



HAL
open science

Modèles Non Linéaires et Prévisions

Gilbert Colletaz, Christophe Hurlin

► **To cite this version:**

| Gilbert Colletaz, Christophe Hurlin. Modèles Non Linéaires et Prévisions. 2007. halshs-00156692

HAL Id: halshs-00156692

<https://shs.hal.science/halshs-00156692v1>

Preprint submitted on 22 Jun 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rapport de Recherche
Institut CDC pour la Recherche

Modèles Non Linéaires et Prévisions

G. COLLETAZ et C. HURLIN
LEO – Laboratoire d’Economie d’Orléans
UMR CNRS 6221

Novembre 2006

Laboratoire d'économie d'Orléans
Rue de Blois
BP6739
45067 Orléans Cedex 2
Tel : 02 38 49 49 88

Gilbert.Colletaz@univ-orleans.fr ; Christophe.Hurlin@univ-orleans.fr

Nous tenons à remercier l'Institut pour la Recherche de la Caisse des Dépôts et Consignation, et plus spécifiquement Isabelle Laudier, pour leur appui à ce travail.

Contents

1. Introduction	1
2. Les Modèles Univariés Non Linéaires	7
2.1 Les Processus Bilinéaires	9
2.2 Les Processus TARMA	12
2.2.0.1 Les processus <i>TMA</i>	14
2.2.0.2 Les processus <i>TAR</i>	15
2.3 Les Processus <i>STAR</i>	19
2.4 Les Processus à Changement de Régimes Markovien	25
2.5 Le Processus <i>ACR</i>	27
3. Comment Construire une Prévion dans un Modèle Non Linéaire ?	31
3.1 Méthodes de Prévion Ponctuelle	32
3.1.1 Méthode Analytique	33
3.1.2 Méthodes de Simulation Numérique	35
3.1.3 Méthode "Naïve" ou Méthode dite du "Skeleton"	38
3.1.4 Méthode <i>Normal Forecast Error</i> (NFE)	39
3.1.5 Méthode de l'Estimation Dynamique ou <i>Dynamic Estimation (DE) Method</i>	41
3.1.6 Application au Modèle SETAR	42
3.2 Méthodes de Prévion par Intervalle de Confiance et par Densité	46
3.2.1 Prévion par Densité	47
3.2.2 Prévion par Intervalle de confiance	49

4. Comment Evaluer une Prévision dans un Modèle non Linéaire ?	51
4.1 Evaluation des Prévisions Ponctuelles	54
4.1.1 Les Tests de Diebold et Mariano (1995)	55
4.1.2 Les Tests de Harvey, Leybourne et Newbold (1997) .	59
4.1.3 Les Tests Non-Paramétriques de Prévisions de Régime : Pesaran et Timmerman (1992)	60
4.1.4 Tests de Prévisions Enveloppées (Forecast Encompassing) : Harvey, Leybourne et Newbold (1998) . .	63
4.1.5 Le Test d'Échec de Prévision : Clements et Hendry (2002)	65
4.2 Evaluation des Intervalles de Confiance	66
4.2.1 Définition de la couverture conditionnelle	68
4.2.2 Tests de Ratios de Vraisemblance : Christoffersen (1998)	69
(a) Test de couverture non conditionnelle	69
(b) Test d'indépendance des violations .	70
(c) Test de couverture conditionnelle . .	71
4.3 Evaluation des Densités de Prévisions	71
4.3.1 Tests de Spécification Correcte	72
4.3.1.1 Le Test de Bai (2003)	75
4.3.1.2 Application au modèle SETAR	77
4.3.2 Tests de Comparaison de Densités Conditionnelles Mal Spécifiées	81
4.3.2.1 Le test de Bao, Lee et Saltoglu (2004)	83
5. Conclusion	89
Appendix A Bibliographie	91

Chapter 1

Introduction

Depuis quelques années, la prise en compte de la non linéarité et plus spécifiquement de l'existence de phénomènes de changement de régimes tend à modifier profondément les approches de l'économétrie appliquée à la macroéconomie et à la finance. Cette évolution en devenir est sans nul doute comparable à celle qu'a pu connaître la micro-économie lorsque l'on a progressivement abandonné l'univers de référence walrassien, que nous pouvons assimiler à la modélisation linéaire en économétrie, pour s'orienter vers les multiples formes de la concurrence imparfaite, auxquelles nous pouvons assimiler les innombrables modélisations non linéaires.

En effet, il existe aujourd'hui un très grand nombre d'études tendant à démontrer la présence de changements structurels permanents et l'existence d'asymétries dans la dynamique des principaux agrégats macroéconomiques ou financiers. Etant donné qu'il est impossible de rendre compte de ce type de phénomènes à partir des modèles linéaires autorégressifs usuels qu'ils soient univariés (*ARMA*) ou multivariés (*VAR*), il a été nécessaire de considérer des processus non linéaires qui soient en mesure de reproduire ces caractéristiques.

Il s'avère que ces nouvelles formes de modélisations conduisent à un profond renouvellement de la problématique de la prévision à court et moyen terme des séries macroéconomiques et financières. Premièrement, ces mod-

élisations à changement de régimes impliquent de repenser la définition même de ce que doit être une prévision (prévision ponctuelle, par intervalle ou par densité). Dans le cas de certains modèles, se posent un certain nombre de problèmes techniques relatifs à la construction de la prévision ponctuelle. Mais au-delà de ces problèmes techniques, l'utilisation des modèles non linéaires à des fins de prévision a conduit à s'intéresser de plus en plus à la construction d'une densité de prévision et non plus à de simples prévisions ponctuelles. L'estimation ou le calcul direct d'une densité de prévision prend en effet tout son sens dans le cas des modèles à changement de régime : cet exercice permet de révéler d'éventuelles asymétries ou autres multi-modalités de la distribution des prévisions que ne pouvaient générer de façon endogène les modèles linéaires. Enfin, au-delà de la prévision sur le niveau même de la variable se pose bien évidemment un problème totalement nouveau dans cette perspective qui est celui de la prévision du régime dans lequel se situera cette variable à l'avenir.

Deuxièmement, ces modélisations à changement de régimes impliquent de repenser les critères d'évaluation de ces prévisions. En effet, si l'on désire juger la qualité des prévisions des modèles à changement de régimes, en particulier les prévisions par densité, il apparaît essentiel de recourir à des critères différents de ceux qui étaient généralement utilisés dans le cadre des prévisions réalisées à partir de modèles linéaires. C'est pourquoi une très abondante littérature est aujourd'hui consacrée aux différents aspects de la comparaison entre les prévisions des modèles linéaires, des modèles non linéaires en général et des modèles à changement de régimes (de type *STAR* ou *MS-AR*) en particulier. C'est de cette diversité des critères d'évaluation des prévisions que peut provenir un sentiment de flou quant à l'apport des modèles non linéaires quant à la prévision économique.

Dans ce rapport de recherche, nous proposons une synthèse de l'abondante littérature technique consacrée aux modèles non linéaires et aux modèles à changement de régimes qui a pour objectif d'éclairer les dif-

férentes dimensions du débat sur l'apport de ces modèles à la prévision. Ce rapport comporte trois parties.

Dans la première partie, nous passons en revue les principaux modèles univariés non linéaires utilisés à des fins de prévisions. En nous limitant au cadre univarié qui est assez simple pour permettre de comprendre le principe des idées proposées et dont généralement la traduction multivariée est une extension immédiate, et en considérant des modèles qui sont une généralisation de la méthodologie *ARMA* dominante dans l'univers linéaire, nous présentons en premier lieu des processus capables d'approximer toute fonction continue non linéaire. Le processus type de cette approche est constitué par le modèle bilinéaire. Outre les difficultés de spécification et d'estimation qui lui sont propre, on reproche souvent à cette approche l'absence de toute référence théorique de description du phénomène étudié. In fine l'utilisateur dispose d'un modèle dont la justification est purement statistique. Il n'est donc pas surprenant que la majorité des travaux appliqués d'économétrie non linéaire aient suivi une autre voie en recourant à des modèles autour desquels une "histoire" économique peut être bâtie. On retrouve alors tout particulièrement des modèles à seuils capables par exemple de structurer une lecture des données au moyen de thèmes tels que "conjoncture basse-conjoncture haute" avec des dynamiques spécifiques au régime concerné, ou bien encore de modèles à "tunnels" pour lesquels, par exemple en raison de coûts de transaction, on va encadrer un régime central au sein duquel les arbitrages spéculatifs n'auront pas lieu par deux régimes où s'imposeront les relations théoriques d'équilibre. La classe de modèles concernés est constituée par les modèles *TARMA* qui peuvent être perçus comme une généralisation simple de la classe *ARMA* où la linéarité est seulement vérifiée par morceaux. Outre les changements de régimes, un apport supplémentaire de ces modèles à seuils est un enrichissement du traitement de la non stationnarité. Les processus peuvent maintenant être globalement stationnaires tout en présentant des épisodes de non stationnarité caractérisés par la présence de racine unitaire dans le polynôme

autorégressif caractéristique d'un des régimes possibles. Sur cette base de processus à seuils, une amélioration notable est fournie par le modèle *STAR* : alors qu'avec un *TARMA* le changement entre chaque régime s'effectue brutalement en fonction de la valeur prise par une variable de transition, on va ici autoriser une transition douce entre (au moins) deux régimes extrêmes. Un avantage est par exemple d'éviter une forte révision des anticipations à la suite d'une modification éventuellement minime des conditions initiales simplement parce que cette modification, même minime, est susceptible de signaler un changement de régime et donc une transformation qui peut être importante de la dynamique future du système étudié. Dans les modèles qui précèdent la variable de transition qui gouverne le régime en vigueur est supposée être observable or cette exigence est susceptible d'être fortement limitative. C'est naturellement trivialement le cas si la variable de transition ne peut être connue. Un autre cas, sans doute plus intéressant se fait jour si l'on suppose que la transition s'effectue selon un processus complexe mettant en jeu plusieurs variables. Dans ces conditions, plutôt que de réaliser une modélisation de cette transition pouvant être cause d'erreurs de spécification qui affectent potentiellement la qualité finale du modèle non linéaire, il peut être préférable de considérer que la variable de transition en question est inobservée. A condition de contraindre les propriétés du processus suivi par cette variable cachée il est alors possible de préciser empiriquement l'historique des régimes auxquels à été soumis une variable. Cette logique est tout particulièrement celle des processus à changement de régimes markoviens. Au total il apparaît donc que les modélisations non linéaires se caractérisent certes par leur richesse, notamment en termes d'interprétation et de compréhension de phénomènes "anormaux" au regard de la modélisation linéaire usuelle, mais en contrepartie l'utilisateur est placé devant une diversité de représentations concurrentes sans procédure de tri simple. Dans ces conditions il serait utile de disposer d'un modèle englobant qui admettrait les précédents comme cas particuliers. Une telle tentative est ainsi offerte par le processus *ACR* (*Autoregressive Conditional Root*) qui peut donc, dans certains cas, se révéler

particulièrement avantageux.

Dans la seconde partie, nous nous intéressons à la manière dont on construit une prévision dans le cadre d'un modèle non linéaire. De façon générale, on sait qu'il existe trois grandes méthodes de prévisions : prévisions ponctuelles, prévisions par intervalle de confiance et densité de prévision. Mais contrairement au cas des modèles linéaires, ces trois méthodes (et non plus uniquement la méthode de prévision ponctuelle) trouvent tout leur intérêt dans le cas des modèles non linéaires. En ce qui concerne les prévisions ponctuelles, deux cas doivent être considérés : pour différents modèles (modèles markoviens, modèles *TMA*) il existe une forme analytique aux prévisions ponctuelles comme dans le cas des modèles linéaires. En revanche, pour tous les autres modèles et en particulier pour les modèles à seuils (*TAR*, *STAR*) il n'existe pas de forme analytique à ces prévisions. Au moins cinq méthodes alternatives (méthode dite du skeleton, méthode Normal Forecast Error, méthode de l'estimation dynamique, simulation numérique de type Monte-Carlo ou Bootstrap) permettent alors de calculer une prévision ponctuelle dans ces modèles. En ce qui concerne les prévisions par intervalle de confiance, la présence d'asymétrie ou de multi-modalités des prévisions engendrées par ces modèles non linéaires peut conduire à des intervalles de prévision non nécessairement symétriques et non nécessairement continus de type *HDR* (*High Density Region*).

Dans la troisième partie, nous évoquons enfin les méthodes d'évaluation des prévisions issues de modèles non linéaires. Pour cela, nous distinguons l'évaluation des prévisions ponctuelles, des intervalles de confiance et des densités de prévision. En ce qui concerne l'évaluation des prévisions ponctuelles, au-delà de la comparaison simple des critères usuels (*RMSFE*, etc.), nous exposons les différents tests qui ont été proposés pour comparer les qualités prédictives de deux modèles concurrents (emboîtés ou non emboîtés) sur la base d'une fonction de perte associée aux prévisions ponctuelles. Nous présentons en outre les tests de prévision em-

boîtés qui peuvent être mis en œuvre dans ce contexte, mais aussi les tests non paramétriques de prévision de régime et les tests d'échec de prévision. Concernant l'évaluation des prévisions par intervalle de confiance, nous définissons les concepts de couverture non conditionnelle, de couverture conditionnelle et d'indépendance proposés par Christoffersen (1998) ainsi que les tests de ratio de vraisemblance associés à ces différentes notions.

