JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

ALI ALAMI
ÉRIC RENAULT

Risque de modèle de volatilité

Journal de la société française de statistique, tome 141, nº 1-2 (2000), p. 103-136
<http://www.numdam.org/item?id=JSFS_2000__141_1-2_103_0>

RISQUE DE MODÈLE DE VOLATILITÉ

Ali. ALAMI *, Eric. RENAULT †

RÉSUMÉ

L'arbitrage rendement – risque étant la substance de la finance, la volatilité a toujours été un paramètre essentiel pour la gestion de portefeuille. La généralisation de l'utilisation de produits dérivés a en outre mis sur le devant de la scène le concept de risque de volatilité, c'est-à-dire en quelque sorte le risque de modèle engendré par la vision de la volatilité comme un paramètre constant, alors que celle-ci est elle-même volatile. Ainsi, des mesures précises et des prévisions fiables de la volatilité sont demandées à l'économètre, non seulement pour l'évaluation et la couverture des actifs dérivés mais aussi plus généralement pour la gestion de portefeuille. La thèse centrale de cet article est que des stratégies opérationnelles de prévision statistique de la volatilité qui seraient « model-free » n'existent pas davantage que les opportunités d'arbitrage (« free lunch ») sur les marchés financiers. D'où le risque de modèle de volatilité contre lequel il est illusoire de vouloir s'immuniser. Plusieurs composantes spécifiques de ce risque de modèle sont analysées. On en déduira que le choix d'un modèle de volatilité pour une application financière donnée confronte toujours à un compromis rendement/risque sur le modèle lui-même.

Mots-clés — Volatilité, Risque de modèle, ARCH,GARCH, Prévision, Modèle structurel.

ABSTRACT

The risk-return trade-off being the very substance of finance, volatility has always been an essential parameter for portfolio management. Moreover, the generalization of the use of derivatives has placed in the forefront the concept of volatility risk : *i.e.* the model risk generated by treating the volatility as a constant parameter, when it is in fact volatile. Hence the econometrician is asked for accurate measures and reliable forecasts of volatility, not only for pricing and hedging derivatives, but also more generally for portfolio management. The central thesis of this paper is that operational model-free methods of volatility forecasting do not exist any more than do arbitrage opportunities (free lunches) in financial markets. It is for this reason that there exists volatility model risk against which it is illusory to try to immunize. Several specific components of this model risk are analyzed. One will imply that the choice of a volatility model for a given financial application always confronts one with a risk-return trade-off on the model itself.

Keywords and phrases — Volatility, Model Risk, ARCH, GARCH, Forecasting, Structural model.

*. INSEA ET BMCE Capital, Casablanca, Maroc

†. Université de Montréal, CRDE, CIRANO ET IFM2, Canada et CREST, Insee, France. e-mail : Eric.Renault@UMontreal.CA

INTRODUCTION

L'arbitrage rendement – risque étant la substance de la finance, la volatilité a toujours été un paramètre essentiel pour la gestion de portefeuille. La généralisation de l'utilisation de produits dérivés a en outre mis sur le devant de la scène le concept de **risque de volatilité**, c'est-à-dire en quelque sorte le risque de modèle généré par la vision de la volatilité comme un paramètre constant, alors que celle-ci est elle-même volatile. Ainsi, des mesures précises et des prévisions fiables de la volatilité sont demandées à l'économètre, non seulement pour l'évaluation et la couverture des actifs dérivés mais aussi plus généralement pour la gestion de portefeuille. La vision de la volatilité comme un processus stochastique nécessite en effet que sa valeur contemporaine soit prise en compte dans l'information conditionnante tant d'une performance de Sharpe conditionnelle que d'une Valeur à Risque (quantile d'une distribution conditionnelle).

La fiabilité des calculs du couple rendement/risque dans la loi de probabilité conditionnelle des rendements dépend donc directement du pouvoir prédictif du modèle de volatilité utilisé. D'où le **risque de modèle de volatilité**. La nature spécifique de ce risque vient du fait qu'il s'agit de modéliser un processus – « la volatilité » – qui, non seulement n'est pas directement observable, mais dont la définition même repose sur la spécification d'un modèle. Le risque de modèle de volatilité a donc deux composantes de natures très différentes :

(i) Il s'agit d'une part de contrôler la spécification de la modélisation par un processus stochastique qui n'est pas directement observable : on ne pourra jamais comparer directement la dynamique réelle de la volatilité à celle produite par le modèle. On se contentera soit de tester des conséquences du modèle telles qu'elles sont identifiables à partir des observations disponibles, soit d'examiner la conformité avec le modèle des valeurs de la volatilité que l'on a pu filtrer. Notons que ce filtrage repose lui-même sur un modèle, d'où un problème classique d'hypothèse jointe.

(ii) Mais on court aussi le risque plus radical de se tromper sur la définition de la quantité d'intérêt. Pour se figurer ce risque, il suffit de se référer à un travail des plus importants, récemment publié sur le risque de modèle de volatilité sous un titre pourtant résolument optimiste (voir Andersen et Bollerslev (1998)). Le problème du risque de modèle de volatilité y est présenté dans les termes suivants que nous traduisons in extenso pour y souligner l'absence d'une définition de la quantité d'intérêt (beaucoup d'exemples de ce type pourraient être trouvés dans la littérature récente !) :

« Ecrivons l'innovation du rendement [1] sous la forme $r_{t+1} = \sigma_t \cdot u_{t+1}$ où u_{t+1} désigne un processus de variables indépendantes d'espérance

1. Nous suivons dans cet article l'abus de langage courant consistant à traduire par « rendement » l'anglais « return », c'est-à-dire le gain en fin de période résultant d'un investissement unitaire en début de période.

nulle et de variance unitaire, tandis que la *volatilité* latente σ_t évolue conformément au modèle retenu. (...) Malheureusement, la volatilité n'est pas directement observée. (...) Cependant, si le modèle pour σ_t est correctement spécifié, alors :

$$E_t(r_{t+1})^2 = (\sigma_t)^2 E_t(u_{t+1})^2 = (\sigma_t)^2, \tag{1}$$

ce qui semble justifier l'utilisation du carré de l'innovation du processus de rendement sur l'horizon d'intérêt comme une approximation de la volatilité *ex post.* Mais, alors que l'innovation au carré fournit un estimateur sans biais du facteur de volatilité latent, elle peut constituer une mesure très imprécise du fait du terme d'erreur idiosyncratique $(u_{t+1})^2$. Cette composante présente généralement des variations aléatoires importantes relativement à $(\sigma_t)^2$, si bien que la part de la variation au carré du rendement attribuable à la volatilité est, en fait, faible».

Cependant, avant de discuter à partir de l'équation (1) les ordres de grandeur relatifs des variations des deux composantes inobservables $(\sigma_t)^2$ et $(u_{t+1})^2$, il conviendrait de réaliser que celles-ci peuvent, sans rien changer à la forme de la spécification (1), être remplacées par $(\sigma_t.z_t)^2$ et $(u_{t+1}/z_t)^2$ avec un large degré d'arbitraire sur le facteur d'échelle stochastique (z_t). En effet, la phrase «l'innovation au carré fournit un estimateur sans biais du facteur de volatilité latent» n'a aucun contenu identifiant (ou falsifiable) tant que l'on ne précise «l'information conditionnante» à la date t, c'est-à-dire la tribu $I(t)$ qui définit l'opérateur d'espérance conditionnelle E_t utilisé dans (1) :

$$E_t(r_{t+1})^2 = E[(r_{t+1})^2/I(t)].$$

Quand les économistes parlent d'information conditionnante (voir par exemple Hansen et Richard (1987)), ils ont en effet le plus souvent en tête une filtration croissante : $I(t) \subset I(t+1)$ (hypothèse dite de mémoire parfaite), qui traduit le fait que les agents du marché ont accumulé dans le passé un certain nombre d'observations (prix, variables de conjoncture macro-économique, etc.) qu'ils considèrent pertinentes pour leurs calculs d'anticipations rationnelles (espérances conditionnelles). Mais l'économètre ignore l'information ainsi utilisée par les agents du marché et le choix de la filtration utilisée pour la modélisation est largement arbitraire et contribue de manière significative au risque de modèle. Ainsi, si $J(t)$ est une structure d'information plus fine

$$(J(t) \supset I(t))$$

et :

$$E[(u_{t+1})^2/J(t)] = (z_t)^2 \quad \text{avec} \quad E[(z_t)^2/I(t)] = 1,$$

on aura à la fois :

$$E[(r_{t+1})^2/I(t)] = (\sigma_t)^2 \quad \text{et} \quad E[(r_{t+1})^2/J(t)] = (\sigma_t.z_t)^2.$$

Autrement dit, «l'innovation au carré fournit un estimateur sans biais du facteur de volatilité latent», que celui-ci soit $(\sigma_t)^2$ ou $(\sigma_t.z_t)^2$!

Bien entendu, ce paradoxe repose sur une différence d'information conditionnante. Mais il convient de souligner que, le plus souvent, rien d'évident ne nous dit quelle information conditionnante se cache derrière le sigle E_t. Le risque de se tromper sur la spécification de cette information est le premier des risques de modèle de volatilité, bien trop souvent négligé comme le montrent les deux exemples suivants :

(i) La réglementation prudentielle des banques, et sans doute bientôt de toutes les entreprises confrontées à des risques importants, leur dicte de calculer une Valeur à Risque qui, dans les modèles les plus simples, se fonde (*via* une hypothèse de normalité conditionnelle) sur un calcul de volatilité. Quelle volatilité : σ_t ou $(\sigma_t.z_t)$? Le régulateur veut-il que l'entreprise contrôle ses risques tels qu'ils sont calculables à partir de l'information $I(t)$ ou de l'information $J(t)$? Si l'on pense par exemple à des découvertes scientifiques *ex post* sur la nocivité de certains médicaments mis sur le marché, on réalise que rien n'est moins objectif ni plus risqué que la définition de l'information conditionnante pertinente ! Notons que ce problème est général pour la définition d'une Valeur à Risque, que celle-ci se calcule simplement à partir d'une volatilité ou plus généralement à partir de quantiles d'une loi de probabilité conditionnelle. Ainsi, l'efficacité d'une mesure de Valeur à Risque doit toujours être testée vis-à-vis d'une information conditionnante donnée (voir Christoffersen, Hahn et Inoue (2001)).

(ii) La droite de marché du CAPM de Sharpe-Lintner nous dit que la « valeur fondamentale » d'un actif (son prix d'équilibre) doit être telle que son rendement net espéré soit proportionnel à son bêta de marché. Ce coefficient bêta, ratio d'une covariance sur une variance (donc relatif à une notion multivariée de volatilité), doit être calculé conditionnellement à quelle information conditionnante ? La seule indication que nous donne la théorie économique à ce sujet est qu'il faudrait utiliser l'information conditionnante résultant de l'agrégation par le marché des informations individuelles des agents ! Il est clair que la traduction empirique d'une telle proposition laisse aussi beaucoup de place au risque de modèle dans l'évaluation de la performance d'un portefeuille.

En fait, si (1) ne définit absolument pas le processus de volatilité stochastique, la définition de la volatilité dans la citation ci-dessus de Andersen et Bollerslev (1998) se trouve cachée dans la phrase « u_{t+1} désigne un processus de variables indépendantes d'espérance nulle et de variance unitaire », c'est-à-dire un bruit blanc au sens strict (on dira aussi « bruit blanc fort »). Il est clair en effet que (u_{t+1}) et (u_{t+1}/z_t) ne peuvent pas en général être conjointement des bruits blancs forts. Cependant, cette condition, éventuellement identifiante, n'est pas non plus satisfaisante car elle interdit de définir la volatilité d'un processus dont la dynamique ne serait pas résumable par un facteur d'échelle. Nous montrerons en particulier que ce risque de modèle est réel.

L'article est organisé en deux grandes parties.

La première partie est consacrée aux modèles qualifiés de « forme réduite », parce que définissant la volatilité comme un processus observable (sous réserve de la connaissance d'un nombre fini de paramètres). Il s'agit essentiellement de la classe des modèles de type GARCH [2]. On discutera d'abord la tentation de limiter le risque de modèle en recourant à la notion de « GARCH faible », c'est-à-dire à une simple modélisation ARMA pour les rendements au carré. Puis on donnera quatre exemples de risques de modèle qu'il est cependant nécessaire de prendre : risque d'estimation statistique, risque de changement structurel, risque d'asymétrie et risque de variance infinie pour les rendements au carré. On souligne dans la deuxième partie l'importance d'une formulation structurelle des processus à volatilité stochastique [3]. Partant d'une simple formulation auto-régressive du processus de volatilité (processus SR-SARV de Andersen (1994) et Meddahi et Renault (1996, 2000)), on donne ensuite quatre exemples de risque de modèle qui ne sont pas couverts par une telle approche semi-paramétrique. Ces risques sont liés au calcul de prévisions de la volatilité et d'intervalles de prévisions sur celle-ci. En conclusion, on discutera l'utopie d'une observation de la volatilité et l'impossibilité corrélative de s'immuniser contre le risque de modèle de volatilité.

