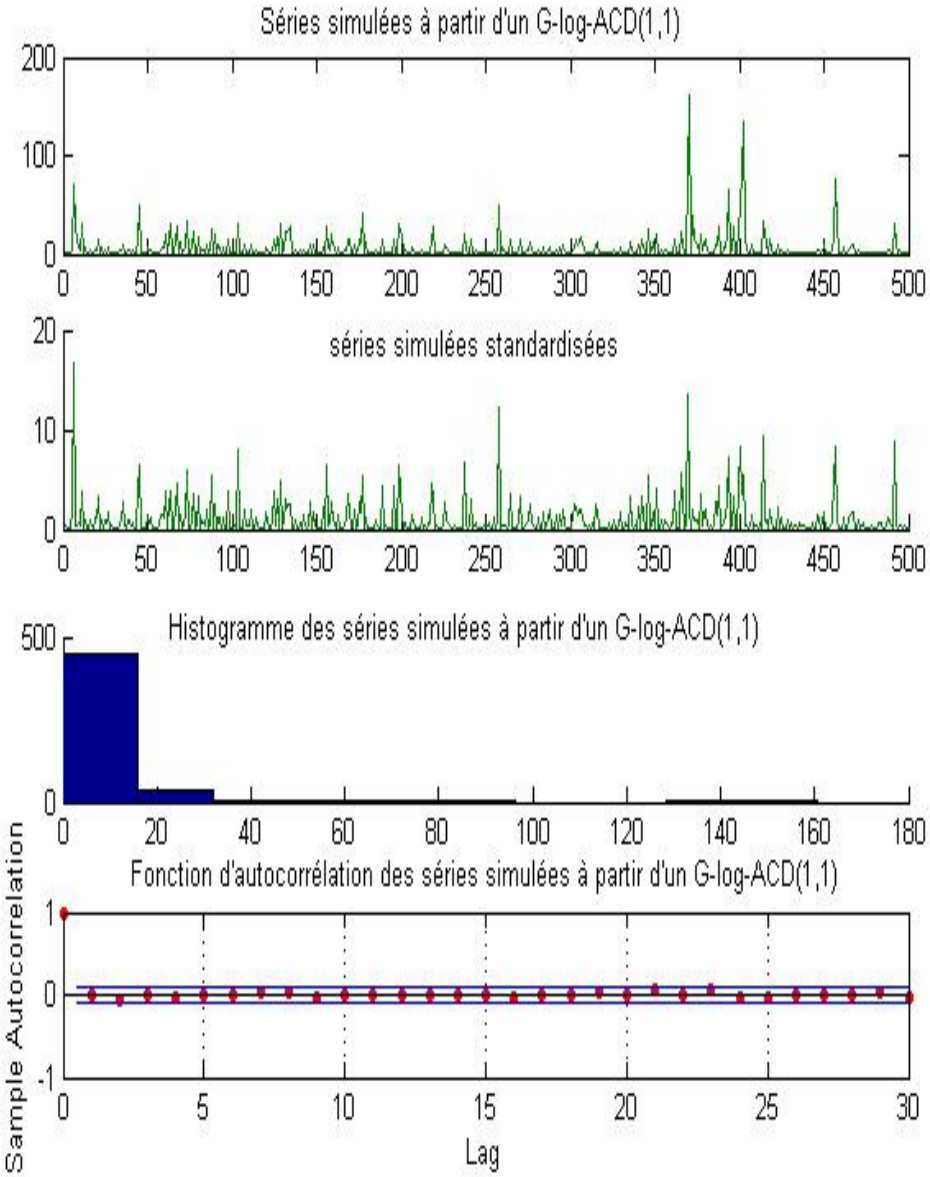


Tableau I : Résultat d'estimation du modèle WACD(1,1) simulé pour 5000 observations

WACD(1,1)				
Paramètres	w	α	β	γ
Vrai	0.3	0.2	0.7	1.5
Estimés	0.3496	0.1986	0.6523	1.5125
l'écart type	0.0420	0.0136	0.0271	0.0167