Enfin, nous présentons les méthodes d'évaluation des densités de prévision en distinguant les tests de spécification correcte et les tests de comparaison de densités. Dans le premier cas, il s'agit de tester l'hypothèse nulle selon laquelle la densité de prévision obtenue à partir du modèle non-linéaire est compatible avec le processus générateur de données de la série étudiée. Dans le second cas, la logique est différente : il s'agit de choisir parmi un ensemble de modèles potentiellement mal spécifiés, le moins mauvais (ou le meilleur modèle) par rapport à un modèle de référence. Différents outils statistiques permettent d'établir cette comparaison non plus uniquement sur une prévision ponctuelle, mais sur la base de l'ensemble de la densité de prévision permettant de prendre en compte ainsi les éventuelles asymétries et multi-modalités.

Chapter 2

Les Modèles Univariés Non Linéaires

Dans cette section, nous présentons les principaux modèles non linéaires utilisés à des fins de prévision. Seule sera traitée ici la modélisation de l'espérance conditionnelle et en particulier nous n'aborderons pas ici les modèles non linéaires employés pour la modélisation du risque tels que les processus *GARCH* et leurs dérivés.

Il convient tout d'abord de rappeler la définition d'un processus linéaire dans la mesure où les principales modélisations non linéaires vont partir de cette représentation de base. Un processus purement stochastique est dit linéaire s'il peut s'écrire comme somme pondérée d'un processus en bruits blancs, soit :

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i u_{t-i} \quad (2.1)$$

où μ est une constante, ψ_i des coefficients réels avec $\psi_0 = 1$ et $\{u_t\}$ des variables aléatoires ayant même distribution, centrées, indépendantes et de variance $\sigma_u^2 < \infty$ et telles que $E[u_t | y_{t-1}, y_{t-2}, \dots] = 0$. Si de plus $\sum_{i=0}^{\infty} |\psi_i| < \infty$ alors les deux premiers moments de y_t sont indépendants du temps et y_t est dit alors faiblement stationnaire.

Dans cet univers des processus linéaires et stationnaires la modélisation de référence est celle des processus *ARMA*(p, q) popularisée par Box et

Jenkins d'écriture :

$$\phi(L)y_t = \theta(L)u_t \quad (2.2)$$

où L est l'opérateur de décalage défini par $L^j x_t = x_{t-j}$, $\phi(L)$ et $\theta(L)$ sont des polynômes en L de degrés respectifs p et q tels que leurs racines sont extérieures au cercle complexe unitaire ce qui assure pour $\phi(L)$ la stationnarité du processus¹ et pour $\theta(L)$ son inversibilité².

L'écriture (2.1) peut naturellement être perçue comme étant un cas particulier dérivée d'une structure dans laquelle y_t est simplement fonction des aléatoires u présente et passées :

$$y_t = f(u_t, u_{t-1}, u_{t-2}, \dots), \quad (2.3)$$

structure dans laquelle on impose la linéarité de la fonction f . Si l'on adopte cette vision, alors on comprend aisément que (2.1) revient à admettre qu'un développement de Taylor du premier ordre suffit pour approximer cette fonction inconnue f . Si tel n'est pas le cas, alors des ordres supérieurs à 1 doivent aussi être considérés et l'approximation linéaire n'est pas satisfaisante. Ce type de considération est bien connu et fonde par exemple les tests de linéarité de WHITE ou RAMSEY appliqués dans le cadre des régressions par moindres carrés. Ici, on peut par ce raisonnement retrouver la logique prévalant à la construction des modèles non linéaires que sont les processus bilinéaires proposés par Granger et Andersen (1978).

Une autre façon d'introduire la non linéarité est de considérer que la relation est linéaire par morceaux. C'est également une approche bien connue dans le cadre de la régression linéaire lorsqu'on pense qu'un événement daté sur l'échelle des temps a pu affecter les coefficients de la régression. Si l'on généralise cette idée à un autre type d'événements observables ou non, alors on dérive sur des processus non linéaires à seuils de type *TAR* (*Threshold*

¹Dans ce cas $\phi(L)$ est inversible et on retrouve une écriture de type (2.1) : $y_t = \phi(L)^{-1}\theta(L)u_t$

²Cette dernière condition assure que le processus possède une représentation autorégressive et rend son estimation possible

AutoRegression) ou à changements markoviens envisagés par Tong (1978, 1983) et Hamilton (1989).

Tous les modèles paramétriques précédents permettent d'arriver à une écriture utile pour la construction de prévisions qui possède la structure :

$$y_{t+1} = g(y_t, y_{t-1}, \dots, u_t, u_{t-1}, \Theta) + u_{t+1} \quad (2.4)$$

où Θ désigne un vecteur de paramètres de dimension finie. On peut déjà noter que le choix d'une spécification pour la fonction f entraîne en théorie celle de g . Pour autant la non linéarité entraîne généralement des difficultés calculatoires pour un horizon de prévision h quelconque si on convient, très classiquement, d'utiliser l'espérance conditionnelle en t de y_{t+h} comme anticipation de la future réalisation y_{t+h} . Ceci est naturellement lié au fait que $E[f(\cdot)] \neq f(E[\cdot])$ si f est non linéaire.

Précisons encore que l'on peut vouloir ne pas spécifier la fonction f . En ce cas on est amené à utiliser des méthodes d'estimation non paramétriques telles que les réseaux de neurones, les régressions à noyaux, les régressions locales (*LOESS* ou *LOWESS* par exemple), etc. En pratique ces techniques nécessitent un nombre d'observations qui dépasse généralement de beaucoup la taille des échantillons disponibles sur données macroéconomiques. Ceci peut expliquer que les études disponibles avec ces outils ne donnent pas, dans ce domaine d'application, de résultats particulièrement probants. Pour cette raison nous ne les présentons pas dans les développements qui suivent.

2.1 Les Processus Bilinéaires

Si on ne retient que certains des termes du second ordre dans le développement de Taylor de la fonction f alors on peut écrire :

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=0}^q \theta_j u_{t-j} + \sum_{i=1}^m \sum_{j=1}^s \beta_{ij} y_{t-i} u_{t-j} \quad (2.5)$$

où l'ensemble des coefficients $\phi_i, \theta_j, \beta_{ij}$ sont des réels et u_t des variables aléatoires indépendantes identiquement distribuées (*i.i.d.*) centrées homoscédastiques.

Dans ce cas y_t est dit suivre un processus bilinéaire, noté $BL(p, q, m, s)$. On remarque immédiatement que l'on retrouve bien le processus $ARMA(p, q)$ lorsque $\beta_{ij} = 0, i = 1, \dots, m, j = 0, \dots, s$. Ces processus sont linéaires en y et linéaires en u mais non linéaires sur le couple (y, u) . Ils apparaissent ainsi comme une généralisation simple des modèles $ARMA$ linéaires. Plus fondamentalement, leur intérêt vient aussi de ce qu'ils sont en mesure d'ajuster le comportement entrée-sortie de processus non linéaires complexes. En effet, les processus bilinéaires sont des formes simplifiées de séries de Volterra. Celles-ci peuvent être vues comme une généralisation du développement de Taylor avec prise en compte de la mémoire et sont en mesure d'approximer toute fonction continue non linéaire³.

Les propriétés de ces processus bilinéaires ont notamment été analysées par Granger et Andersen (1978), Subba Rao (1981), Subba Rao et Gabr (1984). Un de leurs intérêts est de pouvoir rendre compte avec une modélisation relativement simple de non linéarités complexes et de réponses asymétriques aux chocs. Par ailleurs ils peuvent être employés également pour ajuster des modèles à volatilité conditionnelle hétéroscédastique puisque plusieurs processus de type $ARCH$ ou $GARCH$ peuvent être réécrits sous forme de modèles bilinéaires (voir par exemple Kristensen, 2005)⁴. Des conditions de stationnarité et d'ergodicité ont été proposées pour certains modèles bilinéaires (voir notamment Berline et Francq (1990), Bhaskara Rao, Subba Rao et Walker (1983), Granger et Andersen (1978), Guégan et Pham (1987), Liu (1990, 1992), Liu et Brockwell (1988), Pham (1985)). En pratique ces conditions sont d'interprétation

³Sur ces aspects, voir par exemple Brockett (1976), Priestley (1991).

⁴En ce cas on fait plutôt référence à des modèles bilinéaires généralisés dans la mesure où ces processus ne sont pas nécessairement écrits sur le bruit blanc lui-même mais sur des fonctions de ce bruit. En pratique cela a pour conséquence que l'on peut facilement confondre modèle bilinéaire et modèle de type $GARCH$.

difficile sauf dans des cas relativement simples pris dans les sous-classes des modèles bilinéaires diagonaux et super-diagonaux⁵. Par exemple, avec le modèle $y_t = \phi y_{t-i} u_{t-j}$ diagonal ($i = j$) ou super-diagonal ($i > j$), la condition de stationnarité est $\phi^2 E[u_t^2] < 1$ et celle d'inversibilité est $\phi^2 E[u_t^2] < .5$ avec $E[u_t^2] < \infty$.

Dans les applications empiriques des formulations simples sont d'ailleurs souvent employées. Ainsi Granger et Newbold (1976), Granger et Andersen (1978), Tsay (1986) ou Tong (1990) utilisent un modèle diagonal $BL(1, 0, 1, 1)$, soit :

$$y_t = \phi_0 + \phi_1 y_{t-1} + \beta_{11} y_{t-1} u_{t-1} + u_t \quad (2.6)$$

Dans la pratique lorsqu'on autorise des modèles plus complexes la recherche des ordres (p, q, m, s) optimaux est généralement effectuée au moyen d'un critère de sélection et notamment le critère *AIC* d'Akaike.

Notons encore qu'une écriture bilinéaire a été utilisée par Peel et Davidson (1998) dans le cadre d'un modèle de type *VECM* avec prise en compte d'une relation de cointégration. Par ailleurs, des travaux récents envisagent le cas de modèles bilinéaires à coefficients variables (Bibi (2001), Bibi et Oyet (2002)) et périodiques (Bibi et Gauthierm 2005) généralisant ainsi les processus ARMA à coefficients variables. Enfin Subba Rao (1997) et Charemza, Lifshits et Makarova (2002) envisagent le cas de modèles bilinéaires à racine unitaire autorisant ainsi l'ajustement par ces modèles de séries non stationnaires⁶.

Enfin, on peut concevoir le modèle bilinéaire comme ajoutant à une représentation linéaire *ARMA*(p, q) des termes de la forme $y_{t-i} u_{t-j}$ pour

⁵Le processus est dit super diagonal lorsque $i > j$, diagonal lorsque $i = j$, et sous-diagonal lorsque $i < j$.

⁶Ces extensions ne sont d'ailleurs pas équivalentes : alors que le processus à racine unitaire défini par Subba Rao est, comme habituellement, stationnaire en différences premières; celui utilisé par Charemza et al. est non stationnaire en niveaux et en différences premières.

prendre en compte une éventuelle non linéarité. Dans cet esprit d'autres modélisations sont également envisageables. Ainsi le processus *AR exponentiel* de Ozaki (1985) d'écriture :

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \sum_{j=0}^r \beta_{ij} y_{t-i}^j \exp(-\gamma_i y_{t-i}^2) + u_t \quad (2.7)$$

ou encore le modèle *AR* semi-paramétrique (voir notamment Robinson (1983), Engle, Granger, Rice et Weiss (1986)) :

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + h(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + u_t \quad (2.8)$$

2.2 Les Processus TARMA

Dans une série de travaux pionniers, Tong (Tong (1978), Tong et Lim (1980), Tong (1983), Tong (1990)) envisage une linéarité par morceaux de la fonction f . En d'autres termes, dans cette perspective il existe plusieurs processus générateurs chacun s'appliquant à un sous-échantillon. Comme rappelé plus haut, si les dates auxquelles se produisent les sauts sont connues alors la modélisation peut s'effectuer simplement par prise en considération de variables indicatrices. On peut toutefois envisager que l'événement responsable du changement de régime n'est pas simplement observé sur l'échelle des temps mais qu'il peut être signalé par une autre variable, dite variable de transition, supposée exogène. Enfin, il est également possible de considérer que les événements construits sur la variable de transition dépendent de valeurs de paramètres inconnus. Par exemple, si la variable de transition x_t est simplement le temps, $x_t = t$, cela revient à poser que les dates de changement de régimes sont inconnues. Ainsi, dans un modèle à deux régimes, on aura :

$$y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} y_{t-i} + u_{1t} + \sum_{j=1}^{q_1} \theta_{1j} u_{1,t-j} & \text{si } x_{t-d} \leq c \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} y_{t-i} + u_{2t} + \sum_{j=1}^{q_2} \theta_{2j} u_{2,t-j} & \text{si } x_{t-d} > c \end{cases}$$

où u_{1t} et u_{2t} sont deux bruits blancs de variance respective σ_1^2 et σ_2^2 . L'entier d est le paramètre de délai et le réel inconnu c qui coupe le support de x en

deux sous-espaces, est un paramètre de seuil. Ainsi, lorsque x est inférieur à c , alors d périodes plus tard, la variable y sera générée par un certain processus $ARMA(p_1, q_1)$; en revanche, si x est plus élevé que c alors après d périodes y obéira à un autre processus $ARMA(p_2, q_2)$.