1. RISQUE D'UN MODÈLE DE VOLABILITÉ EN FORME RÉDUITE

Il s'agit dans cette section de montrer qu'il n'y a pas de définition « en forme réduite » de la volatilité qui soit à la fois opérationnelle et sans risque. Nous entendons par « définition en forme réduite » une définition qui, conformément à l'idée que « l'innovation au carré fournit un estimateur sans biais du facteur de volatilité latent », prétend définir la dynamique de la volatilité uniquement à travers celle du processus de rendement (r_t), c'est-à-dire par des caractérisations déterministes en fonction de (r_t), évitant la spécification d'une filtration d'information. La terminologie « en forme réduite », non standard dans ce cadre, est usuellement adoptée en Économétrie quand il s'agit d'opposer une formulation purement statistique sur les liens entre les variables économiques observées par l'économètre à une formulation dite « structurelle » parce qu'elle se centre sur les relations proposées par la théorie économique, ce qui peut en particulier conduire à introduire des composantes inobservables.

1.1. Une modélisation « faible » destinée à minimiser le risque de modèle.

Dans son introduction à un recueil des articles fondateurs sur les modèles ARCH, Engle (1995) explique avec humour que la conception de nouveaux modèles est aussi un exercice risqué par les choix de terminologie qu'il implique. Il raconte comment le travail au milieu des années quatre vingt de

2. Voir Bollerslev et Nelson (1994) ou Palm (1996) pour une synthèse sur les processus GARCH.
3. Voir Ghysels, Harvey et Renault (1996) ou Shephard 1996) pour une synthèse sur les processus à volatilité stochastique.

son étudiant Tim Bollerslev pour généraliser les processus ARCH (qui avaient été introduits par Engle (1982)) a failli le conduire à renommer MACH les processus ARCH, alors que les processus GARCH introduits par Bollerslev [4] (1986) auraient été nommés ARMACH par analogie avec l'appellation ARMA. L'écriture ARCH$(q)\sigma_{t-1}^2 = \omega + \sum_{\imath=1}^{q} \alpha_1 r_{t-1}^2$ fait en effet penser à une forme MA(q) alors que l'ajout de valeurs retardées de la variance conditionnelle évoquerait une forme ARMA. Puis Engle (1995) ajoute : « Heureusement, nous n'avons pas suivi cette voie (...) le modèle ARCH est en fait le modèle autorégressif, alors que le GARCH est l'ARMA ! Nous étions proches d'une erreur de terminologie très trompeuse ».

A l'appui de cette dernière affirmation, Engle (1995) développe l'exemple d'un processus GARCH$(1,1)$ dans un contexte où « ε_t est une variable sans corrélation sérielle, comme le rendement d'un actif financier, et h_t est la variance de ce rendement conditionnelle à l'information passée ». L'écriture GARCH$(1,1)$ classique :

$$h_t = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\, h_{t-1}, \quad t = 1, 2, \ldots, T \tag{2}$$

est alors réarrangée de manière équivalente en :

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta(\varepsilon_{t-1}^2 - h_{t-1}) + \varepsilon_t^2 - h_t, \quad t = 1, 2, \ldots, T \tag{3}$$

et Engle (1995) conclut que « la forme réécrite montre clairement qu'un GARCH$(1,1)$ est en réalité un ARMA(1,1) sur les rendements au carré. Les paramètres ARCH sont en réalité les paramètres autorégressifs et les paramètres GARCH contribuent à la fois aux structures autorégressive et moyenne mobile ». Notons que nous sommes restés fidèles ici à la notation classique pour les processus GARCH, même si on peut regretter qu'elle diffère de la notation maintenant en vigueur dans la modélisation à volatilité stochastique et que nous avions utilisée dans l'introduction (h_t correspond à σ_{t-1} car c'est une volatilité qui appartient à l'information conditionnante de la date $(t-1)$ alors que $\varepsilon_t = r_t$ n'est connu qu'en t).

Cette référence à la modélisation ARMA pouvant aisément être généralisée à tout modèle GARCH(p,q), on pourrait être tenté de limiter le risque de modèle de volatilité en affaiblissant la notion de processus GARCH pour n'en conserver que l'interprétation « linéaire » (au sens du théorème de représentation de Wold) de (3) : h_t serait alors interprétée comme une régression affine (et non plus comme une espérance conditionnelle) de ε_t^2 sur l'espace des ε_s^2, $s \leqslant t$. Ainsi, en se contentant d'appliquer le théorème de représentation de Wold au processus ε_t^2 supposé stationnaire au second ordre et régulier, on courrait un risque de modèle possiblement assez faible puisque il n'y a guère que la non-existence éventuelle des moments d'ordre 4

4. Taylor (1986) a indépendamment proposé la terminologie ARMACH pour désigner les processus GARCH. Mais il explique correctement que le modèle ARCH(q) correspond à la sous-classe des processus AR(q) pour les rendements au carré.

des rendements qui pourrait causer une erreur de spécification. En d'autres termes, l'innovation $\varepsilon_t^2 - h_t$ de (3) serait seulement supposée sans corrélation sérielle (ce que l'on conviendra d'appeler un **bruit blanc faible**) alors qu'une vision traditionnelle (c'est-à-dire conforme aux articles fondateurs de Engle (1982) et Bollerslev (1986)) des processus GARCH aurait maintenu sur ce processus d'innovation l'hypothèse beaucoup plus restrictive de **différence de martingale** (tout au moins par rapport à la filtration naturelle).

C'est pourquoi Drost et Nijman (1993) ont convenu d'appeler **GARCH faible** (« weak GARCH ») tout bruit blanc faible ε_t dont le carré est régi par une dynamique du type (3) (ou sa généralisation à tout ordre supérieur) :

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta\ \eta_{t-1} + \eta_t, \quad t = 1, 2, \ldots, T. \tag{4}$$

lorsque le processus d'innovation η_t de ε_t^2 est lui même seulement supposé être un bruit blanc faible. Par opposition, ils appellent **GARCH semi-fort** (« semi-strong GARCH ») le même processus ε_t lorsqu'il s'agit d'une différence de martingale conforme à (4) avec un processus d'innovation η_t du carré ε_t^2 qui est lui même une différence de martingale. Les processus GARCH semi-forts ainsi définis coïncident bien avec l'idée initiale de Engle (1982) et Bollerslev (1986) puisqu'il est clair réciproquement que si l'on suppose dans (4) que η_t est une différence de martingale, on en déduit que :

$$\eta_t = \varepsilon_t^2 - h_t$$

où $h_t = \omega + \alpha\ \varepsilon_{t-1}^2 + \beta\ h_{t-1}$ est bien la variance de ε_t conditionnelle à l'information passée.

Notons que l'on parlera de **GARCH fort** dans le cas d'un GARCH semi-fort tel que l'innovation standardisée $\eta_t = \varepsilon_t\sqrt{h_t}$ soit un bruit blanc fort (suite de variables indépendantes et de même loi).

En fait, la popularité des processus GARCH faibles dans la littérature récente, à la suite des articles fondateurs de Drost et Nijman (1993), Drost et Werker (1996) et Nijman et Sentana (1996), s'analyse sans aucun doute comme le résultat d'une prise de conscience d'un risque de modèle « GARCH semi-fort » d'autant plus manifeste qu'il est aisé de montrer que la classe des processus GARCH semi-forts n'est robuste vis-à-vis d'aucun type d'agrégation. Plus précisément, Drost et Nijman (1993) ont d'abord montré que si des rendements quotidiens sont conformes à un modèle GARCH semi-fort, les mêmes rendements considérés sur un horizon plus long (par exemple hebdomadaire) ne **peuvent pas** l'être. Autrement dit, la classe des processus GARCH semi-forts n'est pas robuste vis-à-vis de l'agrégation temporelle et c'est pourquoi Drost et Nijman (1993) ont proposé de l'étendre à tous les processus qu'ils appellent GARCH faibles pour récupérer cette robustesse. En effet, travailler sous l'hypothèse d'une structure de GARCH semi-fort revient à s'exposer à un risque de modèle dont l'occurrence est quasi-certaine puisqu'il est difficile d'imaginer une quelconque propriété de la (micro)structure des marchés garantissant une dynamique de type GARCH semi-fort précisément pour l'horizon des rendements que l'on considère (sachant que cette structure sera alors invalide à tout autre horizon!). Pour la même raison, l'échantillonnage en temps discret de données de prix d'actifs financiers générées par un processus de diffusion à volatilité stochastique ne pourra en général produire une dynamique de type GARCH semi-fort. Dans ce cas aussi, la structure moins restrictive de GARCH faible fournit une solution pour «réconcilier» («closing the GARCH gap» selon Drost et Werker (1996)) la modélisation financière en temps continu avec les données statistiques disponibles. Nijman et Sentana (1996) ont montré quand à eux que le rendement d'un portefeuille d'actifs individuels suivant tous un processus GARCH semi-fort n'est pas un GARCH semi-fort; la classe des processus GARCH semi-forts n'est pas non plus robuste à l'agrégation contemporaine alors que celle des GARCH faibles l'est.

De façon générale, dans la mesure où il n'existe aucune norme d'agrégation, ni temporelle ni contemporaine, qui s'impose à l'utilisateur, le concept de «GARCH faible» peut apparaître comme la panacée pour évacuer un risque de modèle trop patent. Nous prétendons cependant que ce type d'analyse, purement linéaire, de la dynamique de la volatilité, se heurte aussi à plusieurs risques de modèle.

1.2. Des risques de modèle inévitables.

Nous montrons que la généralité des processus GARCH faibles empêche d'assurer convenablement certains risques. En fait, en restant très général, on ne risque certes pas trop l'erreur de spécification mais on n'a pas de réponse aux questions pertinentes pour la finance appliquée.

1er Risque : *Le processus d'innovation η_t de la représentation ARMA faible des rendements au carré est susceptible de dynamiques non linéaires très marquées, qui rendent très imprécises les contreparties empiriques des corrélations linéaires :*

Après la mise en évidence de la correspondance ARMA/GARCH à travers l'écriture d'une «seconde forme» du modèle GARCH, Engle (1995) ajoute immédiatement : «Alors que la seconde forme peut être utilisée pour estimer le modèle avec des techniques de séries temporelles standards, il est en général très inefficace d'ignorer toute l'hétéroscédasticité dans les suites d'innovations, qui ont non seulement des variances conditionnelles variables mais aussi un support borné et variable dans le temps». Le risque, bien mis en exergue par cette remarque, est la tendance naturelle à se fier à des intuitions courantes qui sont valables pour des ARMA peu ou prou gaussiens (au moins conditionnellement à l'information passée) alors qu'un ARMA écrit pour un processus à valeurs positives (le rendement au carré) aura en général une dynamique très différente.

Une façon de prendre conscience de ce phénomène est d'imaginer le cas où le rendement serait, étant donné l'information passée, conditionnellement gaussien. Le taux de rendement ε_t étant en général calculé comme la variation logarithmique des prix, il s'agit du modèle de log-normalité conditionnelle qui est la référence en finance, que ce soit pour l'évaluation d'options (extensions dynamiques de Black et Scholes) ou l'estimation de processus GARCH ou à volatilité stochastique. Alors, quand la variance conditionnelle de ε_t est h_t celle de ε_t^2 (ou η_t) est $2h_t^2$! Autrement dit, la dynamique de l'hétéroscédasticité conditionnelle du rendement au carré est encore beaucoup plus marquée que celle du rendement lui-même ; et ce phénomène sera encore accentué si la loi conditionnelle de ε_t est leptokurtique et/ou asymétrique.

Pour mesurer à quel point ce risque peut être significatif, il suffit de considérer le cas simple de l'estimation du paramètre α de persistance de la variance dans un modèle ARCH(1), cas particulier du modèle (4) avec $\beta = 0$:

$$\varepsilon_t^2 = \omega + \alpha\, \varepsilon_{t-1}^2 + \eta_t, \quad t = 1, 2, \ldots, T. \tag{5}$$

quand, faute d'une spécification plus complète que celle d'un ARCH faible, on n'a pas d'autre outil de caractérisation de α que celui de la corrélation linéaire entre deux valeurs consécutives du rendement au carré (on négligera ici le fait qu'il pourrait être possible d'améliorer la précision asymptotique de l'estimation de α en tenant compte des auto corrélations d'ordre supérieur. Voir Broze, Francq et Zakoian (1999)). On reporte ci-dessous quelques-uns des résultats d'expériences de Monte-Carlo menées dans Alami (1999) ; ces résultats mettent en évidence des biais d'échantillon fini dans l'estimation de α par le coefficient de corrélation linéaire empirique α^* si dramatiques qu'il n'est même pas utile d'ajouter à ce constat une évaluation des variances confirmant pleinement le pronostic d'inefficacité formulé par Engle (1995) dans la remarque citée ci-dessus. Ces expériences de Monte-Carlo ont toutes été menées en considérant 1000 réplications d'une trajectoire de longueur T tirées dans un modèle conforme à (5) avec $\omega = 0,05$:

(i) Dans un premier groupe d'expériences, on tire les innovations standardisées $\varepsilon_t/\sqrt{h_t}$ comme un échantillon indépendant d'une loi normale centrée réduite. On obtient alors les moyennes de Monte-Carlo (et écarts-types entre parenthèses) de l'estimateur α^* pour des valeurs de α égales à 0,5 et 0,99 et une taille T de l'échantillon 100 et 200 conformes au Tableau 1 ci-dessous.