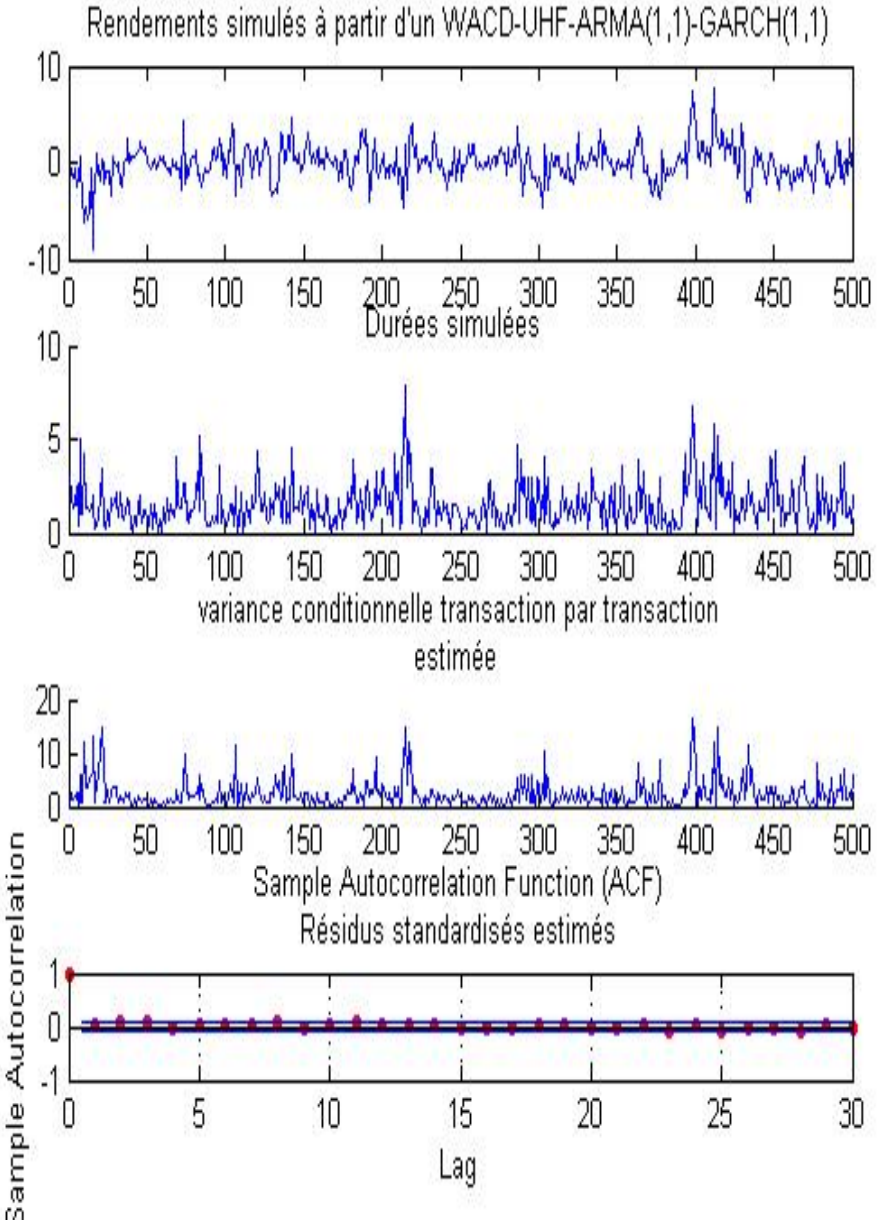
Annexe II

Tableau II : Résultats de l'estimation du modèle *UHF-ARMA(1,1)-GARCH(1,1)* en utilisant 5000 observations simulées 1000 fois à partir du modèle *WACD(1,1)-UHF-ARMA(1,1)-GARCH(1,1)*.

UHF-ARMA(1,1)-GARCH(1,1)							
Paramètres	$\widetilde{w}$	$\widetilde{\alpha}$	$\widetilde{\beta}$	v	c	$\varphi 1$	$\theta 1$
Vrai	0.2	0.2	0.7	1	-0.02	0.5	-0.02
Estimés	0.2878	0.1992	0.6977	1.09	-0.0224	0.4826	-0.0196
l'écart type	0.0282	0.0187	0.0259	0.0248	0.0042	0.011	0.0022

Graphique II: Rendements, durées et volatilité conditionnelle transaction par transaction simulés à partir d'un *WACD(1,1)-ARMA(1,1)-UHF-GARCH(1,1)* et

fonction d'autocorrélation totale pour les résidus standardisés du processus des rendements.