Ce processus à seuil (*Threshold*) à deux régimes se généralise à un nombre quelconque k de régimes et se note $TARMA(2; p_1, p_2; q_1, q_2)$. On a alors en considérant une variable de transition réelle :

$$y_t = \sum_{k=1}^K \left(\phi_{k0} + \sum_{i=1}^{p_k} \phi_{ki} y_{t-i} + u_{kt} + \sum_{j=1}^{q_k} \theta_{kj} u_{k,t-j} \right) \times \mathbb{I}(x_{t-d} \in R_i) \quad (2.9)$$

où u_{kt} est un bruit blanc de variance σ_k^2 pour $k = 1, \dots, K$. Les segments $R_i =]c_{i-1}, c_i]$ forment une partition de l'espace des réels conformément aux valeurs des paramètres de seuil c_i tels que $-\infty = c_0 < c_1 < \dots < c_K = +\infty$. Enfin $\mathbb{I}()$ est une fonction indicatrice prenant la valeur 1 si $x_{t-d} \in R_i$ et 0 sinon.

Le processus $TARMA(K; p_1, \dots, p_k; q_1, \dots, q_k)$ précédent est, dans la logique d'une linéarité par morceaux, l'extension la plus générale du modèle $ARMA$. En pratique seuls des cas simplifiés sont cependant considérés. Ainsi Tong (1983) considère le cas où la variable de transition est l'expliquée elle-même obtenant alors le modèle *Self-Exciting TARMA* (noté *SETARMA*) dans lequel il impose également l'absence de changement dans les processus en bruits blancs entre les régimes, soit $u_{kt} = u_t$. L'écriture générale de ce *SETARMA* ($K; p_1, \dots, p_K; q_1, \dots, q_K$) est donc :

$$y_t = \sum_{k=1}^K \left(\phi_{k0} + \sum_{i=1}^{p_k} \phi_{ki} y_{t-i} + u_t + \sum_{j=1}^{q_k} \theta_{kj} u_{t-j} \right) \times \mathbb{I}(y_{t-d} \in R_i) \quad (2.10)$$

La construction de prévisions à partir de cette structure est discutée notamment par Amendola et Niglio (2003) qui autorisent d'ailleurs une modification des variances des bruits selon les régimes. Des conditions de stationnarité et d'ergodicité ont été obtenues notamment par Brockwell,

Liu et Tweedie (1992) dans le cas particulier où la composante MA n'est pas soumis aux effets de seuils. Face aux difficultés rencontrées dans le cas général, la majorité des travaux se fonde sur des simplifications de ces structures générales pour ne retenir que des processus à seuils en moyennes mobiles (TMA) et surtout autorégressifs (TAR).

2.2.0.1 Les processus TMA

Un processus TMA est une structure composée uniquement de processus MA sur chacun des sous-ensembles disjoints définis par la variable de transition. Ainsi un $SETMA(q_1, q_2, d)$ à un seul changement de régime ($K = 1$) est défini par l'équation suivante :

$$y_t = \sum_{i=1}^{q_1} \delta_{1i} u_{t-i} + \sum_{j=1}^{q_2} \delta_{2j} \mathbb{I}(y_{t-d} \leq c) u_{t-j} + u_t \quad (2.11)$$

Le processus TMA partage avec le processus $MA(q)$ la propriété d'être toujours stationnaire sans qu'il soit besoin d'imposer de contraintes sur les coefficients (Liu et Susko, 1992). Comme pour les MA il est cependant nécessaire d'imposer des conditions d'inversibilité. A notre connaissance, seules des conditions suffisantes ont été énoncées par Ling et Tong (2005) pour un TMA quelconque, et des conditions nécessaire et suffisante pour un $TMA(1)$ par Ling, Tong et Li (2005)⁷. Le processus $SETMA$ est également analysé par De Gooijer (1998) qui propose en particulier un test de Lagrange de l'hypothèse MA versus $SETMA$ ⁸.

Ces processus engendrent naturellement des réponses asymétriques aux chocs : lorsque y_{t-d} est supérieur à c alors le poids de u_{t-i} dans la détermination de y_t est donné par δ_{1i} , en revanche ce poids est égal à $\delta_{1i} + \delta_{2i}$ lorsque y_{t-d} est inférieur ou égal à la valeur du seuil c .

⁷Soit $y_t = u_t + \{\delta_1 + \delta_2 \mathbb{I}(y_{t-1} \leq c)\} u_{t-1}$ un processus stationnaire et ergodique, avec $E|ln|y_t|| < \infty$ alors il est inversible si $|\delta_1|^{1-F_y(c)} |\delta_1 + \delta_2|^{F_y(c)} < 1$ où $F_y(\cdot)$ est la fonction de répartition de y_t . Il est non inversible si $|\delta_1|^{1-F_y(c)} |\delta_1 + \delta_2|^{F_y(c)} > 1$ et Ling, Tong et Li suspectent qu'il ne l'est pas lorsque la quantité précédente est égale à un.

⁸Sur ce point, voir également Ling et Tong (2005) qui offrent un test de quasi-rapport de vraisemblance.

Dans l'esprit, un processus proche est considéré par Wecker (1981) dans lequel deux filtres $MA(q)$ sont utilisés, l'un fondé sur les innovations passées positives et l'autre sur les innovations passées négatives. Ainsi, en définissant $u_t^+ = \max(0, u_t)$ et $u_t^- = \min(0, u_t)$, un processus $asMA(q)$ (*asymmetric MA*) est donné par :

$$y_t = \sum_{i=1}^q \delta_i^+ u_{t-i}^+ + \sum_{i=1}^q \delta_i^- u_{t-i}^- + u_t \quad (2.12)$$

soit encore :

$$y_t = \sum_{i=1}^q \delta_i^+ u_{t-i} + \sum_{i=1}^q \gamma_i \mathbb{I}(u_{t-i} \leq 0) u_{t-i} + u_t$$

avec $\gamma_i = \delta_i^- - \delta_i^+ \neq 0$ pour au moins un i . Un test de linéarité versus une réponse asymétrique au sein d'une telle structure a été proposé par Brännäs, De Gooijer et Teräsvirta (1998).

2.2.0.2 Les processus TAR

Toutefois, dans cette classe de modèles, les travaux ont dans leur très grande majorité porté sur la sous-classe des processus TAR et tout particulièrement celle des $SETAR$ certainement en raison de moindres difficultés d'estimation. Contrairement aux TMA , ces modèles ne retiennent que des écritures AR sur les sous-ensembles disjoints définis par la variable de transition. Ainsi, un $SETAR$ a un seul changement de régime aura pour écriture :

$$y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} y_{t-i} + u_{1t} & \text{si } y_{t-d} \leq c \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} y_{t-i} + u_{2t} & \text{si } y_{t-d} > c \end{cases} \quad (2.13)$$

et plus généralement, un $SETAR(K; p_1, \dots, p_k; d)$ s'écrira :

$$y_t = \sum_{k=1}^K \left(\phi_{k0} + \sum_{i=1}^{p_k} \phi_{ki} y_{t-i} + u_{kt} \right) \times \mathbb{I}(y_{t-d} \in R_i) \quad (2.14)$$

où, comme précédemment, les $R_i =]c_{i-1}, c_i]$ sont des segments de la droite réelle dont les bornes sont données par les valeurs des paramètres de seuil

c_i ordonnées de sorte que $-\infty = c_0 < c_1 < \dots < c_K = +\infty$.

Des conditions suffisantes de stationnarité des *SETAR* ont été données mais ces conditions apparaissent comme très restrictives lorsqu'on les compare aux conditions nécessaires et suffisantes qui sont connues pour des cas particuliers (Chang (1993), De Gooijer et De Bruin (1998), Tong(1990)). A titre d'illustration, on peut reprendre un exemple donné par Terasvirta (2005). Avec une structure *SETAR*($K; p; d$), où tous les *AR* sont du même ordre p , Chang (1993) donne comme condition suffisante de stationnarité $\max_k \sum_i^p |\phi_{ki}| < 1$ et lorsque $p = 1$, cette condition devient $\max_k |\phi_{k1}| < 1$, ce qui équivaut à imposer la stationnarité de chacun des K processus *AR*(1). Or, dans ce cas particulier les conditions à la fois nécessaire et suffisante dans un modèle à deux régimes sont connues et correspondent aux contraintes $\phi_{11} < 1$, $\phi_{21} < 1$ et $\phi_{11}\phi_{21} < 1$. De toute évidence, la condition suffisante de Chang est alors beaucoup trop restrictive et il est par exemple possible de construire un *SETAR* stationnaire dans lequel un des régimes est gouverné par un *AR* non stationnaire. Considérons par exemple le *SETAR*(2; 1; 1) suivant :

$$y_t = \begin{cases} -2 y_{t-1} + u_t & \text{si } y_{t-1} \leq 0 \\ 0.6 y_{t-1} + u_t & \text{si } y_{t-1} > 0 \end{cases}$$

C'est un processus stationnaire alors même que l'un des régimes est gouverné par un *AR* explosif. Du fait de ce coefficient $\phi_{11} = -2$, y_t sera plus souvent positif que négatif de sorte que $E[y_t] > 0$. Ainsi, l'espérance de ce processus sans terme constant est non nulle. Par contraste, on rappelle qu'un *AR* est d'espérance non nulle précisément si et seulement si son équation fait apparaître un terme constant non nul.

Un autre cas de figure souvent employé est celui d'un *SETAR* à 3 régimes dans lequel le processus central possède une racine unitaire. Lorsqu'on se limite à des processus autorégressif d'ordre 1, cela revient à imposer que dans le régime intermédiaire la variable obéit à une marche

au hasard. On a donc une structure du type suivant :

$$y_t = \begin{cases} \phi_1 y_{t-1} + u_t & \text{si } y_{t-1} \leq c_1 \\ y_{t-1} + u_t & \text{si } c_1 < y_{t-1} \leq c_2 \\ \phi_2 y_{t-1} + u_t & \text{si } y_{t-1} > c_2 \end{cases} \quad (2.15)$$

Si $|\phi_1| < 1$ et $|\phi_2| < 1$ alors le processus est globalement ergodique ce qui implique l'existence d'une distribution stationnaire pour y_t (Tong, 1990) même si la série est localement non-stationnaire. Ce modèle a notamment été employé en finance où, lorsque l'on prend en compte des coûts de transaction, les arbitrages ne se réalisent qu'à la condition que les déviations des rendements soient suffisantes et on observe alors un phénomène de retour à la moyenne (ici zéro) du fait des deux processus *AR* stationnaires en vigueur dans les deux régimes extrêmes. Dans ce système le régime central est qualifié de corridor *inactif* ce qui signifie qu'il ne nécessite l'estimation d'aucun paramètre (pour une analyse du système à corridor inactif, voir notamment Kapetanios et Shin (2006)). Un système plus général où chacun des trois régimes est gouverné par un *AR(p)* a été notamment étudié et estimé par Bec, Ben Salem et Carrasco (2004).

On notera que la considération de processus de type SETAR avec racine unitaire donne lieu à des extensions particulièrement intéressantes généralisant la théorie de la cointégration initiée par Engle et Granger (1987) et Johansen (1988, 1991, 1996) qui suppose la linéarité et la symétrie des ajustements⁹. Dans la logique précédente si on considère que y_t est une équation d'équilibre de long terme, alors la cointégration à seuils correspond à un SETAR à trois régimes globalement stationnaire dans lequel le régime

⁹Par exemple l'approche en deux étapes d'Engle-Granger considère l'équation linéaire $x_{1t} = \beta_0 + \beta_2 x_{2t} + \dots + \beta_m x_{mt} + v_t$ où $x_{it} \sim I(1)$ et v_t peut être autocorrélé. La seconde étape consiste en un test de racine unitaire à la Dickey-Fuller sur $\Delta v_t = b v_{t-1} + \epsilon_t$ éventuellement augmentée de décalages $\Delta v_{t-1}, \Delta v_{t-2}, \dots$. Le rejet de l'hypothèse nulle implique que $v_t \sim I(0)$ et dans ce cas le théorème de représentation de Granger assure l'existence d'une écriture de type modèle à correction d'erreur de la forme $\Delta x_{1t} = \alpha(x_{1t} - \beta_0 - \beta_2 x_{2t} - \dots - \beta_m x_{mt}) + \nu_{1t}$, éventuellement augmentée de décalages Δx_{it} . Ainsi les déviations positives et négatives de x_{1t} autour de sa position d'équilibre de long terme donnée par $\beta_0 + \beta_2 x_{2t} + \dots + \beta_m x_{mt}$ sont affectées du même coefficient α .

intermédiaire possède une racine unitaire (Balke et Fomby¹⁰, 1997). Dans ce cadre des réponses asymétriques aux chocs sont possibles, l'asymétrie pouvant là encore être provoquée par l'existence de coûts de transaction comme dans l'étude portant sur les taux d'intérêt réalisée par Anderson (1997). Elle peut, toujours sur ce même thème, être également justifiée par un comportement différencié de la banque centrale qui serait plus réactive lors des périodes où les anticipations d'inflation s'élèvent et plus attentiste en cas de baisse de ces anticipations¹¹.