TABLEAU 1. — Innovations normales.

T \ α	0,5	0,99
100	0,317 (0,157)	0,434 (0,164)
200	0,358 (0,131)	0,470 (0,149)

(ii) Dans un second groupe d'expériences, on tire les innovations standardisées $\varepsilon_t/\sqrt{h_t}$ comme un échantillon indépendant d'une loi gamma $\gamma q_{(1)}$ convenablement translatée (pour obtenir une moyenne nulle) et

TABLEAU 2. — Innovations gamma.

T \ α	0,5	0,99
100	0,223 (0,154)	0,361 (0,187)
200	0,248 (0,128)	0,396 (0,166)

changée d'échelle (pour obtenir une variance unité). On espère ainsi rendre compte des conséquences à attendre d'une asymétrie et d'une leptokurticité de la distribution conditionnelle des rendements assez souvent constatée empiriquement. On obtient alors – voir tableau 2.

On conclut donc que α^* est un estimateur de α **inutilisable en pratique**, même dans le cas le plus favorable de rendements conditionnellement normaux (le constat est encore plus pessimiste dans l'éventualité plausible de leptokurticité ou d'asymétrie conditionnelles). Ceci nous paraît un **risque majeur dans l'utilisation du modèle GARCH faible** pour plusieurs raisons :

(i) Si on ne spécifie pas plus la dynamique que ne le fait le modèle GARCH faible, il y a peu d'espoir d'améliorer en échantillon fini les performances de l'estimateur des moments α^*, contrepartie empirique du coefficient de corrélation linéaire théorique α. En effet, l'amélioration proposée par Broze, Francq et Zakoian (1999) n'est qu'asymptotique et, en utilisant, comme ils le font aussi, les autocorrélations d'ordre supérieur, on ne fera que cumuler des biais d'échantillon fini similaires à celui documenté ci-dessus. Pour réaliser à quel point le comportement de ce coefficient de corrélation linéaire empirique est ici pathologique, il suffit de comparer la précision de l'estimateur α^* avec celle de l'estimateur du coefficient de corrélation linéaire que l'on obtient dans un modèle autorégressif d'ordre un gaussien :

$$y_t = c + \rho\, y_{t-1} + u_t,$$

avec $u_t \quad t = 1, 2, ..., T$ indépendants de même loi normale

On sait alors que la loi du coefficient de corrélation linéaire empirique ρ^* ne dépend ni de la moyenne ni de la variance de y_t et, par exemple avec $T = 100$ et une vraie valeur $\rho = 0,5$, on obtient un estimateur ρ^* qui vaut en moyenne 0,478. Il y a donc encore une sous-estimation mais sans commune mesure avec celle obtenue pour α dans le même contexte ($\alpha^* = 0{,}317$). Il est en outre important de remarquer que, même avec des innovations ARCH ($u_t = \varepsilon_t$ processus ARCH(1) conditionnellement gaussien avec $\alpha = 0,5$ et $\omega = 0,05$ comme ci-dessus), l'estimation de ρ se détériore assez peu ($\rho^* = 0{,}469$ en moyenne). Il n'est donc pas abusif de parler de comportement pathologique des innovations η_t du processus AR(1) ε_t^2 : celles-ci sont à l'évidence pourvues d'une dynamique beaucoup plus erratique que celle d'un ARCH !

(ii) La très mauvaise qualité de l'estimation de α décrite ci-dessus n'est pas une propriété intrinsèque de la modélisation ARCH mais une conséquence de la spécification « faible ». Par exemple (voir Alami (1999) pour une étude de Monte-Carlo plus systématique) le maximum de vraisemblance dans le cas conditionnellement gaussien avec $T = 100$ et une vraie valeur $\alpha = 0,5$ donne un estimateur α^{**} qui vaut en moyenne 0,464 ce qui est bien meilleur, d'une qualité comparable à l'estimation du coefficient autorégressif d'un processus AR(1) gaussien. Cela peut s'expliquer en remarquant que le maximum de vraisemblance (MLE) est équivalent à un estimateur de moindres carrés (dans l'équation de régression (5)) pondérés par une estimation convergente de la variance conditionnelle de ε_t^2 (ou η_t) donnée par :

$$2h_t^2 = 2[\omega + \alpha\ \varepsilon_{t_1}^2]^2$$

Cette interprétation en termes de moindres carrés pondérés explique que cet estimateur soit assez précis même dans le cas d'un ARCH semi-fort pour lequel la vraie loi conditionnelle n'est pas gaussienne (α^{**} est alors un estimateur du quasi ou pseudo maximum de vraisemblance). Il est par exemple vérifié dans Alami (1999) que α^{**} n'est en moyenne pas détérioré si on remplace la loi normale par une loi gamma (même s'il serait plus pertinent dans ce cas de reconsidérer la pondération des carrés en rapport avec la kurtosis de la loi gamma ; voir Meddahi et Renault (1997) pour une théorie générale de ces M-estimateurs). En revanche, si on sait seulement qu'il s'agit d'un ARCH faible, on renonce par là même à l'hypothèse assurant que le pseudo-score (correspondant à la méthode du pseudo-maximum de vraisemblance) soit une différence de martingale, et on n'a plus aucune raison d'espérer la convergence de l'estimateur PMLE α^{**}. Confrontés à ce dilemme, Drost et Nijman (1993) invoquent des expériences de Monte Carlo qui les ont rassurés sur une « relative » convergence de l'estimateur PMLE ; mais cela prouve simplement que les processus GARCH faibles qu'ils ont considérés pour leurs expériences n'étaient pas loin d'être GARCH semi-forts.

(iii) La critique radicale développée ci-dessus pourrait être atténuée en imaginant des tailles T d'échantillon très supérieures à la taille 100 privilégiée ici. En fait, les premières applications de la modélisation ARCH (voir Engle (1982)) concernent la prévision macro-économique où $T = 100$ est déjà une grande taille, et les premières études en échantillon fini de l'estimation de processus ARCH se sont focalisées sur le cas d'échantillons de quelques dizaines de points (voir par exemple Engle, Hendry et Trumble (1985)). Mais il n'est pas rare (en particulier avec le développement des bases de données haute fréquence) de disposer en finance de trajectoires de plusieurs milliers de points. Il serait cependant risqué d'oublier le problème d'échantillon fini, et ce pour au moins deux raisons. D'abord, si on veut utiliser des données très haute fréquence (intra-journalière), on se heurte à la question des effets de microstructure qui perturbent le processus d'intérêt. Ensuite, si on considère des observations sur une période longue (plusieurs mois), il y a un grand risque d'instabilité des paramètres d'un modèle ARCH. C'est le risque dit de changement structurel que nous analysons ci-dessous.

2ème Risque : *Il est peu plausible qu'une valeur donnée des paramètres ARMA décrive correctement la dynamique des rendements journaliers au carré sur une période très longue (plus d'une année). Des changements structurels (par exemple sauts déterministes dans la valeur de ces paramètres à certaines dates) sont au contraire à prévoir :*

Dans une étude récente, Mikosch et Starica (1999) vont jusqu'à affirmer : « une des conclusions évidentes de notre étude est l'**impossibilité de modéliser des séries longues** de rendements d'actifs financiers avec un seul modèle GARCH. Les paramètres du modèle doivent être mis à jour régulièrement (...) ». Leur analyse théorique confirme les résultats empiriques qu'avaient déjà mis en évidence Lamoureux et Lastrapes (1990). En considérant les rendements journaliers sur les actions de trente compagnies du 01 Janvier 1963 au 13 Novembre 1979 et en divisant arbitrairement la séquence des 4228 dates d'observations en 14 séquences consécutives de 302 jours chacune, Lamoureux et Lastrapes ont constaté que pour toutes les compagnies la variance inconditionnelle était significativement non constante d'une période à l'autre. En autorisant en conséquence les paramètres GARCH à varier d'une période à l'autre, ils ont observé qu'en moyenne l'estimation du coefficient de persistance $(\alpha + \beta)$ d'un GARCH$(1,1)$ était ramenée de 0,978 à 0,817. Autrement dit, la présence de changements structurels plausibles avec une périodicité au moins annuelle (penser par exemple aux changements de politique économique) conduit à sévèrement biaiser (dans le sens d'une surestimation) l'estimation de la persistance de la variance. Quand on sait que des valeurs de T de plusieurs milliers seraient en pratique nécessaires pour rendre « raisonnable » la procédure d'estimation linéaire décrite ci-dessus (on a vu par exemple que le passage de $T = 100$ à $T = 200$ n'améliorait que très peu les résultats), on conclut que le risque est réel. Mikosch et Starica (1999) proposent un modèle théorique pour analyser ce risque. Ils font une étude asymptotique en imaginant que l'échantillon des T dates d'observation peut être divisé en K sous échantillons en proportions fixes $q_i, i = 1, 2, \ldots, K$ (avec $\sum_{i=1}^{K} q_i = 1$) sur chacun desquels le processus de rendement est stationnaire ergodique. On appellera régime stationnaire numéro i celui correspondant aux observations numérotées dans le $i^{\text{ème}}$ groupe (en proportion q_i). Autrement dit, quand T tend vers l'infini (pour les besoins de l'étude asymptotique), les $[q_1\ T]$ premières observations (ou $[x]$ désigne la partie entière du nombre réel x) sont conformes à un certain régime stationnaire (par exemple GARCH$(1,1)$ avec une valeur $\theta_1 = (\omega_1, \alpha_1, \beta_1)$ des paramètres), les $[q_2\ T]$ sont conformes à une autre spécification stationnaire (par exemple encore GARCH$(1,1)$ mais avec une valeur θ_2) des paramètres, etc. Ils montrent que, dans ce cadre, l'autocovariance empirique des rendements au carré

$$\Gamma_T(h) = 1/T \sum {\varepsilon_t}^2 \varepsilon_{t+h}^2 - [1/T \sum {\varepsilon_t}^2]^2$$

converge (dans la mesure où le moment d'ordre 4 des rendements existe) **vers** :

$$\sum_i q_i\Gamma^i(h) + \sum_{i<j} q_i\ q_j[E^i - E^j]^2$$

où E^i et $\Gamma^i(h)$ désignent respectivement l'espérance et l'autocovariance théorique de ε_t^2 dans le régime stationnaire numéro i. On voit donc bien, comme anticipé par Lamoureux et Lastrapes (1990), que des valeurs E^i différentes selon le régime i, c'est-à-dire des changements déterministes dans la valeur de la variance inconditionnelle, peuvent augmenter arbitrairement l'autocovariance des rendements au carré. Afin d'illustrer le fait que des estimations de processus GARCH sur longues séries, négligeant le phénomène de changement structurel, peuvent être très trompeuses, Mikosch et Starica (1999) calculent la limite de l'estimateur de Whittle sur la représentation ARMA $(1,1)$:

$$\varepsilon_t^2 = \omega + (\alpha+\beta)\varepsilon_{t-1}^2 - \beta\eta_{t-1} + \eta_t,\ t = 1, 2, \ldots, T.$$

Dans le cas où le vrai processus est ARCH(1) dans chaque régime ($\beta_1 = \beta_2 = 0$ et cette contrainte est prise en compte dans le calcul de l'estimateur) et qu'il se produit un changement structurel dans les coefficients ARCH ($\alpha_1 \neq \alpha_2$ et $\omega_1 \neq \omega_2$), l'estimateur de Whittle de α a pour limite :

$$1 - [(1-\alpha_1)q_1\sigma_1^2 + (1-\alpha_2)q_2\sigma_2^2]/[q_1q_2(\sigma_1^2 - \sigma_2^2)^2 + q_1\sigma_1^2 + q_2\sigma_2^2]$$

où $\sigma_i^2(i = 1,2)$ désigne la variance inconditionnelle de ε_t^2 dans le régime numéro i.

On voit donc bien par exemple que, si seule la variance inconditionnelle $\omega/(1-\alpha)$ de ε_t^2 est soumise à des sauts à travers $\omega_1 \neq \omega_2$ (alors que $\alpha_1 = \alpha_2 = \alpha$), la variation de cette variance, qui se répercute sur les $\sigma_i^2(i = 1,2)$ va conduire l'estimateur de Whittle à surestimer α jusqu'à le croire arbitrairement proche de 1 quand on considère des valeurs de σ_1^2 et σ_2^2 arbitrairement éloignées. Dans le cas général, on peut montrer (ce qui confirme l'évidence empirique obtenue par Lamoureux et Lastrapes (1990)) que c'est l'estimateur du coefficient β de la partie MA qui aura une limite arbitrairement proche de 1 quand augmente l'amplitude du changement structurel entre σ_1^2 et σ_2^2. Ainsi, un tel changement donnera l'impression fallacieuse d'une très grande persistance $(\alpha+\beta)$ de la variance si on ne prend pas le soin d'estimer différents processus GARCH dans différents régimes.