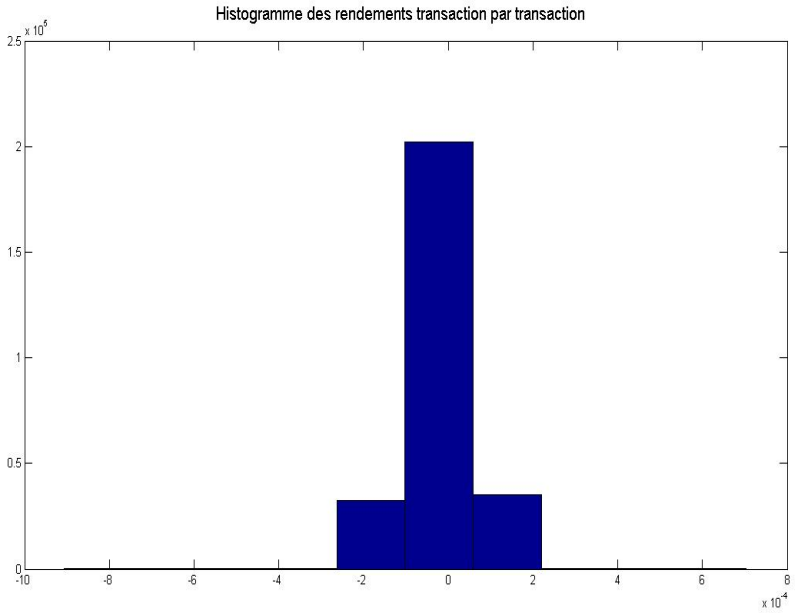
Annexe III

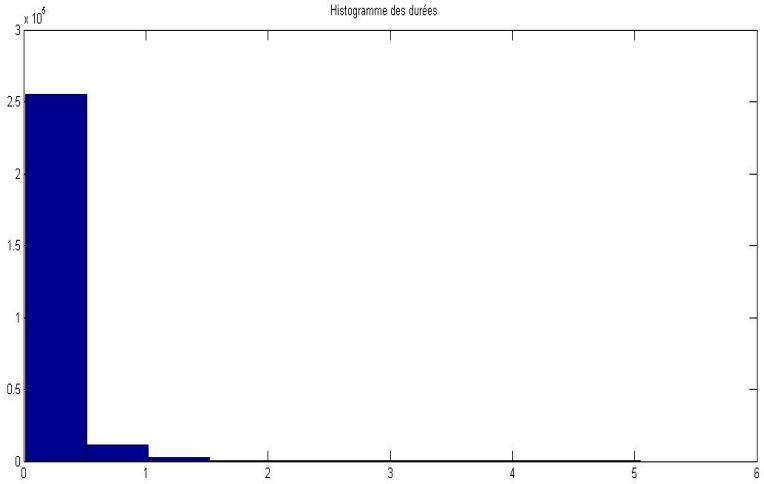
Tableau III.1 : Statistiques descriptives des données à haute fréquence brutes

Échantillon : Contrats à Terme 10Y US

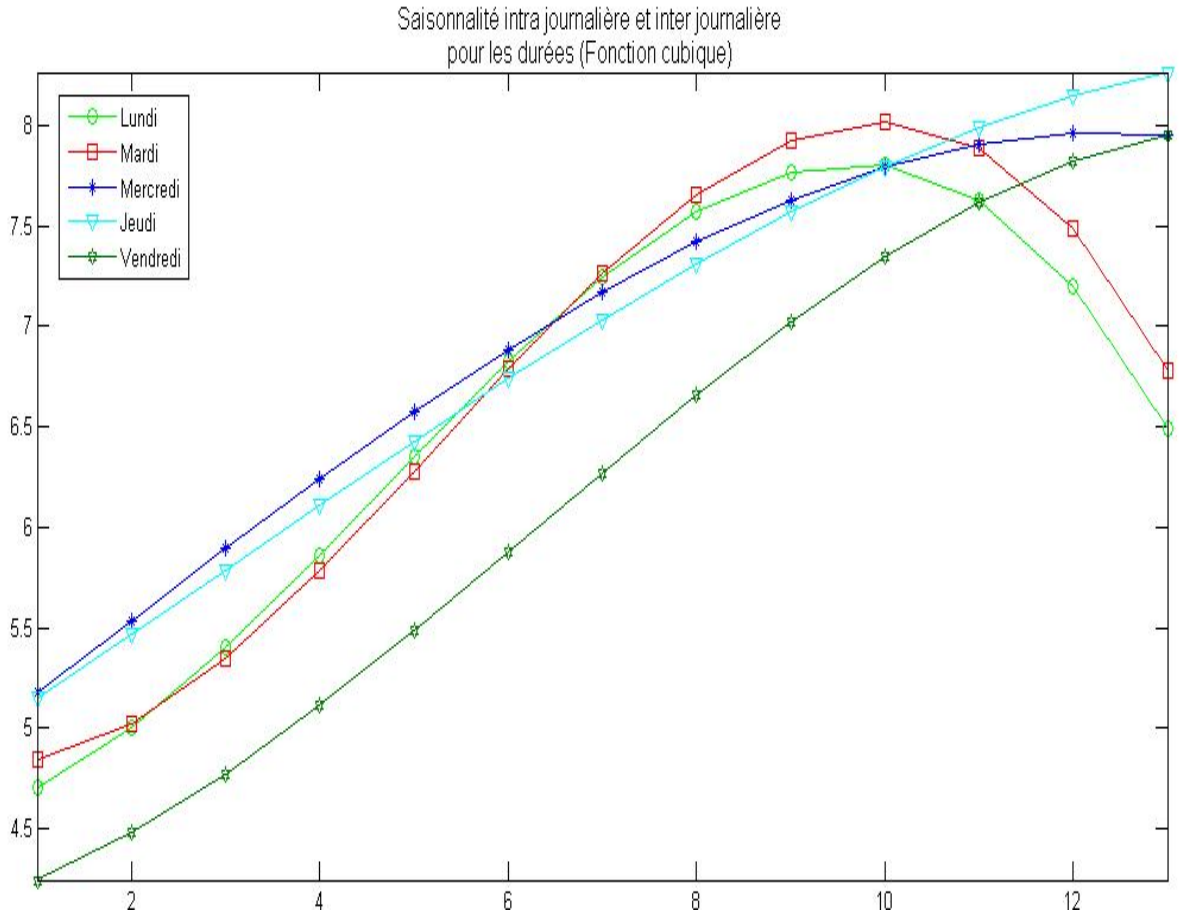
	Durée	Rendement
Moyenne	5.44	0.000000
Médiane	2	0.000000
Maximum	295	0.0007032
Minimum	1.000000	-0.0009071
Écart-type	9.31	0.0000725
Skewness	4.88	-0.00897
Kurtosis	45.832	4.70
Indice de dispersion	1.71	
LM 1(15)	27193	43838
LM2(15)	10125	15337