Dans un modèle *SETAR* la variable de transition est constituée par le niveau décalé de l'expliquée, y_{t-d} . Enders et Granger (1998), Enders et Siklos (2001), Caner et Hansen (2001) ont suggéré d'utiliser également la variation décalée $\Delta y_{t-d} = y_t - y_{t-d}$. Un argument statistique est avancé : si le processus gouvernant y tout en étant stationnaire possède une racine proche de l'unité (*near unit root process*) alors il peut être préférable de considérer la série en différence. Indépendamment de cet argument, il est possible que la variable étudiée soit-elle que les agents font particulièrement attention à l'amplitude des variations plutôt qu'au niveau lui-même. Ce peut par exemple être le cas pour les taux d'intérêt, les taux de change et, plus généralement, les rendements des actifs financiers. Bohl et Siklos (2004) préconisent aussi l'emploi de ce modèle pour la détection des bulles sur les marchés financiers. Un tel modèle, qualifié de *MTAR* (*Momentum Threshold AutoRegression*) a donc pour écriture :

$$y_t = \sum_{k=1}^K \left(\phi_{k0} + \sum_{i=1}^{p_k} \phi_{ki} y_{t-i} + u_{kt} \right) \mathbb{I}(\Delta y_{t-d} \in R_i) \quad (2.16)$$

¹⁰Dans cette approche une procédure en deux étapes est implémentée : test standard de cointégration ou de présence d'une racine unitaire dans la première puis, en cas de rejet de l'hypothèse nulle, test de présence d'une non-linéarité à seuil. La puissance des tests de Dickey-Fuller étant affectée par la non linéarité, des tests évitant cette procédure en deux étapes ont été également proposés (Bec, Ben Salem et Carrasco (2004), Caner et Hansen (2001), Enders et Granger (1998), Kapetanios et Shin (2006), Enders et Siklos (2001), Seo (2005))

¹¹Sur ces aspects, voir par exemple l'étude récente de Haug et Siklos (2005). Pour des applications sur les changes, voir par exemple Taylor et Peel (2000), Taylor, Peel et Sarno(2001)

La recherche de la spécification optimale, pour un nombre de régimes fixé a priori, s'effectue généralement au moyen de critères de sélection de type *AIC* ou *BIC* adaptés par Wong et LI (1998) dont les valeurs sont données par les équations suivantes :

$$AIC(p_1, \dots, p_K) = \sum_{k=1}^K [T_k \log \hat{\sigma}_k^2 + 2(p_k + 2)] \quad (2.17)$$

$$AIC_c(p_1, \dots, p_K) = \sum_{k=1}^K \left[T_k \log \hat{\sigma}_k^2 + \frac{T_k(T_k + p_k + 1)}{T_k - p_k - 3} \right] \quad (2.18)$$

$$BIC(p_1, \dots, p_K) = \sum_{k=1}^K [T_k \log \hat{\sigma}_k^2 + (p_k + 1) \log T_k] \quad (2.19)$$

Des simulations de Monte Carlo réalisées par ces auteurs paraissent indiquer que le critère *BIC* est préférable sur grands échantillons alors que sur petits échantillons l'emploi du critère d'Akaike corrigé *AICc* serait préférable. On trouve également des critères de validation de choix des paramètres d'un *SETAR* dans De Gooijer (2001) et Galeano et Peña (2004).

2.3 Les Processus *STAR*

Un des reproches faits à l'encore des modèles de type *TAR* concerne la brutalité de la transition entre les régimes pouvant notamment mettre en doute la crédibilité des prévisions qui en sont issues. Si l'on suppose par exemple dans un *SETAR* à deux régimes que y_{t-d} est très légèrement inférieur à la valeur du seuil c , mais que l'observation y_{t+1-d} soit légèrement supérieur à c alors la prévision en $t+1$ sera tirée d'une équation différente de celle utilisée en t de sorte qu'une forte révision de ces prévisions n'est pas impossible alors que les conditions initiales des deux configurations sont proches. Par ailleurs la discontinuité au niveau du seuil complique l'étape d'estimation. Afin de pallier à ces difficultés, Chang et Tong (1986) ont proposé de retenir

une transition *douce* entre les différents régimes définissant ainsi la classes des processus *STAR* (*Smooth Transition Autoregressions*).

Dans le cas simple caractérisé par une seule fonction de transition qui vient pondérer les équations *AR* afférentes à deux dynamiques différentes dans lesquelles on suppose pour simplifier l'égalité des variances des résidus, on obtient donc l'écriture suivante :

$$y_t = \left(\phi_{10} + \sum_{i=1}^p \phi_{1i} y_{t-i} \right) \times [1 - G(x_t; \theta)] + \left(\phi_{20} + \sum_{i=1}^p \phi_{2i} y_{t-i} \right) \times G(x_t; \theta) + u_t \quad (2.20)$$

où $G()$ est une fonction continue de paramètres θ telle que $0 \leq G() \leq 1$ et x_t une variable de transition stationnaire.

Comme le font remarquer van Dijk, Teräsvirta et Franses (2000), un modèle *STAR* peut recevoir deux interprétations possibles : soit celle d'un modèle à seuil à deux régimes, chacun étant associé aux valeurs extrêmes de la fonction de transition, $G(x_t, \theta) = 0$ et $G(x_t, \theta) = 1$, avec une transition continue de l'un des régimes vers l'autre. Soit celle d'un *continuum* de régimes, celui en vigueur étant déterminé par une valeur particulière de la fonction de transition et donc de la variable de transition x_t .

L'équation (2.20) peut être ré-écrite de manière équivalente comme :

$$y_t = \beta_{0t} + \beta_{1t} y_{t-1} + \beta_{2t} y_{t-2} + \dots + \beta_{pt} y_{t-p} + u_t \quad (2.21)$$

avec $\beta_{it} = \phi_{1,i} \times [1 - G(x_t; \theta)] + \phi_{2,i} \times G(x_t; \theta)$. En d'autres termes, et ainsi que l'ont souligné Lundbergh et Teräsvirta (2002), le modèle *STAR* peu être également vu comme un *AR* à coefficients variables. En conséquence le modèle *STAR* est localement linéaire lorsque la variable de transition x_t est constante.

Dans l'immense majorité des travaux le choix de la fonction de transition s'effectue entre deux formulations :

- Soit avec une fonction de type logistique, pour donner le modèle *LSTAR* (*Logistic STAR*), qui dans sa version la plus simple a pour expression :

$$G(x_t; \theta) = L(x_t; \gamma, c) = [1 + \exp(-\gamma(x_t - c))]^{-1}, \quad \gamma > 0 \quad (2.22)$$

- Soit avec une fonction de type exponentielle pour donner le modèle *ESTAR* (*Exponential STAR*),

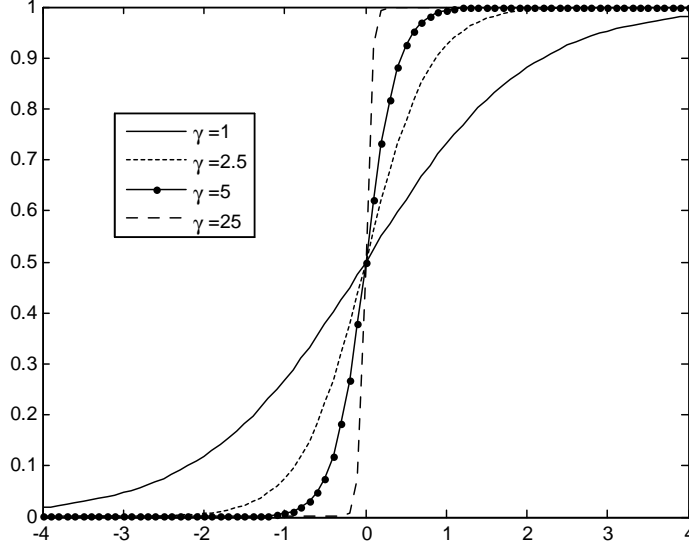
$$G(x_t; \theta) = E(x_t; \gamma, c) = 1 - \exp[-\gamma(x_t - c)^2], \quad \gamma > 0 \quad (2.23)$$

Comme le montre la Figure (2.1), dans un *LSTAR*, la fonction de transition est monotone croissante en x_t avec $L(c; \gamma, c) = 0.5$. Le paramètre de pente γ détermine le rythme de la transition d'un régime vers l'autre. Lorsque γ tend vers 0 alors la fonction $L(x_t; \gamma, c)$ tend vers 0.5 et le *SETAR* dégénère en un *AR* linéaire¹². A l'autre extrémité du domaine de définition, lorsque γ tend vers $+\infty$, alors la fonction de transition $L(x_t; \gamma, c)$ tend vers 1 si $x_t > c$, et tend vers 0 si $x_t < c$. En d'autres termes le processus *TAR* a deux régimes est un cas particulier du *LSTAR* obtenu avec des valeurs élevées de la pente de la fonction de transition¹³. Compte-tenu de ces caractéristiques, on conçoit que le processus *LSTAR* avec une seule transition soit intéressant pour traiter des variables pouvant obéir à deux dynamiques différentes dans des régimes extrêmes avec un comportement asymétrique en fonction de l'écart à l'équilibre. L'exemple type est celui de la modélisation des cycles d'activité avec alternance de phases de récession et d'expansion.

Le comportement de la fonction de transition du processus *ESTAR*, toujours dans le cas simple à deux régimes extrêmes et une seule valeur pour le paramètre de seuil c , est sensiblement différent du cas *LSTAR* comme le montre la Figure (2.2). Du fait de la forme quadratique retenue,

¹²Cela se voit aisément à partir de l'écriture du *LSTAR* comme *AR* à coefficients variables. En reprenant l'équation (2.21), il vient en effet $\beta_{it} \rightarrow \beta_i = (\phi_{1,i} + \phi_{2,i})/2$ lorsque $\gamma \rightarrow 0$.

¹³En considérant à nouveau l'écriture (2.21), lorsque $\gamma \rightarrow \infty$ on vérifie que $\beta_{it} \rightarrow \phi_{2,i}$ si $x_t > c$; et $\beta_{it} \rightarrow \phi_{1,i}$ si $x_t < c$.

Fig. 2.1 Fonction de Transition d'un modèle LSTAR $L(x_t; \gamma, c)$ pour $c = 0$ 

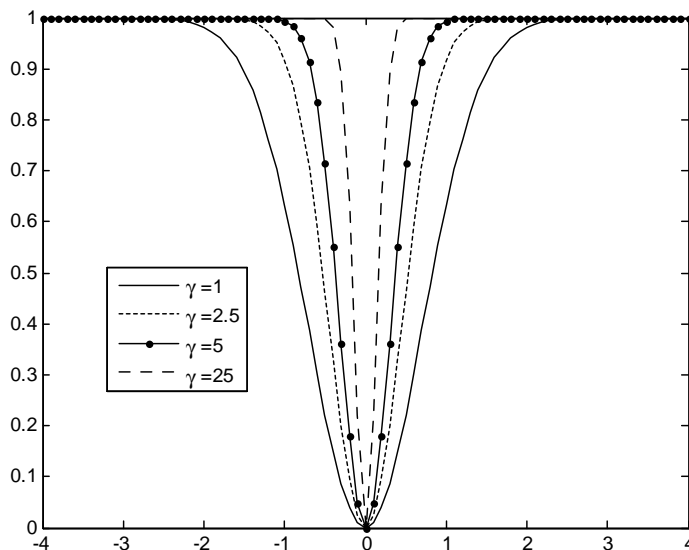
seule la distance entre la variable de transition, x_t , et le paramètre de seuil, c , va intervenir sans considération de leurs positions relatives, puisque :

$$E(c; \gamma, c) = E(x_t; 0, c) = 0 \quad (2.24)$$

$$\lim_{x \rightarrow \pm\infty} E(x; \gamma, c \mid \gamma \neq 0) = \lim_{\gamma \rightarrow \infty} E(x_t; \gamma, c \mid x_t \neq c) = 1 \quad (2.25)$$

Ainsi, en fonction de la distance $x_t - c$, la transition à une forme en U : un régime extrême correspond au cas où $x_t = c$ et l'autre au cas où x_t tend vers $+\infty$ ou $-\infty$. De ce fait, le processus *ESTAR* est plutôt adapté à l'ajustement de systèmes possédant un corridor central tel que vu plus haut avec les *SETAR* à trois régimes. Notons cependant que le processus *ESTAR* ne dégénère pas en un *TAR* à trois régimes selon les valeurs de γ puisqu'aux limites de son espace de définition le modèle devient linéaire¹⁴.