3ème Risque : *La dynamique jointe des rendements d'actifs financiers et de leurs processus de volatilité présente couramment des **asymétries** importantes qui ne sont pas prises en compte par la modélisation GARCH faible :*

Les asymétries dans la dynamique de la volatilité sont souvent référencées dans la littérature sous le vocable «effet de levier» depuis que Black (1976) avait noté que les rendements d'actions sont négativement corrélés avec les variations de leur volatilité, au sens où la volatilité a tendance à augmenter en réponse à des « mauvaises nouvelles » (rendements nets moins élevés que prévu) et à chuter en réponse à des bonnes nouvelles (rendements nets plus élevés que prévu). Le levier financier de l'entreprise émettrice des actions peut en effet expliquer cette corrélation négative entre rendement d'aujourd'hui et volatilité de demain mais, quel que soit le signe de cette corrélation, celle-ci introduira un certain type d'asymétrie dans la distribution observée. On voit souvent reprocher aux modèles GARCH de ne pas capturer de telles asymétries, ce reproche étant fondé sur l'idée simple que, dans une dynamique GARCH telle que (2) (ou toute généralisation à un ordre supérieur), c'est seulement l'amplitude des chocs contemporains ε_t et non leur signe qui détermine la volatilité future h_{t+1}. On retient en particulier des modèles exponentiels GARCH (EGARCH) proposés par Nelson (1991) qu'ils résolvent ce problème en introduisant des effets de seuil par rapport à la valeur zéro de ε_t. Cependant, Nelson (1991) lui-même avait remarqué que c'est seulement dans le cas où la distribution de ε_t est conditionnellement symétrique que la modélisation GARCH implique que la valeur de la volatilité demain est sans corrélation avec le rendement aujourd'hui. Dans le cas général d'un ARCH(1) conforme à (2) (avec $\beta = 0$), la covariance conditionnelle vaut :

$$E_{t-1}[h_{t+1}\ \varepsilon_t] = \alpha E_{t-1}(\varepsilon_t^3)$$

Autrement dit, le seul reproche que l'on puisse faire aux modèles GARCH généraux est, non pas d'ignorer l'asymétrie, mais de confondre deux types d'asymétrie :

(i) D'un côté l'asymétrie « dynamique » que l'on conviendra d'appeler effet de levier (même si en toute rigueur cette terminologie devrait être réservée à la corrélation négative décrite ci-dessus) :

$$\mathrm{Cov}_{t-1}[\varepsilon_{t+1}^2, \varepsilon_t] = E_{t-1}(h_{t+1}\varepsilon_t) = 0 \tag{6}$$

(ii) D'un autre côté l'asymétrie « statique » que l'on conviendra d'appeler « effet de skewness » :

$$\mathrm{Cov}_{t-1}[\varepsilon_t^2, \varepsilon_t] = E_{t-1}(\varepsilon_t^3) = 0 \tag{7}$$

Les propriétés (6) et (7) sont effectivement équivalentes dans le cas GARCH alors qu'il est logiquement possible de les déconnecter dans un modèle à volatilité stochastique général (voir section 2). Il ne faut cependant pas exagérer la portée pratique d'un tel degré de liberté car il est clair que les deux types d'asymétries sont en fait très liés au moins pour les deux raisons suivantes (sans compter une certaine équivalence observationnelle au niveau du smile de volatilité [5] : voir Garcia, Luger et Renault (2000)) :

5. Renault et Touzi (1996) ont démontré que le caractère stochastique de la volatilité, quand il ne s'accompagne pas d'un phénomène d'asymétrie, produit des « smiles de volatilité » symétriques. La terminologie « smile de volatilité » réfère de façon générale à la courbe de la volatilité implicite dans un prix d'option (obtenue par inversion de la formule de Black et Scholes) en fonction du logarithme du prix d'exercice.

- D'abord, l'existence d'un effet de levier introduit automatiquement un effet de skewness par agrégation temporelle. Par exemple, si le taux de rendement ε_{t+1} sur la période $[t,\ t+1]$ est calculé comme la variation logarithmique du prix (ou de l'indice) sur la période, le taux de rendement sur la période « agrégée » $[t-1, t+1]$ sera calculé comme :

$$R_{t-1,t+1} = \varepsilon_t + \varepsilon_{t+1}$$

et donc l'effet de skewness à « basse » fréquence sera caractérisé par la non nullité de :

$$\mathrm{Cov}_{t-1}[(R_{t-1,t+1})^2, R_{t-1,t+1}]$$

non nullité qui peut bien être simplement une conséquence de celle de $\mathrm{Cov}_{t-1}\ [\varepsilon_{t+1}^2, \varepsilon_t]$.

- Ensuite, l'existence de primes de risque fonctions croissantes de la variance conditionnelle (voir modèles ARCH-M de Engle, Lilien et Robins (1987)) rend problématique l'identification séparée des différents effets d'asymétrie.

Quoiqu'il en soit, les modèles GARCH standards (c'est-à-dire semi-forts) sont capables de capturer les différentes asymétries fondamentales pour la mesure des risques. En effet, faute de capturer ces asymétries, on s'expose à un risque de modèle important que ce soit par exemple pour l'évaluation d'une option ou d'une Valeur à Risque qui ne portent que sur un côté de la distribution. Or, la modélisation GARCH faible renonce implicitement (cela est même explicite dans les articles fondateurs de Drost et Nijman (1993) et Drost et Werker (1996)) à capturer de telles asymétries puisque, avec une interprétation « linéaire » de h_t, on confond deux définitions possibles pour une approximation linéaire d'une variance conditionnelle :

- d'une part celle de h_t comme une régression affine de ε_t^2 sur l'espace des $\varepsilon_s^2, s \leqslant t$,
- d'autre part celle de h_t comme une régression affine de ε_t^2 sur l'espace des ε_s^2 et ε_s , $s \leqslant t$.

Les deux définitions ne coïncident précisément qu'à cause du maintien de l'hypothèse :

$$\mathrm{Cov}\ [\varepsilon_t^2, \varepsilon_s] = 0.$$

4ème Risque : *Pour les processus GARCH ε_t couramment estimés, $E(\varepsilon_t^4) = \infty$, si bien que l'auto-corrélation empirique des ε_t^2 n'est pas l'estimateur d'une auto-corrélation théorique :*

Nous avons déjà souligné (voir le premier risque) que le processus ARMA des ε_t^2 était susceptible de présenter des dynamiques non linéaires marquées le rendant très différent des processus ARMA approximativement gaussiens auxquels on est habitués. C'est particulièrement sensible dans le cas d'une persistance importante de la variance (comme il est courant d'en observer avec des données journalières) puisque l'on peut par exemple montrer (voir Bollerslev (1986)) que, dans le cas d'un processus GARCH (1, 1) conditionnellement gaussien, le moment d'ordre 4 ($E\varepsilon_t^4$) est infini dès que :

$$(\alpha + \beta)^2 + 2\alpha^2 > 1 \tag{8}$$

Autrement dit, pour les niveaux de persistance de la volatilité les plus couramment observés avec des séries de rendements d'actifs financiers calculés sur une base journalière, le moment d'ordre 4 du rendement est infini. C'est par exemple le cas pour un processus ARCH(1) dès que le coefficient de persistance sur deux périodes α^2 dépasse 1/3 ($\alpha < 0,5774$). Ce paradoxe est à mettre au débit de la modélisation en forme réduite (voir section 2). Même si on croit au modèle GARCH, il est difficile d'invoquer le théorème de Wold appliqué aux ε_t^2 pour justifier une modélisation faible puisqu'on est précisément dans un cas où le théorème de Wold ne s'applique pas : le processus d'intérêt est de variance infinie. Cela explique aussi le biais catastrophique observé dans l'estimation de α par le coefficient de corrélation linéaire empirique α^* des ε_t^2. Si α est supérieur à 0.5774, le coefficient de corrélation linéaire théorique n'existe pas et la limite du coefficient de corrélation linéaire empirique sera une variable aléatoire non dégénérée [6] (et non pas le paramètre α à estimer). On peut aussi montrer (voir Bollerslev (1986)) dans le cas d'un ARCH(1) conditionnellement gaussien que le moment d'ordre 8 ($E\varepsilon_t^8$) est infini dès que α^2 dépasse $[1/105]^{1/2}$ ($\alpha > 0,3124$). Alors, ε_t^2 ayant un moment d'ordre 4 infini, la théorie asymptotique standard du coefficient de corrélation linéaire empirique ne s'applique plus et α^* converge vers α à une vitesse inférieure à la racine carrée de la taille de l'échantillon (voir Davies et Mikosch (1998)), ce qui explique en partie les mauvaises propriétés en échantillon fini de α^* même pour une valeur de la persistance aussi réduite que $\alpha = 0,5$.

2. MODÉLISATION STRUCTURELLE DE LA VOLABILITÉ STOCHASTIQUE

On souligne dans cette section l'importance d'une formulation structurelle des processus à volatilité stochastique. L'ouvrage de Harvey (1989) a consacré la terminologie « modèle structurel de séries temporelles » dans le cas où le modèle dynamique est spécifié à travers des composantes inobservables qui centrent le travail de prévision et/ou d'analyse sur des composantes importantes pour leur interprétation économique (tendance, cycle, saisonnalité, composantes factorielles, etc.) qui sont identifiées, non pas directement comme fonctions des observables, mais comme caractéristiques de certaines lois conditionnelles d'intérêt. L'approche que nous proposons ici, suivant en cela Meddahi et Renault (2000), est précisément de ce type, en définissant la volatilité stochastique par la condition (1) assortie d'une donnée de la filtration d'information. Alors que la méthodologie proposée par Harvey (1989) s'appuie sur le filtre de Kalman pour traiter des modèles espace-état linéaires, la dynamique de la volatilité stochastique sera caractérisée par un modèle espace-état non linéaire qui peut être estimé par la méthode des moments généralisés de Hansen (1982) (voir aussi Renault (1997)).

6. Autrement dit, la loi des grands nombres ne s'applique pas et la limite des moments empiriques n'est pas un nombre fixe.

2.1. De la forme réduite aux équations structurelles.

Toute l'approche de la section 1 était basée sur l'idée qu'un modèle ARMA des innovations carrées des rendements rendait compte de l'hétéroscédasticité conditionnelle ; c'est en effet la philosophie fondamentale des modèles GARCH qui a présidé aux articles fondateurs de Engle (1982) et Bollerslev (1986). Cependant, si l'on n'oublie pas que la motivation de toute cette branche de la littérature est la prévision de la volatilité, fondée sur l'idée que celle-ci est hautement persistante (le phénomène dit de « volatility clustering »), il est naturel de partir d'une « équation structurelle », c'est-à-dire d'un modèle de la dynamique de la composante d'intérêt, à savoir la volatilité σ_t. Le modèle le plus simple pour rendre compte d'un phénomène de persistance [7] qui s'accompagne cependant d'un retour à la moyenne est un modèle autorégressif d'ordre 1 avec coefficient positif mais inférieur à un :

$$(\sigma_{t+1})^2 = \omega + \gamma(\sigma_t)^2 + \nu_{t+1}, \quad 0 < \gamma < 1, \quad E_t(\nu_{t+1}) = 0. \tag{9}$$

Pour estimer cette équation structurelle, il est évidemment nécessaire de revenir aux quantités « observables », à savoir les innovations carrées des rendements qui fournissent un « estimateur sans biais du facteur de volatilité latent » selon la citation de Andersen et Bollerslev (1998) rapportée plus haut :

$$(\varepsilon_{t+1})^2 = (\sigma_t)^2 + \eta_{t+1}, E_t(\eta_{t+1}) = 0. \tag{10}$$

L'appellation « estimateur sans biais » est cependant trompeuse puisque c'est plutôt la variable $I(t)$-mesurable $(\sigma_t)^2$ qui estime (ou est une anticipation de) la variable future $(\varepsilon_{t+1})^2$ alors qu'une partie de la variance attendue de celle-ci, égale à $\mathrm{Var}_t(\eta_{t+1})$, ne peut être capturée par son anticipation. Rappelons en outre que toutes les « anticipations » sont calculées relativement à la donnée d'une information conditionnante $I(t)$ qui est une filtration croissante par rapport à laquelle toutes les variables d'intérêt sont automatiquement supposées adaptées. Ainsi, $(\sigma_t)^2$ est une anticipation (qualifiée en économie de « rationnelle ») de $(\varepsilon_{t+1})^2$ parce que :

$$E_t[(\varepsilon_{t+1})^2] = E[(\varepsilon_{t+1})^2/I(t)] = (\sigma_t)^2.$$

Le passage de l'équation de régression structurelle (9) à une équation sur les variables observables est alors un problème très classique en économétrie, appelé problème « d'erreur sur les variables ». En effet, en intégrant dans (9) la relation (10) entre les observables et leurs anticipations, on obtient :

$$(\varepsilon_{t+1})^2 = \omega + \gamma(\varepsilon_t)^2 + \nu_t - \gamma\eta_t + \eta_{t+1} \tag{11}$$

7. La persistance de la volatilité, c'est-à-dire le phénomène de « volatility clustering » selon la terminologie anglosaxonne est un phénomène largement observé qui peut être modélisé soit par un processus auto-régressif à mémoire courte à racines proches de l'unité soit par un processus à différenciation fractionnaire (Comte et Renault (1998)).