Graphique III.1 : Histogrammes des rendements intra journaliers et Histogrammes des durées intra journalières





Graphique III.2 : Facteurs intra journaliers estimés pour les durées



Graphique III.3 : Facteurs intra journaliers estimés pour les rendements au carré

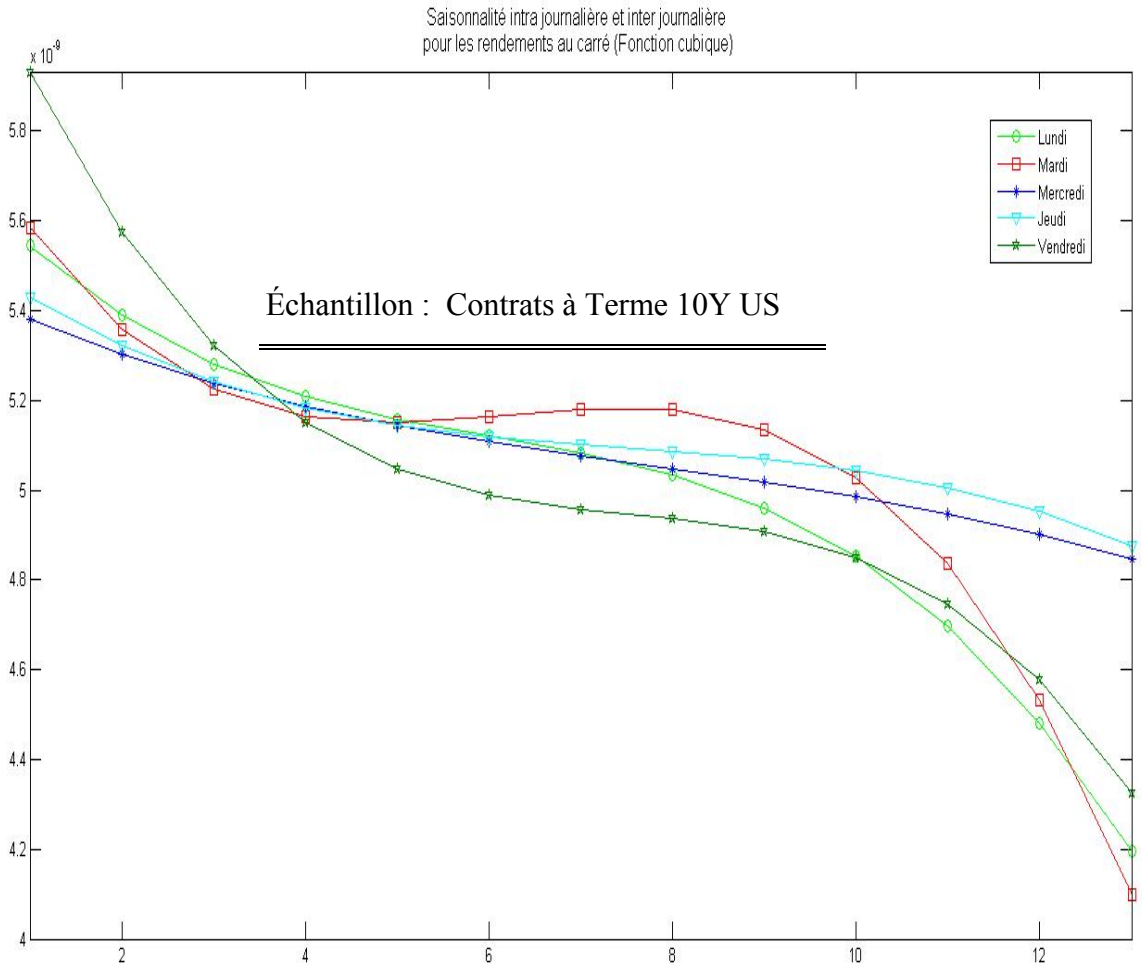
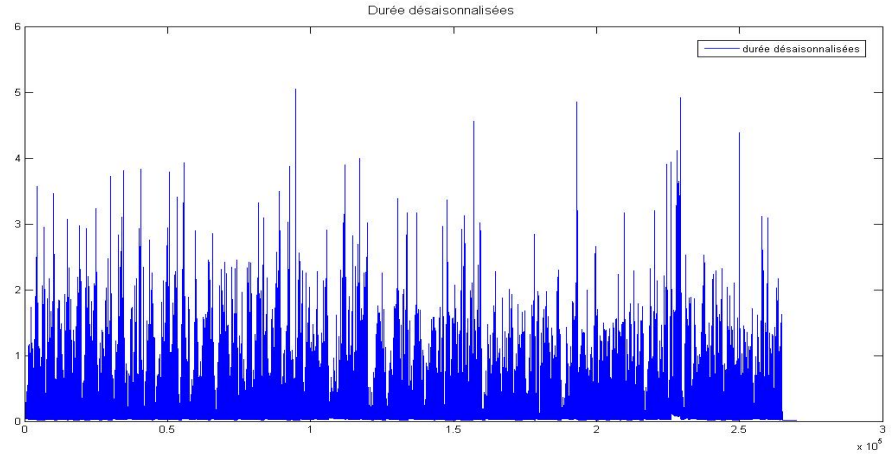