¹⁴Si ce cas particulier doit être couvert, Jansen et Teräsvirta (1996) suggère de retenir un *LSTAR* du second ordre, c'est à dire un processus avec une fonction de transition définie sur deux paramètres de seuil c_1 et c_2 , avec $c_1 < c_2$, d'écriture

Fig. 2.2 Fonction de Transition d'un Modèle ESTAR $E(x_t; \gamma, c)$ pour $c = 0$ 

Les processus précédents se généralisent au moins en théorie aisément. Ainsi, un processus *LSTAR* d'ordre n aura pour fonction de transition :

$$LSTAR(x_t; \gamma, c_1, c_2, \dots, c_n) = \left[1 + \exp \left(-\gamma \prod_{i=1}^n (x_t - c_i) \right) \right]^{-1}, \quad \gamma > 0 \quad (2.26)$$

Van Dijk et Franses (1999) présente également un processus à plusieurs fonctions de transition faisant éventuellement chacune référence à des variables de transition différentes pour obtenir un modèle *MRSTAR* (*Multiple Regime STAR*). Ce modèle possède la particularité, au même titre que les

$LSTAR(x_t; \gamma, c_1, c_2) = (1 + \exp \{-\gamma(x_t - c_1)(x_t - c_2)\})^{-1}$. Ici, $LSTAR(x_t; 0, c_1, c_2) = 1/2$ et $\lim_{\gamma \rightarrow \infty} LSTAR(x_t; \gamma, c_1, c_2 | x_t < c_1) = \lim_{\gamma \rightarrow \infty} LSTAR(x_t; \gamma, c_1, c_2 | c_2 < x_t) = 1$, et $\lim_{\gamma \rightarrow \infty} LSTAR(x_t; \gamma, c_1, c_2 | c_1 < x_t < c_2) = 0$: il y a bien maintenant un basculement entre les deux équations selon la position de la variable de transition avec toutefois une contrainte par rapport au *SETAR* puisque les équations des deux régimes extrêmes sont identiques.

réseaux de neurones, de pouvoir approximer toute fonction continue non linéaire.

Enfin des travaux récents ont proposé des processus à transition douce avec mémoire longue : dans un modèle *FISTAR*, la variable modélisée est la différence fractionnaire de y_t , soit :

$$(1-L)^d y_t = \left(\phi_{10} + \sum_{i=1}^p \phi_{1i} (1-L)^d y_{t-i} \right) \times [1 - G(x_t; \theta)] \\ + \left(\phi_{20} + \sum_{i=1}^p \phi_{2i} (1-L)^d y_{t-i} \right) \times G(x_t; \theta) + u_t \quad (2.27)$$

Un tel processus est étudié par van Dijk, Franses et Paap (2002) avec une fonction de transition logistique, et par Smallwood (2005) au moyen d'une fonction exponentielle. Notons cependant que la possibilité de distinguer les deux aspects que sont la mémoire longue et les changements structurels peut être douteuse. Ainsi Diebold et Inoue (1999) soulignent qu'ils sont "*intimement liés et peuvent être aisément confondus*" (p.22). Ils montrent d'ailleurs que dans des cas particuliers construits sur des processus de markov, les changements structurels et la mémoire longue sont deux appellations différentes qui recouvrent en fait un même phénomène. Sur ces aspects voir encore Guegan (2005), Guegan et Rioublanc (2005) ainsi que Kapetanios (2004), ce dernier considérant également un processus *TAR*.

La démarche à suivre pour l'identification, l'estimation et la mise en oeuvre de tests de validation des modèles *STAR* est notamment présentée par Teräsvirta (1994, 1998) et van Dijk, D., Teräsvirta, T. et Franses (2000). Lorsque la fonction de transition et la variable de transition ont été choisies alors les estimateurs des moindres carrés non linéaires des paramètres peuvent être obtenus. Une procédure en deux étapes est généralement employée : pour des valeurs données de la pente et des paramètres de transition, on trouve les estimateurs des *MCO* des paramètres ϕ . Ces derniers étant alors fixés, on recherche les estimateurs des moindres carrés non linéaires

des premiers. Le processus est itéré jusqu'à ce qu'un critère de convergence soit satisfait. Les propriétés de ces estimateurs et leurs caractéristiques empiriques révélées par des simulations de Monte Carlo ont été étudiées notamment par Chan et McAleer (2002).

2.4 Les Processus à Changement de Régimes Markovien

Dans les modèles de type *TAR* la variable de transition est observée mais on peut imaginer faire dépendre le régime en vigueur d'une variable d'état inobservée s_t . En faveur de cette approche, on peut imaginer que le changement de régime est le résultat d'un ensemble de causes ce qui, dans le cadre d'un modèle *TAR*, obligerait à prendre en compte plusieurs variables de transition et éventuellement contraindrait à la modélisation et à l'estimation d'un système d'équations. La référence à une variable d'état offre alors une description alternative parcimonieuse.

Dans un article fondateur, Hamilton (1989) envisage ainsi un processus autorégressif d'ordre 4 sur le taux de croissance trimestriel du PNB réel américain de la forme :

$$\Delta y_t - \mu(s_t) = \phi_1(\Delta y_{t-1} - \mu(s_{t-1})) + \dots + \phi_4(\Delta y_{t-4} - \mu(s_{t-4})) + u_t \quad (2.28)$$

où u_t est un bruit blanc gaussien de variance constante σ_u^2 , et donc ici seule l'espérance est autorisée à changer. La variable latente $\mu(s_t)$ est supposée obéir à un processus de Markov ergodique en temps discret à deux états de la forme $s_t = 0$ si on est en expansion à la date t , $s_t = 1$ si on est en récession. Avec, par exemple, un taux de croissance moyen positif en expansion et négatif en récession, on aura ainsi :

$$\mu(s_t) = \begin{cases} \mu_0 > 0 & \text{si } s_t = 0 \\ \mu_1 < 0 & \text{si } s_t = 1 \end{cases} \quad (2.29)$$

Le processus stochastique générant l'état inobservable est défini par un

ensemble de probabilités de transition constantes :

$$p_{ij} = \Pr(s_{t+1} = j \mid s_t = i, s_{t-1} = k, \dots) = \Pr(s_{t+1} = j \mid s_t = i), \quad (2.30)$$

avec $p_{ij} + p_{ii} = 1 \forall i, j, k \in \{0, 1\}$ ¹⁵. Ainsi, la probabilité que le système soit dans l'état j en $t + 1$ ne dépend que de l'état dans lequel il se trouvait à la période immédiatement précédente. En d'autres termes, le futur du processus ne dépend que de son état actuel et est indépendant de son passé.

Cette structure de base peut naturellement être amendée par exemple en considérant plus de deux états. La variance du bruit peut aussi être dépendante du régime, soit $\sigma_{u_t}^2 = \sigma^2(s_t)$. Il est également possible de lever la contrainte de constance des probabilités de transition. En effet, la probabilité de sortir d'un épisode de récession ou d'expansion peut se modifier en fonction de la durée de celui-ci (Diebold, Rudebush et Sichel (1993), Diebold, Lee et Weinbach (1994)). Par ailleurs, avec une équation telle que (2.28), le changement d'état provoque un saut immédiat de la série vers son nouveau niveau. Krolzig (1997)¹⁶ montre qu'un ajustement graduel du niveau de la série peut être pris en compte par une structure autorégressive avec un terme constant fonction de cette variable d'état :

$$\Delta y_t = \phi_0(s_t) + \phi_1 \Delta y_{t-1} + \dots + \phi_p \Delta y_{t-p} + u_t \quad (2.31)$$

Il est également possible de faire dépendre les valeurs des coefficients autorégressifs du régime en vigueur indiqué par s_t . En ce cas, il peut arriver que la racine d'un des polynômes autorégressifs soit unitaire, voire explosive, alors que le processus est globalement stationnaire. Comme avec un SETAR, le modèle autorégressif à saut de Markov est ainsi en mesure de

¹⁵Si on note P la matrice constituée des probabilités de transitions, soit pour un système à E états possibles, $P = \{p_{ij}\}$ pour $i, j = 1, 2, \dots, E$, alors cette matrice a par construction une valeur propre égale à l'unité. L'ergodicité de s_t exige que ses autres valeurs propres soient de modules inférieurs à l'unité.

¹⁶Par ailleurs Krolzig généralise le modèle aux processus multivariés VAR éventuellement cointégrés pour définir une typologie de modèles : MSM-VAR (Markov Switching VAR avec espérances non constantes), MSI-VAR (Markov Switching VAR avec termes constants sujets à sauts), MSMH-VAR (MSM-VAR avec hétéroscédasticité),... Dans cette terminologie, le modèle de Hamilton est un MSM(2)-AR(4), *i.e.* un processus de Markov à saut à 2 régimes couplé à une écriture autorégressive d'ordre 4.

représenter des alternances d'historiques stationnaires et non stationnaires (Hall, Psaradakis et Sola, 1999).

L'estimation du maximum de vraisemblance de ces processus fait appel à l'algorithme EM (*Expectation Maximization*) proposé par Demster, Laird et Rubin (1977) qui s'applique précisément lorsque la trajectoire observée est fonction de variables aléatoires inobservées¹⁷. Les propriétés des estimateurs sont discutées notamment par Douc, Moulines et Ryden (2004). Un des intérêts des applications empiriques est qu'il est possible d'obtenir les estimateurs des probabilités associées aux régimes inobservés conditionnellement à l'ensemble d'information disponible et donc de fournir une indication sur l'historique de ces régimes. On distingue alors trois types de probabilité : la probabilité associée au filtrage (*filtered regime probabilities*), $\Pr[s_t|Y_t]$, la probabilité lissée (*smoothing regime probabilities*), $\Pr[s_t|Y_T]$, et les probabilités prévues (*predicted regime probabilities*), $\Pr[s_t|Y_{t-1}]$, où Y_t désigne l'historique des réalisations de la variable y jusqu'à la date t . Il est ainsi possible de comparer les enseignements du modèle avec des informations tierces, l'exemple type étant la datation des cycles US du NBER.

2.5 Le Processus *ACR*

Récemment, Gouriéroux et Robert (2001), Rahbek et Shephard (2002), Bec, Rahbek et Shephard (2005) ont proposé un modèle capable de reproduire des épisodes de non-stationnarité au moyen d'un processus géométriquement ergodique et globalement stationnaire. Un des intérêts de cette modélisation *ACR* (*Autoregressive Conditional Root*) est d'offrir une écriture englobante des processus *SETAR*, *STAR* et autorégressif à changement de régime markovien.

¹⁷Cet algorithme appliqué au processus à sauts markoviens est exposé en détail par Hamilton (1990) et Krolzig (1997).

Le plus simple des processus *ACR* est l'*ACR*(1) d'expression :

$$y_t = \rho^{s_t} y_{t-1} + u_t \quad (2.32)$$

où s_t est une indicatrice binaire prenant ses valeurs dans l'ensemble $\{0, 1\}$, ρ un coefficient réel et u_t un bruit blanc gaussien. Cette équation est couplée à une hypothèse relative à la probabilité de l'état décrit par s_t qui est supposée ne dépendre que de la valeur passée de la variable d'intérêt y :

$$\Pr(s_t = 1 \mid y_{t-1}, u_t) = \Pr(s_t = 1 \mid y_{t-1}) = P(y_{t-1}) \quad (2.33)$$

Ainsi y se comporte comme une marche au hasard si $s_t = 0$ et comme un *AR*(1) stationnaire si $s_t = 1$ et $|\rho| < 1$. La seule exigence est que la fonction $P(y_t)$ tende vers l'unité lorsque $|y_t|$ tend vers l'infini.