Il s'agit bien d'une équation de régression mais elle montre que le coefficient de corrélation linéaire entre l'innovation carrée du rendement et sa valeur passée, qui serait la quantité d'intérêt s'il s'agissait seulement de prévoir les rendements au carré, donne en général une perception biaisée du paramètre γ de persistance de la variance puisque :

$$\mathrm{Cov}[(\varepsilon_t)^2, \nu_t - \gamma\eta_t] \neq 0. \tag{12}$$

C'est de façon générale la différence entre une forme réduite (juste utile pour la prévision) et une forme structurelle qui décrit la dynamique des composantes d'intérêt pour les applications économiques. Un exemple classique en économie depuis Friedman (1957) est celui de l'estimation d'une « fonction de consommation » vue comme une relation en première approximation linéaire entre la « consommation permanente » C_t^* et le « revenu permanent » Y_t^* , avec un terme d'erreur ν_t qui résume les effets autres que le « revenu permanent » pris en compte par le consommateur dans sa décision de consommation [8] . Ces effets ne biaisent pas la définition de la propension marginale à consommer γ dans la mesure où ils sont non corrélés avec le revenu permanent. La propension à consommer est ainsi définie par l'équation structurelle de régression suivante :

$$C_t^* = \omega + \gamma Y_t^* + \nu_t, \quad \mathrm{Cov}[Y_t^*, \nu_t] = 0.$$

Cette équation structurelle ne fournit pas directement une procédure d'estimation de γ car « consommation permanente » et « revenu permanent » sont des constructions théoriques dont consommation et revenu observés, notés respectivement C_t et Y_t ne sont que des mesures bruitées. Mais, les déviations transitoires étant par définition non corrélées avec les quantités dites permanentes, on aura :

$$C_t = C_t^* + \eta_{Ct} \quad \text{avec} \quad E[\eta_{Ct}/C_t^*] = 0,$$

$$Y_t = Y_t^* + \eta_{Yt} \quad \text{avec} \quad E[\eta_{Yt}/Y_t^*] = 0.$$

En reportant dans l'équation structurelle, on obtient :

$$C_t = \omega + \gamma Y_t + \eta_{Ct} - \gamma\eta_{Yt} + \nu_t.$$

Exactement comme dans le cas de (11), cette équation montre que la régression simple de C_t sur Y_t ne fournira pas un estimateur convergent du paramètre d'intérêt γ car, en général :

$$\mathrm{Cov}[Y_t, -\gamma\eta_{Yt} + \nu_t] \neq 0$$

8. Voir par exemple Romer (1996), p. 311, pour les fondements économiques et Gouriéroux et Montfort (1989), p. 321 pour le problème économétrique. Romer (1996) explique en particulier comment les quantités permanentes doivent être comprises comme des moyennes sur le cycle de vie, intégrant les anticipations sur les valeurs futures.

puisque par construction

$$\mathrm{Cov}[Y_t, -\gamma\eta_{Yt}] \neq 0.$$

Évidement, ce biais, dit de «simultanéité», peut s'annuler si :

$$-\gamma\eta_{Yt} = \nu_t.$$

Mais il n'y a aucun argument structurel pour justifier que la partie ν_t de la consommation permanente qui n'est pas prévisible comme fonction du revenu permanent coïncide exactement avec le produit de la propension marginale à consommer par l'erreur de mesure sur le revenu permanent.

On doit pourtant recourir à une hypothèse similaire pour obtenir un modèle ARCH semi- fort :

$$\mathrm{Cov}[(\varepsilon_t)^2, \nu_t - \gamma\eta_t] = 0 \quad \text{parce que} \quad \nu_t = \gamma\eta_t. \tag{13}$$

Cette dernière égalité est en effet nécessaire et suffisante pour que l'équation structurelle (9) se réduise à la forme réduite ARCH(1) :

$$(\sigma_{t+1})^2 = \omega + \gamma(\varepsilon_{t+1})^2, \quad 0 < \gamma < 1 \tag{14}$$

Si l'on se souvient de l'approche traditionnelle en économétrie pour le biais de simultanéité, il est pourtant possible d'estimer directement l'équation structurelle (9) sans recourir à une telle forme réduite. Le biais de simultanéité est traité de façon générale par l'utilisation de variables instrumentales convenablement choisies ; il s'agit en fait de trouver des variables qui, à la différence du régresseur $(\varepsilon_t)^2$ de (11), soient non corrélées au terme d'erreur $(\nu_t - \gamma\eta_t + \eta_{t+1})$ de cette équation. Il apparaît en particulier immédiatement que toute variable $I(t-1)$ mesurable convient. Autrement dit, alors que, du fait du problème d'erreur sur les variables, on ne peut pas espérer en général (sauf à maintenir le postulat (13) du modèle ARCH) :

$$E[(\varepsilon_{t+1})^2 - \omega - \gamma(\varepsilon_t)^2 / I(t)] = 0, \tag{15}$$

le modèle structurel implique en revanche :

$$E[(\varepsilon_{t+1})^2 - \omega - \gamma(\varepsilon_t)^2 / I(t_1)] = 0. \tag{16}$$

Cette approche en variables instrumentales, déjà bien connue dans la littérature macro-économétrique des modèles à anticipations rationnelles (voir, par exemple, Gourieroux et Monfort (1989), Exemple 9.46, p 326), a été développée dans le contexte des modèles de volatilité stochastique par Meddahi et Renault (2000). Il convient de remarquer que si, négligeant le biais de simultanéité, on utilise pour l'estimation par la «méthode des moments généralisés»[9] les

9. Voir Renault (1997) pour une synthèse sur les applications de la méthode des moments généralisés en finance.

conditions de moments (15) plutôt que (16), on sera le plus souvent conduit à une sous-estimation du paramètre d'intérêt γ. Ce phénomène, bien documenté dans le contexte des fonctions de consommation et des modèles à anticipations rationnelles, est mis en évidence par les expériences de Monte Carlo de Meddahi et Renault (2000). Cela s'explique par le fait que le biais de simultanéité est par construction du signe de l'espérance de la covariance conditionnelle

$$\mathrm{Cov}_{t-1}[(\varepsilon_t)^2, \nu_t - \gamma\eta_t] = \mathrm{Cov}_{t-1}[(\varepsilon_t)^2, \nu_t] - \gamma \mathrm{Var}_{t-1}[(\varepsilon_t)^2]$$

dont on s'attend à ce qu'elle soit le plus souvent négative. Pour préciser ce point, notons $\alpha_{t-1}[(\varepsilon_t)^2 - (\sigma_{t-1})^2]$ la régression affine, conditionnelle à $I(t-1)$, de ν_t sur $(\varepsilon_t)^2$:

$$\nu_t = \alpha_{t-1}[(\varepsilon_t)^2 - (\sigma_{t-1})^2] + \xi_t \text{ avec } \mathrm{Cov}_{t-1}[(\varepsilon_t)^2, \xi_t] = 0 \text{ et } E_{t-1}[\xi_t] = 0. \quad (17)$$

Le terme d'erreur ξ_t de cette régression représente la part de l'innovation de la volatilité qui n'est pas linéairement expliquée par le carré du rendement. Comme expliqué plus haut, cette part peut permettre de capturer soit des non-linéarités importantes (voir par exemple les phénomènes d'asymétrie soulignés dans la section 1) soit la nécessité d'étendre l'information conditionnante à d'autres variables explicatives de la volatilité d'intérêt. La pertinence empirique de ces deux questions a été soulignée récemment comme centrale dans l'appréciation des modèles de volatilité aujourd'hui disponibles par Engle et Patton (2000).

Quoiqu'il en soit, la valeur de ce terme d'erreur ne joue pas de rôle dans le niveau du biais de simultanéité qui est du signe de :

$$\mathrm{Cov}_{t-1}[(\varepsilon_t)^2, \nu_t - \gamma\eta_t] = (\alpha_{t-1} - \gamma)\mathrm{Var}_{t-1}[(\varepsilon_t)^2] \quad (18)$$

L'intuition de ce signe vient de l'interprétation qu'en fournit l'équation structurelle (9) quand on y reporte la décomposition (17) :

$$(\sigma_{t+1})^2 = \omega + (\gamma - \alpha_t)(\sigma_t)^2 + \alpha_t(\varepsilon_{t+1})^2 + \xi_{t+1} \quad (19)$$

La décomposition (19) met en évidence la régression affine (conditionnelle à l'information courante $I(t)$) de la volatilité future sur le carré de l'innovation du rendement. Cette régression, écrite sous la forme

$$EL_t\ [(\sigma_{t+1})^2/(\omega_{t+1})^2] = \omega + (\gamma - \alpha_t)(\sigma_t)^2 + \alpha_t(\varepsilon_{t+1})^2, \quad (20)$$

généralise le modèle GARCH (1,1) à deux niveaux :

(i) D'abord, on ne suppose pas que la volatilité future coïncide avec sa régression affine conditionnelle à $I(t)$ sur le carré du rendement (le terme d'erreur ξ_{t+1} de la régression peut être non nul). Autrement dit, le modèle GARCH(1,1) restreint *a priori* l'information conditionnante considérée comme pertinente pour la prévision de la volatilité.

(ii) Ensuite, les coefficients $\beta_t = \gamma - \alpha_t$ et α_t ne sont pas supposés constants comme dans (2). En fixant *a priori* le coefficient α_t de la régression conditionnelle (17) la modélisation GARCH introduit un risque supplémentaire d'erreur de spécification.

On notera que, comme annoncé dans l'introduction, le risque de modèle de volatilité a deux composantes très différentes : risque sur le choix de l'information conditionnante qui identifie la volatilité d'intérêt (point (i) ci-dessus) et risque sur la spécification de la dynamique de cette volatilité (point (ii)). Alors que le second risque peut être contrôlé par des tests statistiques, le premier a trait à la structure du problème financier lui-même. Évidemment, l'approche en forme réduite, qui consiste à ajouter à l'équation (11) la contrainte $\nu_t = \alpha\eta_t$ pour que cette équation se ramène au modèle ARMA$(1,1)$ de l'équation (4), combine implicitement les deux risques.

On notera plus généralement que, par analogie avec l'interprétation des modèles GARCH, on aura tendance à considérer que le poids $(\gamma - \alpha_t)$ de $(\sigma_t)^2$ dans « l'intercept » de la régression conditionnelle (20) est non négatif, d'où le signe négatif attendu pour le biais de simultanéité conformément à (18).

Enfin, il convient de remarquer que la spécification en forme réduite GARCH $(1,1)$ n'est pas la seule compatible avec le modèle structurel (9).
Meddahi et Renault (2000) exhibent plusieurs exemples de modèles de volatilité récemment proposés dans la littérature avec une approche purement « forme réduite » (spécification directe d'une forme fonctionnelle reliant la valeur courante $(\sigma_t)^2$ de la volatilité avec ses valeurs passées et les valeurs courantes et passées de l'innovation carrée des rendements $(\varepsilon_s)^2, s \leqslant t$) qui peuvent s'interpréter comme le choix (sans fondement structurel) d'une spécification particulière de l'innovation ν_t dans l'équation structurelle (9). Considérons ici à titre d'exemple le modèle « GARCH asymétrique » tel que proposé par Engle et Ng (1993) :

$$(\sigma_{t+1})^2 = \omega^* + \beta(\sigma_t)^2 + \alpha(\varepsilon_{t+1} + \delta)^2. \tag{21}$$

En identifiant avec l'équation structurelle (9) on voit que :

$$\omega = \omega^* + \alpha\delta^2$$

$$\gamma = \beta + \alpha$$

$$\nu_{t+1} = \alpha[(\varepsilon_{t+1})^2 - (\sigma_t)^2] + 2\alpha\delta_{t+1}.$$

Par rapport à la modélisation GARCH standard, le terme correctif $(2\alpha\delta_{t+1})$ dans la spécification du processus d'innovation de l'équation structurelle permet de déconnecter les questions d'asymétrie dynamique et statique puisqu'on a alors :

$$E_{t-1}[(\sigma_t)^2.\varepsilon_t] = \alpha E_{t-1}[(\varepsilon_{t+1})^3] + 2\alpha\delta E_{t-1}[(\varepsilon_{t+1})^2].$$

Il n'y a évidemment pas de limite à l'industrie de l'ajout de tels termes correctifs. Il s'agit toujours, conformément à la philosophie des modèles GARCH, d'une spécification qui impose une corrélation parfaite entre la volatilité et les rendements observés (il n'y a pas d'autres variables d'état dans l'information conditionnante) mais qui relâche éventuellement la contrainte initialement introduite par Engle (1982) et Bollerslev (1986) selon laquelle cette corrélation parfaite serait en outre linéaire par rapport au carré du rendement.