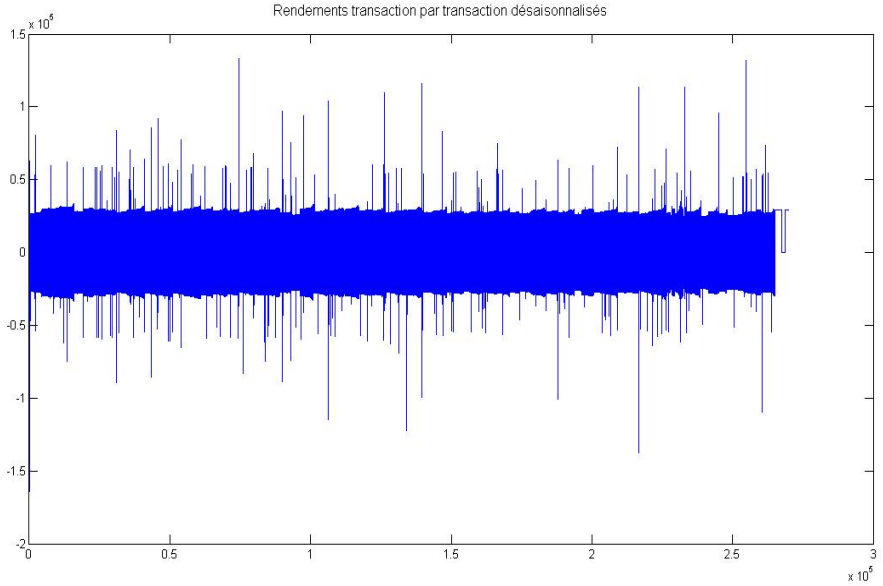


Tableau III.2 : Statistiques descriptives des données à haute fréquence désaisonnalisées

	Durée	Rendement
Moyenne	0.1377	418.6
Médiane	0.0517	0
Maximum	5.044	1.33 e +005
Minimum	0.0139	-1.639 e+005
Écart-type	0.2249	1.443 e+004
Skewness	4.496	0.0306
Kurtosis	36.5724	4.33761
Indice de dispersion	1.59	
LM1(15)	22140	44057
LM2(15)	7516	15033