Il est aisé de voir qu'un cas particulier de processus *SETAR* à 3 régimes peut être dérivé de cette structure. En effet, en posant :

$$P(y_t) = \begin{cases} 1 & \text{si } |y_t| > c > 0 \\ 0 & \text{sinon} \end{cases} \quad (2.34)$$

il vient¹⁸ :

$$y_t = \begin{cases} \rho y_{t-1} + u_t & \text{si } |y_t| > c > 0 \\ y_{t-1} + u_t & \text{sinon} \end{cases} \quad (2.35)$$

Le rapport avec les processus à transition douce *STAR* peut être illustré en considérant une spécification de type logistique sur la probabilité prévisionnelle :

$$\log \left[\frac{P(y_t)}{1 - P(y_t)} \right] = \delta_0 + \delta_1 f(y_t) \quad (2.36)$$

où $f()$ est une fonction croissante avec $|y_t|$. Comme d'après (2.32), on a :

$$\Delta y_t = (\rho^{s_t} - 1)y_{t-1} + u_t = s_t(\rho - 1)y_{t-1} + u_t = s_t b y_{t-1} + u_t \quad (2.37)$$

¹⁸Il s'agit d'un cas particulier dans la mesure où les deux régimes extrêmes stationnaires sont gouvernés par la même équation. Dans un *SETAR* on peut faire apparaître deux coefficients différents sur chacune de ces équations.

avec $b = \rho - 1$, il vient alors :

$$E[\Delta y_t \mid y_{t-1}] = b \times \left\{ \frac{\exp[\delta_0 + \delta_1 f(y_{t-1})]}{1 + \exp[\delta_0 + \delta_1 f(y_{t-1})]} \right\} \times y_{t-1} \quad (2.38)$$

et finalement on retrouve sur les variations de y_t une expression typique du modèle *STAR* :

$$\Delta y_t = b \times \left\{ \frac{\exp[\delta_0 + \delta_1 f(y_{t-1})]}{1 + \exp[\delta_0 + \delta_1 f(y_{t-1})]} \right\} \times y_{t-1} + \epsilon_t \quad (2.39)$$

où ϵ_t est un bruit à éléments indépendants. Rappelons cependant que contrairement au *STAR*, un processus *ACR* est en mesure de connaître des périodes de non stationnarité.

Enfin, il suffit de considérer l'écriture (2.33) pour noter la proximité de l'*ACR* avec les processus autorégressifs à sauts markoviens présentés auparavant. Une différence fondamentale est que la probabilité associée à un état de la variable s_t est une fonction explicite de la variable y_{t-1} .

Dans sa forme la plus générale, le modèle *ACR* est un processus vectoriel tel que :

$$y_t = s_t(\phi_{11}y_{t-1} + \dots + \phi_{1k}y_{t-k}) + (1-s_t)(\phi_{21}y_{t-1} + \dots + \phi_{2k}y_{t-k}) + u_t \quad (2.40)$$

où $y_t = (y_{1t}, \dots, y_{nt})'$ et $u_t = (u_{1t}, \dots, u_{nt})'$ un processus en bruits blancs. Comme dans le cas particulier précédent, l'une des deux matrices $\phi_1 = (\phi_{11}, \dots, \phi_{1k})$ et $\phi_2 = (\phi_{21}, \dots, \phi_{2k})$ peut avoir des racines unitaires alors même que y_t est globalement stationnaire. Les hypothèses nécessaires à ce résultat ainsi qu'une discussion des estimateurs du maximum de vraisemblance sont présentées dans Rahbek et Shephard (2002). Une application univariée au taux de change franc-mark est réalisée par Bec, Rahbek et Shephard (2005). Les auteurs discutent notamment des probabilités conditionnelles des deux régimes tirées respectivement d'un *ACR* et d'un *SETAR*. En particulier il n'est pas fait état de résultats relatifs à une utilisation à des fins de prévision de l'*ACR*.

Chapter 3

Comment Construire une Prévision dans un Modèle Non Linéaire ?

De façon générale, que ce soit dans un modèle linéaire ou non linéaire, il existe trois principales formes de prévision : les prévisions ponctuelles, les intervalles de confiance et la densité de prévision. Dans le cas des modèles linéaires de type ARMA, seules les deux premières formes de prévisions sont généralement utilisées. Les intervalles de confiance supposés symétriques et continus permettent alors de rendre compte de l'incertitude autour de l'erreur de prévision au même titre que la variance de l'erreur de prévision. En revanche pour ces modèles, la fonction de densité de la prévision est rarement, voire jamais reproduite, puisque comme nous l'avons évoqué en introduction les modèles linéaires ne permettent pas de reproduire les phénomènes, tels que l'asymétrie ou la multi modalités, susceptibles d'être analysés et interprétés lors de l'examen de la densité de prévision.

Au contraire, dans le cas des modèles non linéaires les trois formes de prévisions retrouvent toutes leurs justifications et leur intérêt. En effet, un des messages essentiels de la littérature est qu'il totalement réducteur de se contenter d'une prévision ponctuelle dans le cas d'un modèle non linéaire susceptible par exemple de générer des asymétries. Les intervalles de confiance associés aux prévisions peuvent alors être non symétriques voire même discontinus avec l'émergence de concepts tels que la *HDR* (*High Density Region*). La fonction de densité de la prévision peut de la même façon révéler des évolutions possibles des variables économiques beaucoup

plus fines et nuancées, liées notamment aux asymétries de cette distribution, que ne le permet une simple prévision ponctuelle moyenne.

Enfin, se pose un certain nombre de difficultés techniques liées à la construction même d'une prévision ponctuelle dans le cadre de certains modèles non linéaires qui nécessitent souvent le recours à des méthodes de simulation numérique. C'est pour cette raison que nous évoquerons successivement d'une part les méthodes de prévision ponctuelle et d'autre part les méthodes d'intervalle de confiance et de densité de prévision.

3.1 Méthodes de Prévision Ponctuelle

Tout comme dans le cas des modèles linéaires, les prévisions établies à partir de modèles non linéaires peuvent être exprimées sous la forme d'une valeur ponctuelle traduisant une notion de prévision au «point moyen». Considérons un modèle non linéaire s'écrivant sous la forme suivante :

$$y_t = F(x_t, \theta) + \varepsilon_t \quad (3.1)$$

où ε_t est *i.i.d.* $(0, \sigma_\varepsilon^2)$ et x_t désigne un ensemble de variables pouvant inclure à la fois des variables exogènes et des valeurs retardées de la variable endogène. On appelle skeleton, l'espérance conditionnelle du processus y_t , *i.e.* $F(x_t, \theta)$. Soit $\hat{y}_{t+h|t}$ le prédicteur optimal de y_{t+h} compte tenu de l'ensemble d'information disponible à la date t , noté Ω_t , et soit $\hat{\varepsilon}_{t+h|t}$ l'erreur de prévision associée à $\hat{y}_{t+h|t}$. Par définition, on a :

$$\hat{y}_{t+h|t} = E[y_{t+h} \mid \Omega_t] \quad h \geq 1 \quad (3.2)$$

$$\hat{\varepsilon}_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t} \quad (3.3)$$

Quel que soit le modèle non linéaire envisagé, la prévision à l'horizon d'une période ($h = 1$) ne pose pas de problème particulier. Puisque $E[\varepsilon_{t+1} \mid \Omega_t] = 0$, la prévision ponctuelle optimale à l'horizon un corre-

spond toujours au skeleton¹ :

$$\widehat{y}_{t+1|t} = F(x_{t+1}, \theta) \quad (3.4)$$

Dès lors, la prévision ponctuelle à l'ordre un de tout modèle non linéaire peut s'obtenir de façon analytique comme s'était le cas pour les modèles linéaires. En revanche, pour un ordre de prévision $h > 1$, dans le cas de certains modèles non linéaires, il n'existe pas de formule analytique permettant d'obtenir aisément cette prévision ponctuelle. C'est en particulier le cas des modèles à seuil pour lesquels on est obligé d'avoir recours à différentes méthodes d'approximation ou méthodes numériques pour obtenir les prévisions. Avant d'évoquer ces dernières, commençons par considérer le cas le plus simple dans lequel il existe une forme analytique aux prévisions ponctuelles.

3.1.1 Méthode Analytique

Il n'est possible de donner une expression analytique à la prévision pour un horizon supérieur à une période que pour quelques spécifications particulières de modèles non linéaires. Teräsvirta (2005) recense en particulier deux types de modèles : le modèle moyenne mobile à seuil *TMA* et le modèle à changement de régime markovien.

Considérons le cas le plus simple d'un modèle de type moyenne mobile à seuil ou moyenne mobile asymétrique tel que :

$$y_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{j=1}^q \psi_j \varepsilon_{t-j} \mathbb{I}_{(\varepsilon_{t-j} > 0)} + \varepsilon_t \quad (3.5)$$

où $\mathbb{I}_{(\varepsilon_{t-j} > 0)}$ désigne la fonction indicatrice valant 1 si $\varepsilon_{t-j} > 0$ et 0 dans le cas contraire. On suppose que le résidu ε_t est distribué selon une loi normale $N(0, \sigma_\varepsilon^2)$. Comme mentionné précédemment, la prévision à l'ordre 1 est

¹Rappelons que par définition du vecteur x_t , la quantité x_{t+1} est soit exogène, soit connue à la date t . Si, par exemple, le modèle est autorégressif d'ordre un, *i.e.* $x_t = y_{t-1}$, alors $\widehat{y}_{t+1|t} = F(y_t, \theta)$.

équivalente au skeleton :

$$\widehat{y}_{t+1|t} = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t+1-j} + \sum_{j=1}^q \psi_j \varepsilon_{t+1-j} \mathbb{I}_{(\varepsilon_{t+1-j} > 0)} \quad (3.6)$$

La formule de la prévision à l'ordre 2 est similaire à celle d'un modèle de type moyenne mobile linéaire à l'exception du terme correspondant à l'espérance de la variable censurée, *i.e.* $\varepsilon_{t+1} \mathbb{I}_{(\varepsilon_{t+1} > 0)}$.

$$\begin{aligned} \widehat{y}_{t+2|t} = & \mu + \sum_{j=2}^q \theta_j \varepsilon_{t+2-j} \\ & + \sum_{j=2}^q \psi_j \varepsilon_{t+2-j} \mathbb{I}_{(\varepsilon_{t+2-j} > 0)} + \psi_1 E [\varepsilon_{t+1} \mathbb{I}_{(\varepsilon_{t+1} > 0)}] \end{aligned} \quad (3.7)$$

Rappelons que même si ε_t est un processus centré sur 0, l'espérance de la variable normale tronquée $\varepsilon_t \mathbb{I}_{(\varepsilon_t > 0)}$ n'est pas nulle² et vérifie $E [\varepsilon_t \mathbb{I}_{(\varepsilon_t > 0)}] = \sigma/\sqrt{2\pi}$. On obtient ainsi :

$$\begin{aligned} \widehat{y}_{t+2|t} = & \mu + \sum_{j=2}^q \theta_j \varepsilon_{t+2-j} \\ & + \sum_{j=2}^q \psi_j \varepsilon_{t+2-j} \mathbb{I}_{(\varepsilon_{t+2-j} > 0)} + \psi_1 \frac{\sigma}{\sqrt{2\pi}} \end{aligned} \quad (3.8)$$

De façon générale si l'ordre de prévision h est inférieur à l'ordre q de la moyenne mobile, la prévision ponctuelle à l'ordre h , notée $\widehat{y}_{t+h|t} = E [y_{t+h} \mid \Omega_t]$, vérifie :

$$\begin{aligned} \widehat{y}_{t+h|t} = & \mu + \sum_{j=h}^q \theta_j \varepsilon_{t+h-j} \\ & + \sum_{j=h}^q \psi_j \varepsilon_{t+h-j} \mathbb{I}_{(\varepsilon_{t+h-j} > 0)} + \frac{\sigma}{\sqrt{2\pi}} \sum_{j=1}^{h-1} \psi_j \end{aligned} \quad (3.9)$$

Pour un horizon de prévision supérieur à l'ordre q de la moyenne mobile, la prévision correspond alors à l'espérance non conditionnelle du processus

²De façon générale, si $z \sim N(m, \sigma^2)$, alors $E [z \mathbb{I}_{(z > 0)}] = m\Phi\left(\frac{m}{\sigma}\right) + \sigma\phi\left(\frac{m}{\sigma}\right)$ où $\Phi(\cdot)$ et $\phi(\cdot)$ désignent respectivement les fonctions de répartition et de distribution de la loi normale centrée réduite.

y_t , c'est à dire à :

$$\hat{y}_{t+h|t} = E(y_t) = \mu + \frac{\sigma}{\sqrt{2\pi}} \sum_{j=1}^q \psi_j \quad h > q \quad (3.10)$$

Reste que seules quelques spécifications très particulières des modèles non linéaires permettent d'établir de telles formules analytiques pour les prévisions à un horizon supérieur à une période. Dans la plupart des cas, et en particulier dans le cas des modèles à seuil, il n'existe pas de formule analytique et l'on est obligé d'avoir recours à différentes méthodes d'approximation ou méthodes numériques pour obtenir les prévisions. Parmi ces méthodes, figurent au moins cinq méthodes alternatives :

- Méthode de simulation par Monte-Carlo
- Méthode de simulation par Bootstrap
- Méthode dite "Naïve" ou méthode dite "du Skeleton"
- Méthode *Normal Forecast Error* (NFE)
- Méthode dite de l'Estimation Dynamique.

Nous allons à présent, présenter successivement ces différentes méthodes dans le cas particulier des modèles à seuils.