2.2. La question d'une modélisation structurelle paramétrique.

Le modèle structurel de volatilité défini par (9) est semi-paramétrique. Il permet des prévisions de la volatilité dans la mesure où :

(i) Les restrictions des moments conditionnels (16) permettent d'obtenir un estimateur convergent asymptotiquement normal (estimateur GMM de Hansen (1982)) des paramètres d'intérêt ω et γ à partir des équations estimantes :

$$E[[(\varepsilon_{t+1})^2 - \omega - \gamma(\varepsilon_t)^2]z(s)] = 0 \quad \text{pour} \quad s < t,$$

pour tout choix « d'instruments » $z(s)$, c'est-à-dire de variables $I(s)$-adaptées (en particulier $z(s) = \varepsilon_s, (\varepsilon_s)^2, (\varepsilon_s)^3$).

(ii) Une fois les paramètres ω et γ estimés, la volatilité future $(\sigma_{t+h})^2$ à attendre pour la période $[t+h, t+h+1]$ peut être prévue à la date t par une combinaison convexe de la volatilité présente $(\sigma_t)^2$ et de la variance inconditionnelle $\sigma^2 = \omega/(1-\gamma)$:

$$E[(\sigma_{t+h})^2/I(t)] = (1-\gamma^h)\sigma^2 + \gamma^h(\sigma_t)^2. \tag{22}$$

Mais, si le choix d'un modèle très général diminue le risque d'erreur de spécification (voir section 1), il ne donne pas de réponse à certaines questions cruciales pour la finance appliquée. Typiquement, le modèle semi-paramétrique (9) laisse sans couverture plusieurs risques de modèle qu'il faudra en pratique courir en spécifiant un modèle plus paramétré. Analysons ces risques sous l'hypothèse, assez réaliste, de stationnarité de tous les processus considérés.

1^er^ Risque : *La prévision à la date t de la volatilité future par la formule (22) repose sur une évaluation de la volatilité présente* $(\sigma_t)^2 = E[(\varepsilon_{t+1})^2/I(t)]$ *:*

Si on ne spécifie pas davantage le modèle, seule une régression non paramétrique de $(\varepsilon_{t+1})^2$ sur les variables de $I(t)$ peut permettre de répondre à cette question. Le principal problème pratique posé par la régression non paramétrique est la « malédiction de la dimension » (« curse of dimensionality »), c'est-à-dire le fait que la précision de l'estimation de cette régression chute très vite quand le nombre de régresseurs augmente. Le problème se posera puisque le processus ε_t est en général loin d'être markovien [10] (d'ordre un) et donc $(\sigma_t)^2$ risque d'être très mal approché par :

$$E[(\sigma_t)^2/(\varepsilon_t)^2] = E[(\varepsilon_{t+1})^2/(\varepsilon_t)^2].$$

En fait, ε_t n'est markovien d'ordre 1 que dans le cas ARCH(1), c'est-à-dire précisément le cas qui caractérise la volatilité comme fonction déterministe

10. X_t est un processus markovien d'ordre p si la loi conditionnelle de $(X_t,\ s < t)$ coïncide avec la loi de X_t sachant $(X_{t-i}, i = 1, 2, \ldots, p)$.

de l'innovation contemporaine du rendement $(E[(\sigma_t)^2/(\varepsilon_t)^2] = \omega + \alpha(\varepsilon_t)^2 = (\sigma_t)^2)$. Dans le cas général, l'équation (20) suggère d'approcher $(\sigma_t)^2$ par :

$$E[(\sigma_t)^2/(\varepsilon_t)^2, I(t-1)] = \omega + (\gamma - \alpha_{t-1})(\sigma_{t-1})^2 + \alpha_{t-1}(\varepsilon_t)^2.$$

Cela ne dispense pas d'une régression non paramétrique, sauf si l'on est prêt à supposer (comme le fait en particulier la modélisation GARCH) que le coefficient α_t est une constante α. Dans ce cas une équation de récurrence permet, en éliminant de proche en proche les valeurs passées de la volatilité, d'exprimer l'approximation de $(\sigma_t)^2$ comme une moyenne mobile infinie (à poids décroissant à vitesse exponentielle) de la suite des $(\varepsilon_s)^2, s \leqslant t$. Cependant cette série ne donnera qu'une approximation de la valeur $(\sigma_t)^2$ de la « volatilité réalisée » puisqu'elle ne prend pas en compte les termes d'erreur ξ_s, $s \leqslant t$, de l'équation (19). Le modèle GARCH(1, 1) est précisément le modèle en forme réduite caractérisé par la nullité de tous ces termes d'erreur.

2ème Risque : *Même évaluée avec précision, la volatilité ne fournit pas une prévision précise de la variabilité finalement (ex post) constatée pour les rendements :*

Cette remarque a donné lieu à beaucoup de débats sur l'utilité des modèles de volatilité. L'argument des détracteurs est que le risque subi *ex post* dans l'investissement dans un actif sur la période $[t, t+1]$ est mesuré par l'écart (ε_{t+1}) entre son rendement anticipé en t et le rendement réalisé en $(t+1)$. En termes de variance *ex post*, on se trouve donc confronté à $(\varepsilon_{t+1})^2$ alors que l'on n'avait prévu que $(\sigma_t)^2$. De nombreuses études empiriques (voir Andersen et Bollerslev (1998)) ont montré que les modèles usuels de volatilité fournissent une valeur de $(\sigma_t)^2$ qui n'est qu'une approximation pauvre du risque $(\varepsilon_{t+1})^2$ constaté *ex post*. Andersen et Bollerslev (1998) proposent une défense des « modèles usuels de volatilité », en l'occurrence du modèle GARCH(1, 1), en arguant que la question posée n'est pas la bonne : le modèle de volatilité se préoccupe de prévoir la volatilité et non pas la variabilité *ex post* des rendements (« il n'y a, en fait, pas de contradiction entre de bonnes prévisions de la volatilité et un faible pouvoir prédictif des rendements au carré »). Mais cette plaidoirie revient à prendre acte du fait que le modèle prévoit certes bien une quantité qu'il se donne mais que celle-ci n'est qu'une mauvaise approximation de la quantité qui intéresse le praticien. Reprenons cette défense telle qu'elle est explicitée dans Andersen et Bollerslev (1998) (avec une traduction libre et une adaptation à nos notations) :

« La majorité des évaluations de prévision de volatilité qui sont proposées dans la littérature reposent sur un critère d'erreur quadratique moyenne mettant en jeu des rendements au carré (ou en valeur absolue) calculés ex post sur l'horizon de prévision d'intérêt. Une évaluation particulièrement populaire est obtenue par la régression *ex post* rendement carré / volatilité :

$$(\varepsilon_{t+1})^2 = a + b(\sigma_t)^2 + \zeta_{t+1}. \tag{23}$$

Cette équation de régression est l'analogue d'une procédure courante pour évaluer les prévisions d'une moyenne conditionnelle, appelée régression de

Mincer-Zarnowitz à la suite de Mincer et Zarnowitz (1969). Si le modèle de la variance conditionnelle est bien spécifié et $E_t[(\varepsilon_{t+1})^2] = (\sigma_t)^2$, les vraies valeurs des coefficients a et b sont 0 et 1 respectivement. Bien sûr, en pratique les valeurs de $(\sigma_t)^2$ sont sujettes à des erreurs d'estimation, ce qui a pour conséquence un problème standard d'erreur sur les variables (...). Cependant, le coefficient de détermination R^2 de la régression (23) fournit une évaluation directe de la part de la variabilité des rendements *ex post*, telle que mesurée par $(\varepsilon_{t+1})^2$, qui est expliquée par l'estimateur proposé pour $(\sigma_t)^2$. Le coefficient R^2 est ainsi souvent interprété comme une mesure simple du degré de prévisibilité du processus de volatilité, et par conséquent de la pertinence économique potentielle des prévisions de volatilité. L'utilisation de ce R^2 comme un guide sur la précision des prévisions de volatilité est cependant problématique».

A l'appui de cette contestation, Andersen et Bollerslev (1998) proposent un calcul de la valeur théorique de ce coefficient R^2 dans le cas d'un processus GARCH (1, 1) fort. Il est possible de proposer une démonstration plus simple, qui non seulement montre que la formule reste valable dans le cas d'un GARCH(1, 1) semi-fort mais souligne aussi comment la formule peut se trouver modifiée dans le cadre général du modèle de volatilité stochastique (9). Partons pour cela de l'équation d'analyse de la variance :

$$\mathrm{Var}[(\sigma_t)^2] = E[\mathrm{Var}[(\sigma_t)^2/I(t-1)]] + \mathrm{Var}[E[(\sigma_t)^2/I(t-1)].$$

On a :

$$\mathrm{Var}[E[(\sigma_t)^2/I(t-1)] = \mathrm{Var}[\gamma(\sigma_{t-1})^2] = \gamma^2\mathrm{Var}[(\sigma_t)^2]$$

alors que :

$$E[\mathrm{Var}[(\sigma_t)^2/I(t-1)]] = E[\mathrm{Var}[\nu_t/I(t-1)]]$$

$$= E[(\alpha_{t-1})^2\mathrm{Var}[(\varepsilon_t)^2/I(t-1)] + \mathrm{Var}[\xi_t/I(t-1)]]$$

Ainsi, dans le cas particulier où le coefficient de régression conditionnelle (α_t) est constant :

$$E[\mathrm{Var}[(\sigma_t)^2/I(t-1)]] = \alpha^2[\mathrm{Var}[(\varepsilon_t)^2] - \mathrm{Var}[(\sigma_t)^2]] + \mathrm{Var}[\zeta_t]$$

si bien que :

$$[1+\alpha^2-\gamma^2]\mathrm{Var}[(\sigma_t)^2] = \alpha^2\mathrm{Var}[(\varepsilon_t)^2] + \mathrm{Var}[\zeta_t].$$

Puisque la vraie valeur du coefficient b de la régression (23) est 1, le coefficient de détermination R^2 de cette régression vaut :

$$R^2 = \mathrm{Var}[(\sigma_t)^2][\mathrm{Var}[(\varepsilon_t)^2]]^{-1}$$

$$= \alpha^2[1+\alpha^2-\gamma^2]^{-1} + [1+\alpha^2-\gamma^2]^{-1}[\mathrm{Var}[(\varepsilon_t)^2]]^{-1}\mathrm{Var}[\zeta_t]. \quad (24)$$

Ce résultat est valide dans les conditions générales du modèle structurel (9) à condition de supposer que le coefficient de régression conditionnelle (α_t) est constant. Ainsi, dans le cas particulier du modèle GARCH$(1,1)$ semi-fort, il est exact, comme l'ont remarqué Andersen et Bollerslev (1998) (dans le cas particulier du modèle GARCH$(1,1)$ fort), que :

$$R^2 = \alpha^2[1 + \alpha^2 - \gamma^2]^{-1} = [1 + (1 - \gamma^2)/\alpha^2]^{-1}. \qquad (25)$$

Loin d'être une défense du modèle GARCH$(1,1)$, cette formule met au contraire en évidence à notre avis certaines lacunes de ce modèle.

D'abord, pour une persistance donnée ($\gamma = \alpha + \beta$ de la variance conditionnelle, la part de la variance expliquée par la régression (23) n'excédera jamais γ (quand $\gamma = \alpha$ et $\beta = 0$) et ce maximum ne sera atteint que dans le cas extrême d'un modèle ARCH(1). La formule générale (24) montre que cette part aurait pu être supérieure si l'on n'avait pas imposé la nullité du terme d'erreur ζ_t dans la régression (17), c'est-à-dire si l'on n'avait pas renoncé à incorporer une information conditionnante plus large dans la prévision de la volatilité. Autrement dit, dans la mesure où l'on choisit une définition de la volatilité comme une fonction affine des valeurs passées du carré du rendement (ou de son innovation), il est exact que l'on s'impose a priori de ne prévoir qu'une partie de la variabilité constatée *ex post* pour le rendement. Ce n'est pas une surprise si l'on se souvient en particulier du quatrième risque mis en évidence dans la section 1. Il est par exemple hors de question de vouloir expliquer avec un modèle GARCH(1,1) fort (et une innovation standardisée normale ou leptokurtique) ne serait – ce que plus du tiers de la variabilité *ex post*! On vérifie en effet aisément que :

$$[1 + (1 - \gamma^2)/\alpha^2]^{-1} > 1/3 \Longleftrightarrow \gamma^2 + 2\alpha^2 > 1$$

ce qui est précisément la condition nécessaire et suffisante (8) pour que la variabilité à expliquer, $\mathrm{Var}[(\sigma_t)^2]$, soit infinie dans le cas conditionnellement gaussien.

En conclusion, le modèle GARCH$(1,1)$ place l'utilisateur devant l'alternative suivante : ou bien la volatilité qu'il définit est peu persistante ($\gamma^2 + 2\alpha^2 \leqslant 1$) et les prédictions que l'on en fera seront par conséquent peu informatives (la prédiction calculée selon la formule (22) sera proche de la variance inconditionnelle), ou bien la variance inconditionnelle de la volatilité que l'on cherche à prévoir est infinie (puisque, d'après (24), $\mathrm{Var}[(\sigma_t)^2]$ est proportionnelle à $\mathrm{Var}[(\varepsilon_t)^2]$; voir aussi $3^{\text{ème}}$ risque ci-dessous).