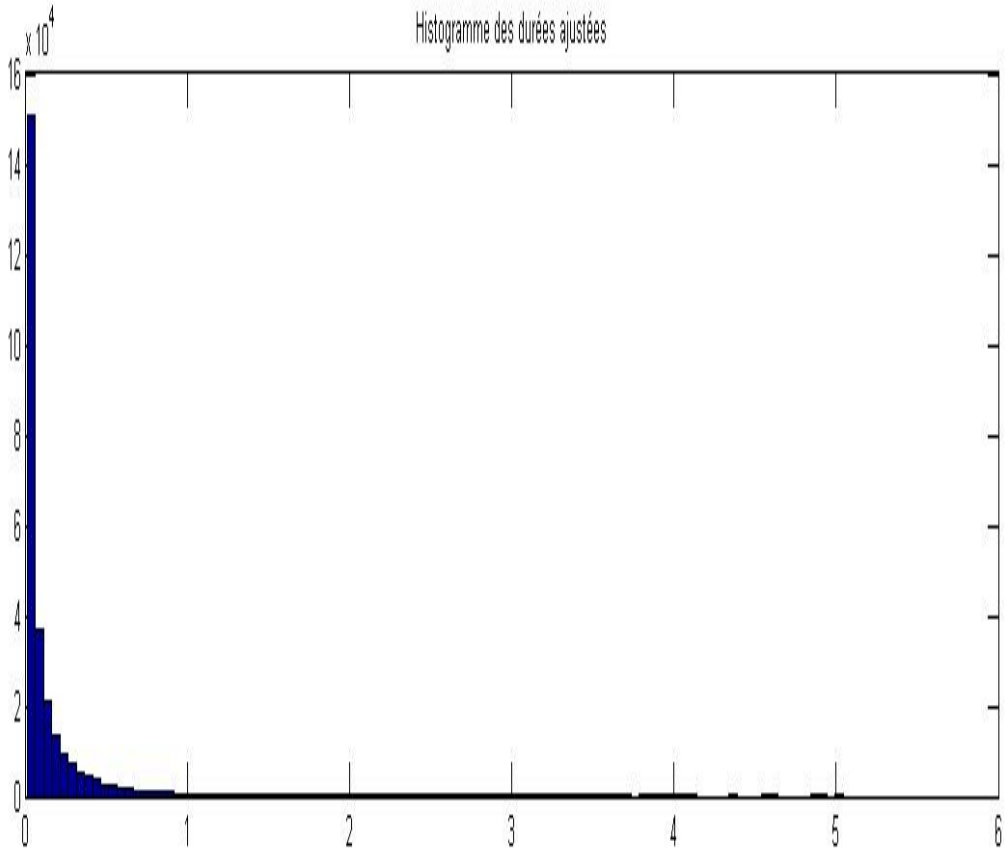
Graphique III.4 : Séries des données à haute fréquence désaisonnalisées





Graphique III.5 : Histogramme des durées ajustées.

,


Histogramme des durées ajustées
x 10^4
16
14
12
10
8
6
4
2
0
0
1
2
3
4
5
6

Annexe IV

Tableau IV.1 : Résultats de l'estimation du modèle *WACD(1,1)- ARMA(1,1)- UHF-GARCH(1,1)*

	Contrats à Terme 10Y US		
Paramètres	*Estimés*	*t-Student*	*P-value*
w	0.00938	33.30	0
α	0.214	117.2646	0
β	0.7256	245.0240	0
γ	0.9275	482.859	0
$\widetilde{w}$	1.8012 e+7	1.8011 e+7	0
$\widetilde{\alpha}$	0.070	107.4100	0
$\widetilde{\beta}$	0.8938	1693.4808	0
c	26.078	533.1637	0
φ	-0.20550	-0.2055	0.8371
θ	-0.020	-0.0200	0.9840
v	0.3532	969.9639	0

Test D'adéquation Statistique	Statistique de Kih-Deux	P-Value
p-value $LM1_d(15)$	262	0.00
p-value LM1(15)	14905	0.00
p-value LM2(15)	2475	0.00

Tableau IV.2 : Résultats de la régression des termes d'erreurs provenant du modèle ARMA(1,1)-UHF-GARCH(1,1) sur les termes d'erreurs standardisées provenant du modèle WACD(1,1), sur les trois retards de la durée.