3.1.2 Méthodes de Simulation Numérique

Considérons un modèle autorégressif d'ordre 1 de type STAR(1, 1, 1) :

$$y_t = F(y_{t-1}, \theta) + \varepsilon_t \quad (3.11)$$

où $F(y_{t-1}, \theta)$ est une fonction non linéaire et θ un ensemble de paramètres. Considérons le cas d'un modèle STAR où la variable de transition correspond au niveau de l'endogène retardé d'une période, *i.e.* $x_t = y_{t-1}$. La fonction $F(y_{t-1}, \theta)$ est alors définie par :

$$\begin{aligned} F(y_{t-1}, \theta) &= (\phi_{1,0} + \phi_{1,1}y_{t-1}) \times [1 - G(y_{t-1}; \gamma, c)] \\ &\quad + (\phi_{2,0} + \phi_{2,1}y_{t-1}) \times G(y_{t-1}; \gamma, c) \end{aligned} \quad (3.12)$$

Comme nous l'avons mentionné, à l'horizon d'une période, sachant que $E[\varepsilon_{t+1} | \Omega_t] = 0$, la prévision ponctuelle optimale correspond simplement au skeleton :

$$\widehat{y}_{t+1|t} = E[y_{t+1} | \Omega_t] = F(y_t, \theta) \quad (3.13)$$

Toutefois, pour un ordre h supérieur à une période, il n'est plus possible d'obtenir une forme analytique de la prévision. Par exemple, si l'on souhaite prévoir le niveau de y_{t+2} , le prédicteur optimal s'écrit alors sous la forme suivante :

$$\widehat{y}_{t+2|t} = E[y_{t+2} | \Omega_t] = E[F(y_{t+1}, \theta) | \Omega_t] \quad (3.14)$$

Tout le problème tient alors au fait que de façon générale l'opérateur linéaire espérance ne peut pas être interchangé avec la fonction non linéaire $F(\cdot)$. En d'autres termes, tout le problème provient du fait que :

$$E[F(y_{t+1}, \theta) | \Omega_t] \neq F[E(y_{t+1} | \Omega_t); \theta] \quad (3.15)$$

et donc par conséquent :

$$\widehat{y}_{t+2|t} \neq F(\widehat{y}_{t+1|t}; \theta) \quad (3.16)$$

Ainsi, contrairement à un modèle linéaire, il n'existe pas de relation récursive simple entre la prévision à des horizons successifs qui permettrait de calculer aisément la prévision à un horizon h quelconque. La prévision optimale à l'ordre deux est ainsi définie par la relation suivante³ :

$$\widehat{y}_{t+2|t} = E[F(\widehat{y}_{t+1|t} + \varepsilon_{t+1}; \theta) | \Omega_t] \quad (3.17)$$

Dès lors, si l'on souhaite prévoir la valeur de y_{t+2} conditionnellement à l'information disponible à la date t il convient d'évaluer la quantité :

$$\widehat{y}_{t+2|t} = \int_{-\infty}^{\infty} F(\widehat{y}_{t+1|t} + \varepsilon_{t+1}; \theta) f(\varepsilon) d\varepsilon \quad (3.18)$$

³En effet, en omettant les paramètres dans l'expression de $F(\cdot)$, il vient

$$\widehat{y}_{t+2|t} = E[F(y_{t+1}) | \Omega_t] = E[F(F(y_t) + \varepsilon_{t+1}) | \Omega_t] = E[F(\widehat{y}_{t+1|t} + \varepsilon_{t+1}) | \Omega_t]$$

où $f(\varepsilon)$ désigne la fonction de densité du choc ε_{t+1} . Etant donné qu'il n'est généralement pas possible de dériver une expression analytique de cette intégrale, l'évaluation de la prévision ponctuelle requiert l'utilisation de méthodes numériques fondées sur des simulations.

La méthode de simulation numérique la plus simple est la méthode de Monte-Carlo. Cette méthode nécessite que l'on fasse une hypothèse sur la distribution des résidus ε_{t+1} . On simule alors N tirages indépendants $\{\varepsilon_{t+1}^{(1)}, \dots, \varepsilon_{t+1}^{(i)}, \dots, \varepsilon_{t+1}^{(N)}\}$ dans cette distribution et la prévision ponctuelle de y_{t+2} , notée $\hat{y}_{t+2|t}^{MC}$, est alors définie par la quantité moyenne :

$$\begin{aligned}\hat{y}_{t+2|t}^{MC} &= \frac{1}{N} \sum_{i=1}^N \hat{y}_{t+2|t}^{(i)} \\ &= \frac{1}{N} \sum_{i=1}^N F(\hat{y}_{t+1|t}; \theta) + \varepsilon_{t+1}^{(i)}\end{aligned}\quad (3.19)$$

Une méthode alternative consiste à ré-échantillonner les résidus historiques, notés $\hat{\varepsilon}_t$, en tirant des séquences $\{\hat{\varepsilon}_{t+1}^{(1)}, \dots, \hat{\varepsilon}_{t+1}^{(i)}, \dots, \hat{\varepsilon}_{t+1}^{(N)}\}$ dans la séquence des résidus historiques $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_t\}$. La prévision ponctuelle de y_{t+2} par méthode de Bootstrap, notée $\hat{y}_{t+2|t}^B$, est définie de façon similaire par la quantité moyenne :

$$\begin{aligned}\hat{y}_{t+2|t}^B &= \frac{1}{N} \sum_{i=1}^N \hat{y}_{t+2|t}^{(i)} \\ &= \frac{1}{N} \sum_{i=1}^N F(\hat{y}_{t+1|t}; \theta) + \hat{\varepsilon}_{t+1}^{(i)}\end{aligned}\quad (3.20)$$

Les avantages et inconvénients respectifs de ces deux méthodes sont bien connus. L'avantage essentiel de la méthode Bootstrap par rapport à la méthode de Monte Carlo est qu'elle ne nécessite a priori aucune hypothèse sur la distribution de ε_{t+1} .

La généralisation de ces deux formules conduit des relations de récurrence qui permettent (par simulations) d'obtenir la valeur de la prévision

ponctuelle à un horizon h quelconque. Ainsi, de façon générale, les prévisions ponctuelles obtenues par méthode de Monte Carlo et de Bootstrap sont définies par les relations suivantes pour $h \geq 2$:

$$\widehat{y}_{t+h|t}^{MC} = \frac{1}{N} \sum_{i=1}^N \widehat{y}_{t+h|t}^{(i)} \quad (3.21)$$

où dans le cas de la méthode de Monte Carlo, on :

$$\widehat{y}_{t+j|t}^{(i)} = F(\widehat{y}_{t+j-1|t}; \theta) + \varepsilon_{t+j-1}^{(i)} \quad (3.22)$$

et dans le cas de la méthode de Bootstrap :

$$\widehat{y}_{t+j|t}^{(i)} = F(\widehat{y}_{t+j-1|t}; \theta) + \widehat{\varepsilon}_{t+h-1}^{(i)} \quad (3.23)$$

où $\varepsilon_{t+h-1}^{(i)}$ et $\widehat{\varepsilon}_{t+h-1}^{(i)}$ désignent respectivement deux tirages du choc ε_{t+h-1} issus soit d'une distribution postulée a priori, soit d'un tirage au hasard dans la séquence historique des chocs estimés. La condition initiale de ces relations de récurrence étant alors fournie par le skeleton pour l'ordre $h = 1$ (équation 3.13).

3.1.3 Méthode "Naïve" ou Méthode dite du "Skeleton"

Une autre solution permettant d'obtenir une prévision ponctuelle dans un modèle non linéaire consiste à négliger⁴ le terme d'erreur de prévision ε_{t+1} dans la formule (3.17). En posant $\varepsilon_{t+1} = 0$, la prévision optimale de y_{t+2} , dite prévision "naïve" et notée $\widehat{y}_{t+2|t}^N$, correspond alors simplement au skeleton du modèle non linéaire :

$$\widehat{y}_{t+2|t}^N = E[F(\widehat{y}_{t+1|t}; \theta) | \Omega_t] = F(\widehat{y}_{t+1|t}; \theta) \quad (3.24)$$

avec $\widehat{y}_{t+1|t} = F(y_t; \theta)$ (cf. équation 3.13). On obtient alors une formule récursive facile à mettre en œuvre et qui ne nécessite aucune simulation numérique. Pour un ordre h quelconque la valeur de la prévision "naïve"

⁴Autrement dit cela revient à interchanger les opérateurs $F(\cdot)$ et $E(\cdot)$ dans l'équation 3.14 de définition de $\widehat{y}_{t+2|t}$.

est définie par une relation récursive de la forme :

$$\hat{y}_{t+h|t}^N = F\left(\hat{y}_{t+h-1|t}^N; \theta\right) \quad h \geq 1 \quad (3.25)$$

Cette méthode "naïve", ou méthode dite du "skeleton", a été utilisée entre autres par Tong (1990) et Brown et Mariano (1989). Naturellement, elle conduit à une forte perte d'efficacité et à des prévisions biaisées. Lin et Granger (1994) montrent ainsi que les méthodes de type Bootstrap ou Monte Carlo dominent largement la méthode naïve en termes de biais de prévision dans de nombreux exercices de simulations.

3.1.4 Méthode Normal Forecast Error (NFE)

Comme nous l'avons vu, le calcul de la prévision optimale de y_{t+2} nécessite d'évaluer une intégrale définie notamment à partir de la fonction de densité des erreurs de prévisions commises sur la prévision à l'ordre un (équation 3.18). Il existe toutefois une autre façon d'écrire la prévision optimale. Celle-ci revient à exprimer la prévision optimale en fonction de la distribution conditionnelle du "vrai" processus y_{t+1} par rapport à l'ensemble d'information disponible à la date t . Formellement, la définition de la prévision optimale $\hat{y}_{t+2|t}$ (équation 3.18) peut se réécrire sous deux formes alternatives :

$$\hat{y}_{t+2|t} = \int_{-\infty}^{\infty} F(\hat{y}_{t+1|t} + \varepsilon_{t+1}; \theta) f(\varepsilon) d\varepsilon \quad (3.26)$$

ou encore

$$\hat{y}_{t+2|t} = \int_{-\infty}^{\infty} F(y_{t+1}; \theta) g(y_{t+1} | \Omega_t) dy_{t+1} \quad (3.27)$$

où $g(y_{t+1} | \Omega_t)$ désigne la distribution conditionnelle de y_{t+1} par rapport à l'ensemble d'information Ω_t . Comme le note Van Dijk (2002), cette distribution correspond à la distribution du résidu⁵ ε_{t+1} avec une moyenne égale à $F(y_t; \theta)$. En effet, compte tenu de la définition de la prévision à l'ordre un (équation 3.13), l'erreur de prévision ε_{t+1} est définie par l'écart entre

⁵De par la définition de la prévision à l'ordre un, l'erreur de prévision correspond ici au choc ε_{t+1} .

le vrai processus y_{t+1} et le skeleton, *i.e.* $\varepsilon_{t+1} = y_{t+1} - F(y_t; \theta)$. Par conséquent, la fonction de densité associée à l'erreur de prévision ε_{t+1} , notée $f(\varepsilon_{t+1})$, peut se réécrire sous la forme $f[y_{t+1} - F(y_t; \theta)]$. Conditionnellement à l'ensemble d'information Ω_t , le skeleton $F(y_t; \theta)$ est connu. Ainsi, la densité non conditionnelle des erreurs de prévision $f(\varepsilon_{t+1})$ correspond en fait à la densité conditionnelle du vrai processus y_{t+1} centré sur $F(y_t; \theta)$. Enfin, sachant que le skeleton $F(y_{t+1}; \theta)$ correspond à l'espérance conditionnelle de y_{t+2} sachant y_{t+1} , la prévision optimale peut s'écrire sous la forme suivante :

$$\hat{y}_{t+2|t} = \int_{-\infty}^{\infty} E(y_{t+2}|y_{t+1}) g(y_{t+1}|\Omega_t) dy_{t+1} \quad (3.28)$$

Naturellement, cette définition ne résout en rien le problème de l'évaluation de cette intégrale. Mais elle présente certaines propriétés intéressantes lorsque que l'on envisage la formule de révision des prévisions entre deux dates successives.