Enfin, pour une persistance donnée $\gamma = \alpha + \beta$ de la variance conditionnelle, la part de la variance expliquée par la régression (23) est une fonction décroissante du coefficient β qui tend vers 0 quand β tend vers γ (α tend vers 0). Comme la plupart des études empiriques conduisent à une valeur estimée de β au moins triple de celle de α (voir en particulier Bollerslev (1986)), on en déduit effectivement des valeurs infimes pour le coefficient de détermination (25).

3ème Risque : *Une prévision de la volatilité doit être accompagnée d'un intervalle de prévision qui repose sur un modèle de la variance conditionnelle de la variance conditionnelle :*

Engle (1982) avait introduit les processus ARCH pour rendre opérationnelle l'idée que «le passé récent fournit de l'information au sujet de la variance des prévisions», citant McNees (1979) pour remarquer que «l'incertitude inhérente à la prévision à différentes périodes semble varier sensiblement dans le temps». Autrement dit, l'accent était mis sur la prévision d'une série chronologique observable d'intérêt (en l'occurrence le niveau de l'inflation au Royaume-Uni dans les années 1970) et le modèle ARCH ne servait qu'à calculer l'intervalle de prévision à fournir autour d'une prévision donnée. La question est très différente quand la variable d'intérêt à prévoir (pour les applications financières) est la volatilité elle-même. On doit en effet alors se préoccuper, pour évaluer la fiabilité de cette prévision, de la distribution conditionnelle, et en particulier de la variance conditionnelle, de la volatilité future $(\sigma_{t+1})^2$ sachant l'information $I(t)$ disponible.
D'après l'équation structurelle (9) :

$$\mathrm{Var}[(\sigma_{t+1})^2/I(t)] = \mathrm{Var}[(\nu_{t+1})/I(t)]$$

et c'est, pour ainsi dire fortuitement, que la modélisation GARCH(1,1), en choisissant la forme réduite :

$$\nu_{t+1} = \alpha[(\varepsilon_{t+1})^2 - (\sigma_t)^2]$$

conduit à :

$$\mathrm{Var}[(\sigma_{t+1})^2/I(t)] = \alpha^2\mathrm{Var}[(\varepsilon_{t+1})^2/I(t)] = \alpha^2(\sigma_t^4)\mathrm{Var}[(u_{t+1})^2/I(t)] \quad (26)$$

où $u_{t+1} = (\sigma_{t+1})/(\sigma_t)$ est l'innovation standardisée. En particulier, dans le cas d'un GARCH fort, celle-ci est un bruit indépendant du passé, si bien que la variance conditionnelle de $(\sigma_{t+1})^2$ est proportionnelle à (σ_t^4). En d'autres termes, les périodes où le rendement est prévu avec le plus d'imprécision sont aussi celles où sa volatilité est elle-même prévue avec le plus de risque d'erreur. Cette propriété, sans doute naturelle et conforme à l'observation, est en revanche sûrement exagérée par la modélisation GARCH qui suppose que l'amplitude des intervalles de prévision de la volatilité va varier comme le carré de l'amplitude des intervalles de prévision du rendement. On peut imaginer des variations beaucoup moins marquées de la volatilité de la volatilité. Il s'agit en fait de compléter la spécification structurelle (9) par une hypothèse sur l'hétéroscédasticité conditionnelle de cette équation en proposant une modélisation structurelle, centrée sur la quantité d'intérêt, plutôt qu'une forme réduite, comme l'hypothèse «GARCH fort» dont on montrera (voir 4e risque ci-dessous) qu'elle ne peut pas avoir un fondement structurel. On donne ci-dessous deux exemples d'une telle modélisation.

Exemple 1 : Barndorff - Nielsen et Shephard (2000) :

Le processus autorégressif de la volatilité représenté par l'équation (9) est particulier parce qu'il décrit la dynamique d'un processus présumé stationnaire mais à valeurs positives. Il n'est donc pas question d'imaginer par exemple que l'innovation ν_t de cette équation soit un bruit blanc gaussien. Rien n'interdit en revanche de supposer que cette innovation est un bruit indépendamment identiquement distribué. Cela signifie simplement que la distribution stationnaire de la volatilité centrée :

$$x_{t+1} = (\sigma_{t+1})^2 - [\omega/(1-\gamma)]$$

est celle d'une variable aléatoire X telle que, pour tout γ, $0 < \gamma < 1$, on ait :

$$X =^d \gamma X + \vartheta$$

où ϑ est une variable aléatoire indépendante de X et $=^d$ signifie égalité en distribution. Les distributions de probabilité qui vérifient cette propriété sont par définition les distributions auto-décomposables ; il s'agit d'une sous-classe (voir par exemple Bondesson (1992) pour des exemples et caractérisations) de l'ensemble des distributions infiniment divisibles (distributions qui, pour tout entier positif n, peuvent être vues comme la somme de n variables aléatoires indépendantes de même loi). Une modélisation structurelle paramétrique de la volatilité peut alors être déduite du résultat suivant (Bondesson (1992), p 18) : Si $N(\tau)$ est un processus de Lévy [11], pour k positif donné, le processus :

$$X(t) = \int_{]-\infty,t]} \exp[-k(t-\tau)]N(d\tau)$$

est auto-décomposable au sens où, pour tout h positif :

$$X(t+h) = \exp(-kh)X(t) + \vartheta_h(t+h),$$

$\vartheta_h(t+h)$ et $X(t)$ sont indépendants, et $\vartheta_h(.)$ est un processus stationnaire. Les qualités structurelles d'une telle modélisation de la volatilité sont :

(i) D'une part l'interprétation d'un processus de Lévy comme processus de cumul de sauts d'amplitudes indépendantes arrivant selon un processus de Poisson. Ceci est conforme à l'interprétation économique de la volatilité stochastique en termes d'arrivée sur le marché de chocs informationnels (Clark (1973)).

(ii) D'autre part, le formalisme en temps continu sous-jacent garantit que le modèle est robuste à l'agrégation temporelle, c'est-à-dire ne dépend pas du pas de temps sur lequel les rendements sont calculés.

Ce modèle de volatilité a été étudié en détail par Barndorff-Nielsen et Shephard (2000). En termes de modèle semi-paramétrique, cela suggère de compléter le modèle (9) par l'hypothèse :

$$\mathrm{Var}[(\sigma_{t+1})^2/I(t)] = \mathrm{Var}[(\nu_{t+1})/I(t)] \quad \text{indépendant de } I(t). \qquad (27)$$

11. Un processus de Lévy est un processus à accroissements indépendants et stationnaires.

En termes de réalisme empirique, on peut se demander si la vérité ne se situe pas à mi-chemin entre les deux extrêmes (26) et (27). Si les variations de la volatilité de la volatilité ne sont pas aussi marquées que celles prévues par un modèle GARCH fort, il serait cependant surprenant que dans une période où les prévisions du rendement sont plus imprécises (comme les « chaotiques années 70 » de Engle (1982)), les prévisions de la volatilité ne le soient pas simultanément. D'où l'exemple 2.

Exemple 2 : Meddahi et Renault (1996, 2000)

Le recours aux processus de Lévy dans l'exemple 1 a été motivé par le souhait de rendre plus paramétrique le modèle structurel (9) tout en étant confronté à l'impossibilité de modéliser la volatilité par un processus AR(1) gaussien. Mais une autre piste, exploitée par Meddahi et Renault (2000), consiste à voir celle-ci comme le carré d'un processus AR(1) gaussien. En effet, si :

$$Y_{t+1} = \rho Y_t + \delta v_{t+1} \quad \text{avec} \quad v_t \quad \text{bruit blanc gaussien standard}$$

on voit que, en posant $(\sigma_{t+1})^2 = (Y_{t+1})^2$, on a :

$$(\sigma_{t+1})^2 = \delta^2 + \rho^2(\sigma_t)^2 + \nu_{t+1}$$

avec :

$$\nu_{t+1} = \delta^2[(v_{t+1})^2 - 1] + 2\delta\rho Y_t v_{t+1}.$$

Le terme d'erreur ϑ_{t+1} vérifie :

$$E[\vartheta_{t+1}/I(t)] = 0$$

$$\mathrm{Var}[\vartheta_{t+1}/I(t)] = 2\delta^4 + 4\delta^2\rho^2(\sigma_t)^2.$$

On a donc bien ainsi construit un modèle paramétrique de la volatilité compatible avec (9) avec :

$$\omega = \delta^2, \gamma = \rho^2$$

et

$$\mathrm{Var}[(\sigma_{t+1})^2/I(t)] \quad \text{fonction affine de} \quad (\sigma_t)^2. \tag{28}$$

Ce modèle « affine » de la volatilité, tant pour sa moyenne conditionnelle que sa variance conditionnelle, a été développé par Meddahi et Renault (1996). Sans doute plus réaliste que (26) et (27), il partage avec (27) l'avantage d'être compatible avec un modèle en temps continu sous-jacent comme le processus « racine carrée » ou ses extensions affines proposées récemment par Duffie, Pan et Singleton (2000) pour les processus de taux d'intérêt : $d\sigma_t^2 = k(\theta - \sigma_t^2)dt + (a + b\sigma_t^2)^{1/2}dW_t$ où W_t est un processus de Wiener.

Un autre avantage du modèle (28), mis en exergue par Meddahi et Renault (1996), est que, si on imagine des primes de risque, fonctions affines de $(\sigma_t)^2$, qui s'intègrent à l'espérance conditionnelle des rendements quand ceux-ci sont calculés sur une certaine durée, le modèle ne sera pas modifié par changement de la fréquence d'échantillonnage. On a donc une extension structurelle du modèle ARCH-M de Engle, Lilien et Robins (1987).

4$^{\text{ème}}$Risque : *L'estimation efficace, par variables instrumentales, des paramètres ω et γ de la dynamique de la volatilité, repose sur la caractérisation de l'hétéroscédasticité conditionnelle des erreurs sur les variables (de volatilité) et donc sur celle de la kurtosis conditionnelle :*

Il s'agit en effet, quand on veut exploiter efficacement les conditions de moments conditionnels (16), de savoir évaluer l'hétéroscédasticité conditionnelle (sachant $I(t-1)$) de :

$$(\varepsilon_{t+1})^2 - \omega - \gamma(\varepsilon_t)^2 = \nu_t - \gamma\eta_t + \eta_{t+1}. \tag{29}$$

Ainsi, outre la volatilité de la volatilité Var $[\vartheta_{t+1}/I(t)]$ dont la modélisation a été discutée ci-dessus, on doit s'intéresser à la variance conditionnelle Var $[\eta_{t+1}/I(t)]$ des erreurs sur les variables (et aussi aux corrélations conditionnelle entres ν_t et η_t). Il n'y a que dans le cas GARCH que les deux problèmes sont confondus puisque $\nu_{t+1} = \alpha[(\varepsilon_{t+1})^2 - (\sigma_t)^2] = \alpha\eta_{t+1}$.

Dans le cas général, une façon usuelle de spécifier la variance conditionnelle Var $[\eta_{t+1}/I(t)]$ est de remarquer que :

$$\text{Var}[\eta_{t+1}/I(t)] = (\sigma_t)^4\text{Var}[(u_{t+1})^2/I(t)]$$

puis de supposer que l'innovation standardisée (u_{t+1}) est un bruit blanc indépendant de loi fixe de coefficient de kurtosis K si bien que :

$$\text{Var}[\eta_{t+1}/I(t)] = (K-1)(\sigma_t)^4.$$

Bien que cette approche soit la plus courante, on va montrer qu'il est difficile d'en imaginer une interprétation structurelle parce qu'elle n'est pas robuste à l'agrégation temporelle. En effet, si l'on divise par deux la fréquence d'observation des prix avec lesquels sont calculés les rendements en actualisation continue ($\varepsilon_{t+1} = \text{Log}[p_{t+1}/p_t]$), on se trouve confronté à une innovation standardisée :

$$u_t^{(2)} = [\sigma_{t-2}u_{t-1} + \sigma_{t-1}u_t][(\sigma_{t-2})^2 + E[(\sigma_{t-1})^2/I(t-2)]]^{-1/2}$$

$$u_t^{(2)} = [\sigma_{t-2}u_{t-1} + \sigma_{t-1}u_t][\omega + (1+\gamma)(\sigma_{t-2})^2]^{-1/2}$$

Ainsi, si l'on veut maintenir l'hypothèse :

$$E[(u_t^{(2)})^4] = K^{(2)} \quad \text{alors que} \quad E[(u_t)^4] = K, \tag{30}$$

on va nécessairement introduire une contrainte entre ces coefficients de kurtosis et la dynamique de la volatilité. Une telle contrainte, reliant deux phénomènes d'origines totalement différentes ne peut pas avoir de justification structurelle. Dans la mesure où il s'agit juste ici d'énoncer un contre-exemple, on peut se contenter d'expliciter cette contrainte dans un contexte particulier caractérisé par les hypothèses suivantes :

(i) absence d'asymétrie, aussi bien statique que dynamique (nullité des covariances considérées dans (6) et (7)),
(ii) $\mathrm{Cov}[(u_t)^2, (\sigma_t)^2/I(t-1)] = \delta(\sigma_{t-1})^2$.