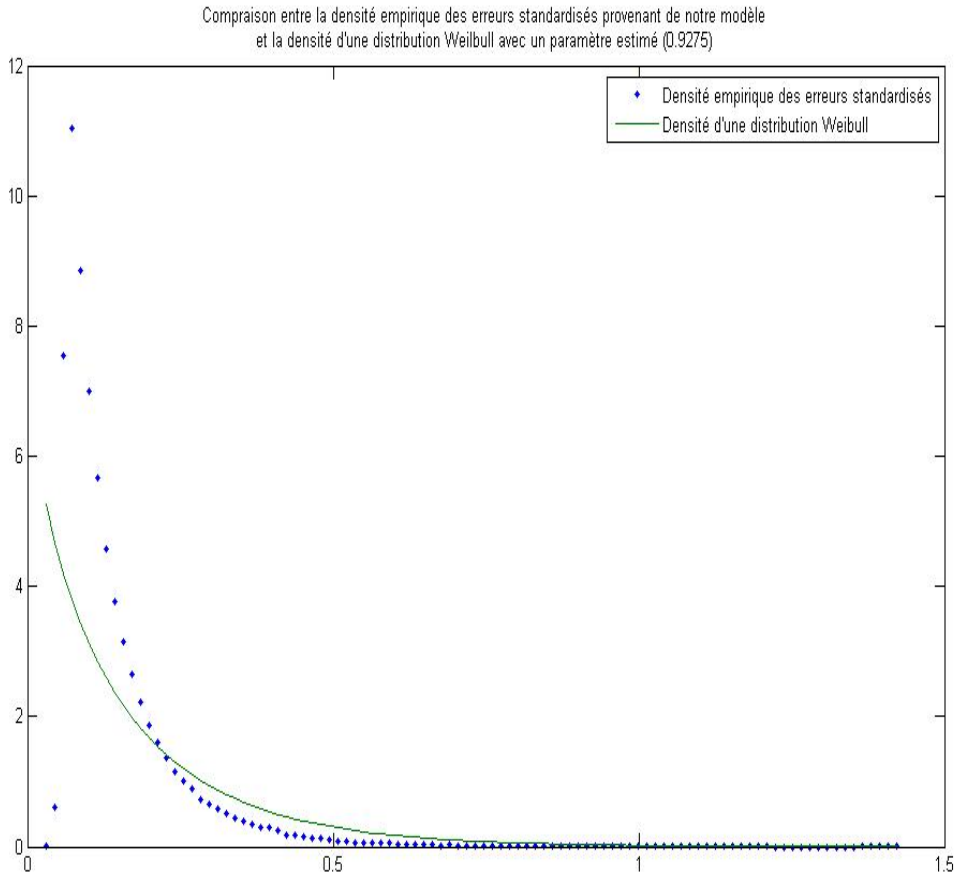
Variable Dépendante : « e_i »