Nous utilisons ces hypothèses juste pour obtenir une formule explicite des contraintes alléguées, mais la généralité de l'argument n'en dépend pas. Notons cependant que la seconde hypothèse peut être justifiée de deux façons :

- d'abord, elle est vérifiée dans le cas d'un processus GARCH(1,1) semi-fort (avec $\delta = 2\alpha$),
- ensuite, elle est impliquée (avec $\delta = 0$) par une hypothèse forte d'absence d'effet de levier (ν_t indépendant de u_t).

On vérifie alors, par un calcul élémentaire, que :

$$\mathrm{Var}[(\sigma_t)^2/I(t-1)] = a(\sigma_{t-1})^4 + b(\sigma_t)^2 + c$$

avec :

$$a = [(K^{(2)}/K) - 1](1+\gamma)^2 + 2\gamma(1 - 3/K) - 6(\delta/K)$$

$$b = 2\omega[(K^{(2)}/K) - 1]\gamma + 2(\omega/K)(K^{(2)} - 3)$$

$$c = \omega^2[(K^{(2)}/K) - 1].$$

Ainsi, pour justifier ces contraintes, on devrait justifier que la dynamique de la volatilité de la volatilité soit définie par un polynôme (en la volatilité passée) dont les coefficients sont reliés aux coefficients de kurtosis conditionnelle sur différentes périodes. Une telle hypothèse peut convenir pour un filtrage en forme réduite mais sûrement pas pour une modélisation structurelle. C'est encore plus frappant si l'on imagine que, conformément à la pratique de modélisation la plus courante, on postule a priori une loi fixe (normale, Student à nombre donné de degrés de liberté, etc.) pour l'innovation standardisée. En effet, la contrainte additionnelle $K^{(2)} = K$ implique immédiatement que :

$$(3/K) = [2(\sigma_{t-1})^2 E[(\sigma_t)^2/I(t-1)] \\ - \mathrm{Var}[(\sigma_t)^2/I(t-1)]][2(\sigma^2_{t-1}) E[(\sigma_t)^2/I(t-1)]]^{-1}.$$

En particulier, cela n'est possible que dans le cas d'une distribution conditionnelle leptokurtique ($K > 3$).

Le message de ce quatrième risque est selon nous l'impossibilité structurelle d'imposer une kurtosis conditionnelle constante. Il s'agit donc d'ajouter au modèle de volatilité celui de la kurtosis conditionnelle. Cela rejoint la difficulté dont nous sommes partis en introduction. La décomposition de l'innovation du rendement en :

$$\varepsilon_{t+1} = \sigma_t u_{t+1}$$

ne garantit en aucun cas que toute la dynamique d'intérêt passe par la volatilité. Les innovations standardisées u_{t+1} présentent elles aussi des aspects dynamiques importants que l'on ne pourra pas modéliser de façon satisfaisante sans recourir à une définition précise de l'information conditionnante et à la spécification d'un processus multivarié en temps continu (voir les deux exemples de référence donnés dans le troisième risque ci-dessus).

3. CONCLUSION

Une pratique ancienne, perpétuée dans les méthodologies les plus modernes, consiste à détecter les variations de la volatilité en en calculant des « estimateurs » sur des fenêtres glissantes. Ces fenêtres peuvent être rétrospectives (point de vue de la **volatilité historique** calculée à partir d'une moyenne mobile des innovations au carré des rendements passés, pour une donnée d'un système de poids, comme par exemple le lissage exponentiel proposé par *Riskmetrics*) ou prospectives (point de vue de la volatilité implicite dans un prix d'option censée traduire la prévision du marché sur la volatilité future). Le rôle de cette fenêtre mobile serait – selon une « utopie » souvent attribuée à Merton (1980) (voir le titre de l'article récent de Bai, Russel et Tiao (2000)) – d'obtenir une « observation » de la volatilité, ce qui simplifierait évidemment beaucoup la question de sa modélisation et des tests de spécification. Andersen, Bollerslev, Diebold et Labys (2000) ont ainsi popularisé le terme de « volatilité réalisée » pour l'évaluation de la volatilité sur une période donnée, par exemple la journée, à travers la variation quadratique interne à cette journée calculée grâce à des données haute-fréquence. Cela leur permet de tester la spécification du modèle de volatilité (log-normalité, dynamique auto-régressive, etc.) sans être confrontés au problème de l'hypothèse jointe puisque l'observation de la volatilité est censée « *model-free* », c'est-à-dire immunisée contre le risque de modèle.

La thèse centrale de cet article est précisément que les stratégies opérationnelles de prévision statistique (de la volatilité) « *model-free* » n'existent pas davantage que les opportunités d'arbitrage (« *free lunch* ») sur les marchés financiers. Qu'il s'agisse de « *free lunch* » ou de prévision « *model-free* », leur éventuelle apparition ne résiste pas à l'examen sérieux des conditions réelles de la transaction. Ainsi, toute mesure de la volatilité repose sur un choix, toujours contestable comme expliqué plus haut, d'une certaine structure d'information conditionnante. De plus, la justification d'une approximation comme celle de la variation quadratique et l'évaluation de sa précision se fondent nécessairement sur un modèle restrictif de la dynamique sous-jacente de la volatilité. Enfin, la volatilité implicite dans un prix d'option ne peut être interprétée qu'en se référant à un modèle d'évaluation d'options particulier. On est donc bien loin d'avoir réalisé l'immunisation contre le risque de modèle. Par conséquent, le choix d'un modèle de volatilité pour une application financière particulière confrontera toujours à un compromis rendement / risque sur le modèle lui-même. Nous avons explicité ici cet arbitrage tant au niveau d'une approche en forme réduite que d'une modélisation structurelle. On a montré dans les deux cas que la minimisation des risques de modélisation de la volatilité ne peut se faire qu'au détriment du profit que l'on espère tirer de ces modèles.

Les auteurs sont très reconnaissants à John Galbraith, René Garcia, Nour Meddahi et un rapporteur anonyme pour de précieux commentaires sur ce texte. Les positions prises ici ainsi que les erreurs ou omissions éventuelles sont bien sûr de l'entière responsabilité des auteurs.

RÉFÉRENCES BIBLIOGRAPHIQUES

ALAMI A. [1999], « Estimation de la persistance de la volatilité : biais, variance et traitement des séries financières de haute fréquence », *Thèse de Doctorat en Science Économique*, Université Paris 9 Dauphine.

ANDERSEN T. G. [1994], « Stochastic autoregressive volatility : A framework for volatility modeling », *Mathematical Finance* 4, 75-102.

ANDERSEN T. G. et T. BOLLERSLEV [1998], « Answering the skeptics : Yes, standard volatility models do provide accurate forecasts », *International Economic Review* 39-4, 885-905.

ANDERSEN T. G., T. BOLLERSLEV, F. X. DIEBOLD et P. LABYS [2000], « The Distribution of Realized Exchange Rate Volatility », *Journal of the American Statistical Association*, à paraître.

BAI X., J. R. RUSSELL et G. C. TIAO [2000], « Beyond Merton's Utopia : effects of non-.normality and dependence on the precision of variance estimates using high-frequency financial data », Document de Travail GSB Chicago.

BARNDORFF-NIELSEN O. E et N. SHEPHARD [2000], « Non-Gaussian OU based models and some of their uses in financial economics », *Journal of the Royal Statistical Society*, à paraître.

BLACK F. [1976], « Studies of Stock Market Volatility Changes », *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

BOLLERSLEV T. [1986], « Generalized Autoregressive Conditional Heteroskedasticity », *Journal of Econometrics* 31, 307-327.

BOLLERSLEV T., R. ENGLE et D. NELSON [1994], « ARCH Models » in R.F. Engle et D. McFadden (eds), *Handbook of Econometrics*, 4, Elsevier Science.

BONDESSON L. [1992], *Generalized Gamma convolutions and related classes of distributions and densities*, Springer Verlag.

BROZE L., C. FRANCQ et J.M. ZAKOIAN [1999], « Efficient use of high order autocorrelations for estimating autoregressive processes », Document de travail, Université de Lille.

CHRISTOFFERSEN, P., J.HAHN et A. INOUE [2001], « Testing and Comparing Value-at-Risk Measures », Document de travail, CIRANO, 2001-03.

CLARK P. K. [1973], « A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices », *Econometrica* 41, 133-155.

COMTE F. et E. Renault [1998], « Long memory in continuous time stochastic volatility models », *Mathematical Finance*, 8, 4, 291-323.

DAVIES R. A. et T. MIKOSCH [1998], « The sample autocorrelations of heavy-tailed processes with applications to ARCH », *Annals of Statistics* 26-5, 2049-2080.

DROST F. C. et Th. E. NIJMAN [1993], « Temporal Aggregation of GARCH processes », *Econometrica* 61, 909-927.

DROST F. C. et B. J. M. WERKER [1996], « Closing the GARCH gap : Continuous Time GARCH Modeling », *Journal of Econometrics* 74, 31-58.

DUFFIE D., J. PAN et K. SINGLETON [2000], « Transform analysis and asset-pricing for affine jump-diffusions », *Econometrica* 68-6, 1343-1376.

ENGLE R. F. [1982], « Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation », *Econometrica* 50-4 , 987-1006.

ENGLE R. F. [1995], *ARCH selected readings, Introduction*, Oxford University Press.

ENGLE R. F., D. HENDRY et D. TRUMBLE [1985], « Small Sample Properties of ARCH estimators and tests », *Canadian Journal of Economics* 18, 66-93.

ENGLE R. F., D.LILIEN et R.P. ROBINS [1987], « Estimating Time-Varying Risk Premia in the Term Structure », *Econometrica* 55-2, 391-407.

ENGLE R. F. et V. K. Ng [1993], « Measuring and Testing the Impact of News on Volatility », *Journal of Finance* 48, 1749-1778.

ENGLE R. F. et A. J.PATTON [2000], « What good is a volatility model ? »Working Paper Stern *School of Business.*

FRIEDMAN M. [1957], *A Theory of the Consumption Function*, Princeton University Press.

GARCIA R., R. LUGER et E. RENAULT [2000], « Asymmetric Smiles, Leverage Effects and Structural Parameters », Document de Travail CIRANO.

GHYSELS E., A. HARVEY et E.RENAULT [1996], « Stochastic Volatility »in Maddala G.S. et Rao C.R. (eds), *Handbook of Statistics*, 14, 119-191, Elsevier Science.

GOURIEROUX C. et A. MONFORT [1989], « Statistique et Modèles Econométriques », Vol 1, *Economica.*

HANSEN L. P. [1982], « Large Sample Properties of Generalized Method of Moments », *Econometrica* 50, 1029-1054.

HANSEN L. P. et S. F. RICHARD [1987], « The Role of Conditionning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models », *Econometrica* 55,587- 614.

HARVEY A. C. [1989], *Forecasting structural time series models and the Kalman Filter*, Cambridge University Press.

LAMOUREUX C. G. et D. LASTRAPES [1990], « Persistence in Variance, Structural Change and the GARCH Model », *Journal of Business & Economic Statistics* 8-2, 225-234.

MCNEES S. S. [1979], « The Forecasting Record for the 1970's », *New England Economic Review*, September 33-53.

MEDDAHI N. et E. RENAULT [1996], « Aggregation and Marginalization of GARCH and Stochastic Volatility Models », Document de travail GREMAQ 96-30-433.

MEDDAHI N. et E. RENAULT [1997], « Quadratic M-estimators for ARCH-type processes », Document de travail CRDE 3197.

MEDDAHI N. and E. RENAULT [2000], « Temporal Aggregation of Volatility Models », Document de Travail CIRANO, 2000-22.

MERTON R. [1980], « On estimating the expected return on the market », *Journal of Financial Economics* 8, 323-361.

MIKOSCH T. et C. STARICA [1999], « Change of structure in financial time series, long range dependence and the GARCH model », Document de travail, Université de Groningen.

MINCER J. et V. ZARNOWITZ [1969], « The evaluation of economic forecasts », dans J. Mincer ed., *Economic Forecasts and Expectations*, NBER.

NELSON D.B. [1991], « Conditional Heteroskedasticity in Asset Returns : A New Approach », *Econometrica* 59-2, 347-370.

NIJMAN Th. et E. SENTANA [1996], « Marginalization and Contemporaneous Aggregation of Multivariate GARCH Processes », *Journal of Econometrics* 71, 71-87.

PALM F.C. [1996], « GARCH Models of Volatility », in Maddala G.S. et Rao C.R. (eds), *Handbook of Statistics*, 14, 209-240, Elsevier Science.

RENAULT E. [1997], « Econométrie de la Finance : la méthode des moments généralisés », dans Y. Simon ed., *Encyclopédie des Marchés Financiers*, Tome 1, Chap19, 330-407, Economica.

RENAULT E. et N. TOUZI [1996], « Option hedging and implied volatilities in a stochastic volatility model », *Mathematical Finance* 6, 259-302.

ROMER D. [1996], *Advanced Macroeconomics*, Mc Graw-Hill.

SHEPHARD, N. [1996], « Statistical aspects of ARCH and stochastic volatility » in D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielson (Eds.), *Time Series Models in Econometrics*, Finance and Other Fields, 1-67. London : Chapman & Hall.

TAYLOR S. [1986], Modelling Financial Time Series, John Wiley.