Méthode: Moindre Carrée Ordinaire

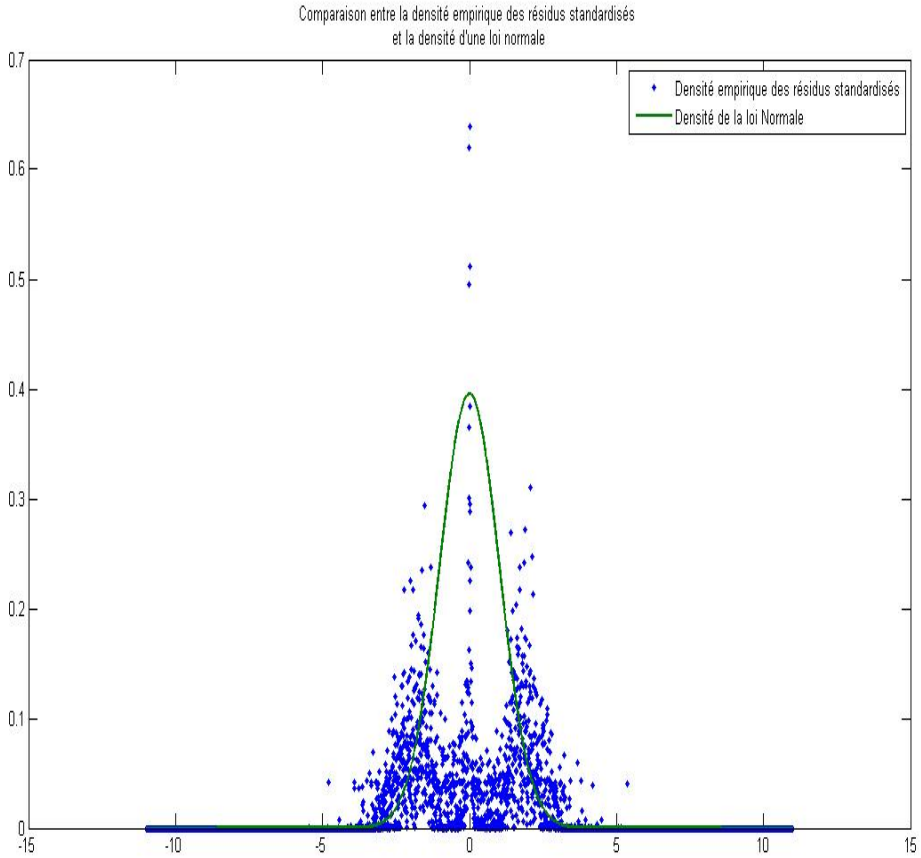
Observations: 88947

Variables	Coefficients	Std. Error	t-Statistic	P-Value
C	-35.74946	105.9618	-0.337381	0.7358
ε_i	-893.0562	1357.597	-0.657821	0.5107
X (-1)	751.4529	387.1330	1.941071	0.0522
X (-2)	158.9332	329.3502	0.482566	0.6294
X (-3)	181.6606	298.0795	0.609437	0.5422

Graphique IV.1 : Densité empirique des erreurs standardisés provenant du modèle WACD(1,1) versus la densité d'une distribution « Weibull ».

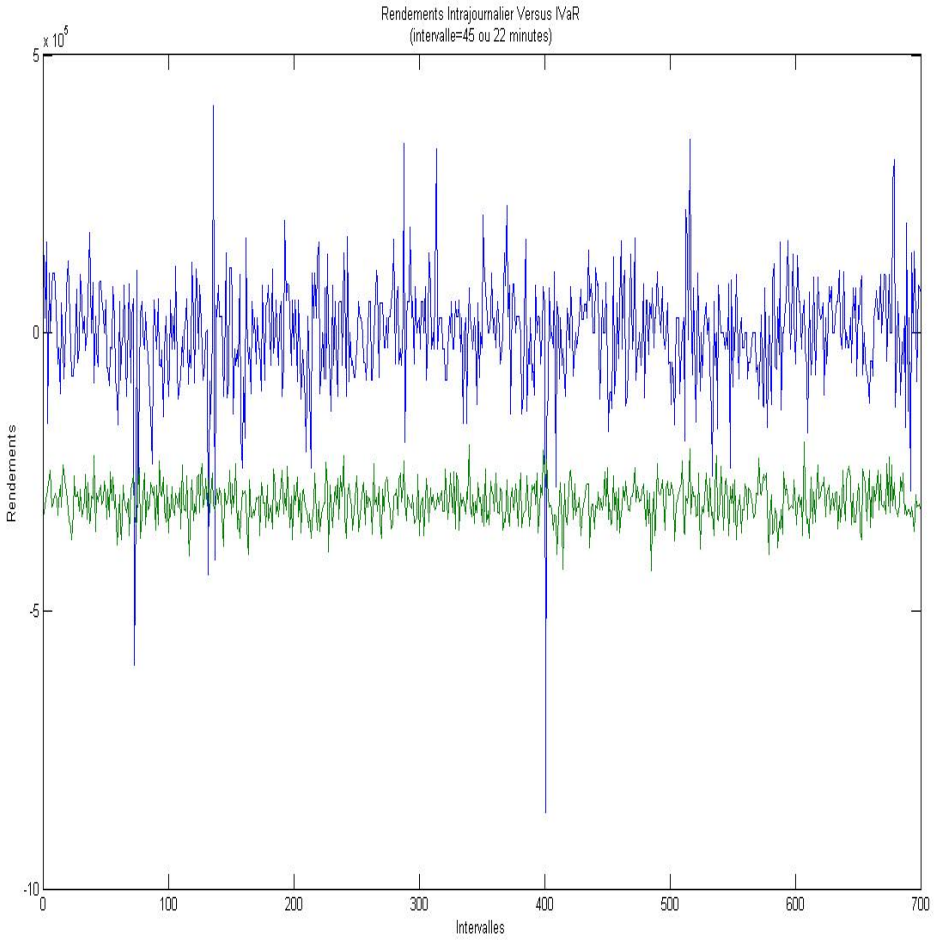


Graphique IV.2 : Densité empirique des erreurs standardisés provenant du modèle ARMA(1,1)-UHF-GARCH(1,1) versus la densité d'une distribution Normale.



Annexe V

Graphique V : Rendements Intra journaliers Versus IVaR (Intervalle =45 ou 22 minutes, niveau de confiance de 95%)



Annexe VI

Tableau VI : *P-value* du test de Kupiec

Niveau de la VaR / Intervalle de temps	5%	2.5%	1%	0.5%
15	0.000	0.1685	0.000	0.0432
25	0.000	0.0288	0.0255	0.2727
35	0.000	0.3323	0.0457	0.4021
45	0.000	0.000	0.0865	0.7836
90	0.000	0.0622	0.999	0.999