En effet, dès lors que l'on cherche à évaluer la prévision optimale à un ordre h quelconque, se pose immédiatement le problème de la caractérisation de la distribution conditionnelle $g(y_{t+h}|\Omega_t)$. Toutefois, celle-ci peut s'écrire en fonction de la distribution conditionnelle $g(y_{t+h}|y_{t+h-1})$ sous la forme :

$$g(y_{t+h}|\Omega_t) = \int_{-\infty}^{\infty} g(y_{t+h}|y_{t+h-1}) g(y_{t+h-1}|\Omega_t) dy_{t+h-1} \quad (3.29)$$

Ainsi l'équation 3.28 peut être généralisée à la prévision optimale $\hat{y}_{t+h|t}$ pour un ordre h quelconque.

$$\hat{y}_{t+h|t} = E(y_{t+h}|\Omega_t) = \int_{-\infty}^{\infty} E(y_{t+h}|y_{t+h-1}) g(y_{t+h-1}|\Omega_t) dy_{t+h-1} \quad (3.30)$$

Cette équation signifie que la prévision optimale à l'ordre h dépend de la distribution conditionnelle de y_{t+h-1} . Par conséquent, une procédure itérative peut être envisagée à partir de la relation récursive (3.29) : pour caractériser $g(y_{t+h-1}|\Omega_t)$ en fonction de $g(y_{t+h-2}|\Omega_t)$ (et ainsi de suite jusqu'à

$g(y_{t+1}|\Omega_t)$, il suffit de calculer la quantité $g(y_{t+h-1}|y_{t+h-2})$. Or rappelons que celle-ci n'est rien d'autre que la distribution du choc ε_{t+h-1} dont la moyenne est fixée à $F(y_{t+h-2}; \theta)$. La procédure récursive peut démarrer pour $h = 2$ en utilisant le fait que $g(y_{t+1}|\Omega_t) = f[y_{t+1} - F(y_t; \theta)]$. La caractérisation de $g(y_{t+h-1}|\Omega_t)$ pour $h > 2$ implique alors de répéter la phase d'intégration numérique dans l'équation (3.29). Toutefois, cette méthode peut s'avérer très lourde à implémenter notamment pour des valeurs de h élevées.

Une solution alternative consiste à supposer que l'erreur de prévision commise à l'horizon $h - 1$ est normalement distribuée. Formellement, on suppose que l'erreur $e_{t+h-1} = y_{t+h-1} - \hat{y}_{t+h-1|t}$ est distribuée selon une loi normale de moyenne nulle et de variance σ_{h-1}^2 . Sous cette hypothèse, la distribution conditionnelle $g(y_{t+h-1}|\Omega_t)$ correspond à une distribution normale d'espérance égale à $\hat{y}_{t+h-1|t}$, et de variance égale à σ_{h-1}^2 . Cette méthode, fondée sur l'hypothèse de normalité de la dernière erreur de prévision et proposée initialement par Pemberton (1987), est connue sous le nom de méthode *NFE* pour *Normal Forecast Error*.

3.1.5 Méthode de l'Estimation Dynamique ou *Dynamic Estimation (DE) Method*

Dans l'esprit des méthodes d'estimation dynamique appliquées aux modèles linéaires (Clements et Hendry, 1996), il est possible de construire une prévision ponctuelle pour y_{t+h} à partir de l'information disponible en t en établissant directement au sein de l'échantillon d'estimation (*in sample*) une relation entre y_t et y_{t-h} . Par exemple, si l'on souhaite construire une prévision pour un horizon h de deux périodes, on commence par établir la forme de la relation entre y_t et y_{t-2} . Formellement, si l'on considère le cas d'un modèle non linéaire autorégressif d'ordre un, on a :

$$y_t = F(y_{t-1}; \theta) + \varepsilon_t = F[F(y_{t-2}; \theta) + \varepsilon_{t-1}; \theta] + \varepsilon_t \quad (3.31)$$

Ce modèle peut parfois se récrire sous une forme simplifiée de type :

$$y_t = G(y_{t-1}; \theta) + v_t \quad (3.32)$$

où v_t désigne une combinaison éventuellement non linéaire des chocs ε_t et ε_{t-1} . Ce modèle simplifié peut être estimé et dès lors, à partir de cette relation estimée, il est possible de prévoir directement y_{t+2} à partir uniquement de l'observation de y_t :

$$\hat{y}_{t+2|t} = G(y_t; \hat{\theta}) \quad (3.33)$$

Cette fonction $G(y_t; \hat{\theta})$, sauf pour le cas particulier d'une prévision à l'ordre $h = 1$ ne correspond naturellement pas au skeleton $F(y_t; \theta)$ du modèle non linéaire. Cette méthode discutée dans Granger et Teräsvirta (1993) a notamment été utilisée par Clements et Smith (1997) dans le cadre des modèles SETAR.

3.1.6 Application au Modèle SETAR

Afin d'illustrer ces différentes méthodes de calcul de prévision ponctuelle, nous proposons ici une application au cas d'un modèle simple type SETAR(1, 1, 1). On considère le processus non linéaire suivant :

$$y_t = \phi_{1,0} + \phi_{1,1}y_{t-1} + (\phi_{2,0} + \phi_{2,1}y_{t-1}) \times \mathbb{I}_{y_{t-1} > c} + \varepsilon_t \quad (3.34)$$

où $\mathbb{I}_{y_{t-1} > c}$ désigne la fonction indicatrice telle que $\mathbb{I}_{y_{t-1} > c} = 1$ si $y_{t-1} > c$ et 0 sinon. On suppose que les paramètres de ce modèles notés $\theta = \{\phi_{1,0} : \phi_{1,1} : \phi_{2,0} : \phi_{2,1} : c\}$ sont connus. On suppose que les résidus sont normalement distribués *N.i.d.* $(0, \sigma_\varepsilon^2)$. Soit $F(y_{t-1}, \theta)$ le skeleton associé à y_t tel que :

$$y_t = F(y_{t-1}, \theta) + \varepsilon_t \quad (3.35)$$

$$F(y_{t-1}, \theta) = \phi_{1,0} + \phi_{1,1}y_{t-1} + (\phi_{2,0} + \phi_{2,1}y_{t-1}) \times \mathbb{I}_{y_{t-1} > c} \quad (3.36)$$

Nous avons vu que pour la prévision à l'ordre 1, notée $\hat{y}_{t+1|t} = E[y_{t+1}|\Omega_t]$, il n'y pas de divergence de méthode puisque la prévision optimale se ramène au skeleton :

$$\hat{y}_{t+1|t} = F(y_t, \theta) = \phi_{1,0} + \phi_{1,1}y_t + (\phi_{2,0} + \phi_{2,1}y_t) \times \mathbb{I}_{y_t > c} \quad (3.37)$$

En revanche, pour un horizon h au moins égal à deux périodes, les différentes méthodes de prévision ponctuelle donnent des résultats différents. La prévision "naïve" ou du "skeleton" est obtenue en négligeant le terme d'erreur de prévision dans la formule (3.17). Ainsi, la prévision "naïve" à l'ordre h , notée $\hat{y}_{t+h|t}^N$, est définie par une formule récursive en fonction de la prévision "naïve" à l'horizon $h-1$:

$$\begin{aligned} \hat{y}_{t+h|t}^N &= F(\hat{y}_{t+h-1|t}^N; \theta) \\ &= \phi_{1,0} + \phi_{1,1}\hat{y}_{t+h-1|t}^N + (\phi_{2,0} + \phi_{2,1}\hat{y}_{t+h-1|t}^N) \times \mathbb{I}_{\hat{y}_{t+h-1|t}^N > c} \end{aligned} \quad (3.38)$$

La prévision optimale à l'ordre $h \geq 2$ obtenue par la méthode de Monte Carlo suppose que l'on réalise N tirages indépendants de séquences de chocs $\{\varepsilon_{t+1}^{(i)}, \varepsilon_{t+2}^{(i)}, \dots, \varepsilon_{t+h}^{(i)}\}_{i=1}^N$ dans une distribution supposée pour les résidus. A partir de ces séquences de chocs, on peut alors reconstruire des séquences "d'histoires" de la variables endogène, notées $\{\hat{y}_{t+h|t}^i\}_{i=1}^N$, selon la formule de récurrence :

$$\hat{y}_{t+h|t}^i = \phi_{1,0} + \phi_{1,1}\hat{y}_{t+h-1|t}^i + (\phi_{2,0} + \phi_{2,1}\hat{y}_{t+h-1|t}^i) \times \mathbb{I}_{\hat{y}_{t+h-1|t}^i > c} + \varepsilon_{t+h-1}^{(i)} \quad (3.39)$$

$$\hat{y}_{t+1|t}^i = \hat{y}_{t+1|t} = F(y_t, \theta) \quad \forall i = 1, \dots, N \quad (3.40)$$

A partir de ces N séquences simulées, on construit un estimateur de Monte Carlo $\hat{y}_{t+h|t}^{MC}$ à l'horizon h en considérant tout simplement la moyenne empirique des N valeurs simulées $\hat{y}_{t+h|t}^i$:

$$\hat{y}_{t+h|t}^{MC} = \frac{1}{N} \sum_{i=1}^N \hat{y}_{t+h|t}^i \quad (3.41)$$

La prévision optimale obtenue par la méthode de Bootstrap, notée $\widehat{y}_{t+h|t}^B$, sera obtenue par une formule de récurrence identique à la différence près que les chocs utilisés dans les simulations auront été obtenus par ré-échantillonnage des résidus historiques.

La prévision ponctuelle à l'ordre h obtenue par la méthode *NFE*, notée est obtenue par résolution de l'intégrale suivante :

$$\widehat{y}_{t+h|t}^{NFE} = \int_{-\infty}^{\infty} E(y_{t+h}|y_{t+h-1}) g(y_{t+h-1}|\Omega_t) dy_{t+h-1} \quad (3.42)$$

où $g(y_{t+h-1}|\Omega_t)$ désigne une distribution normale de moyenne nulle et de variance σ_{h-1}^2 . Clements et Smith (1997) montrent que dans le cas du modèle SETAR, cette récurrence se ramène à la formule suivante :

$$\begin{aligned} \widehat{y}_{t+h|t}^{NFE} &= \Phi\left(\frac{c - \widehat{y}_{t+h-1|t}^{NFE}}{\sigma_{h-1}}\right) \times (\phi_{1,0} + \phi_{1,1}\widehat{y}_{t+h-1|t}^{NFE}) \\ &+ \Phi\left(\frac{\widehat{y}_{t+h-1|t}^{NFE} - c}{\sigma_{h-1}}\right) \times (\phi_{2,0} + \phi_{2,1}\widehat{y}_{t+h-1|t}^{NFE}) \\ &+ \phi\left(\frac{c - \widehat{y}_{t+h-1|t}^{NFE}}{\sigma_{h-1}}\right) \times (\phi_{2,1} - \phi_{1,1}) \times \sigma_{h-1} \end{aligned} \quad (3.43)$$

où $\Phi(\cdot)$ et $\phi(\cdot)$ désignent respectivement les fonctions de répartition et de densité de la loi normale standard et où $\widehat{y}_{t+h-1|t}^{NFE}$ désigne la prévision optimale à l'ordre $h-1$. Cette formule indique que la prévision optimale n'est rien d'autre qu'une somme pondérée des prévisions obtenues dans chacun des deux régimes du SETAR. Les poids sont définis par la probabilité que l'observation y_{t+h} appartienne à un régime particulier. En effet, conditionnellement à la valeur de y_{t+h-1} , la probabilité que l'observation y_{t+h} appartienne au régime 1 (défini par les paramètres $\phi_{1,0}$ et $\phi_{1,1}$) est égale à la probabilité que y_{t+h-1} soit inférieur au paramètre de seuil c , soit sous l'hypothèse de normalité $\Phi[(c - y_{t+h-1})/\sigma_{h-1}]$.

Notons que dans l'équation (3.43) ces probabilités sont établies (sous

l'hypothèse de normalité) par comparaison de la prévision⁶ $\widehat{y}_{t+h-1|t}^{NFE}$ au seuil c . Le dernier terme correspond alors à un terme de correction. De Gooijer and de Bruin (1998) dérivent les formules de récurrence pour la variance de l'erreur de prévision σ_{h-1}^2 d'un modèle SETAR(q, p_1, p_2) et montrent en particulier que :

$$\begin{aligned} \sigma_h^2 = & \left[\left(\phi_{1,0} + \phi_{1,1} \widehat{y}_{t+h-1|t}^{NFE} \right)^2 + \phi_{1,1}^2 \sigma_{h-1}^2 \right] \times p_{h-1} \\ & + \left[\left(\phi_{2,0} + \phi_{2,1} \widehat{y}_{t+h-1|t}^{NFE} \right)^2 + \phi_{2,1}^2 \sigma_{h-1}^2 \right] \times (1 - p_{h-1}) \\ & \times \left[\phi_{2,1}^2 \left(c - \widehat{y}_{t+h-1|t}^{NFE} \right) + 2\phi_{2,1} \left(\phi_{2,0} + \phi_{2,1} \widehat{y}_{t+h-1|t}^{NFE} \right) \right. \\ & \left. - \phi_{1,1}^2 \left(c - \widehat{y}_{t+h-1|t}^{NFE} \right) - 2\phi_{1,1} \left(\phi_{1,0} + \phi_{1,1} \widehat{y}_{t+h-1|t}^{NFE} \right) \right] \times \sigma_{h-1} \times p_{h-1} \\ & + \sigma_\varepsilon^2 - \left(\widehat{y}_{t+h|t}^{NFE} \right)^2 \end{aligned} \quad (3.44)$$

où σ_ε^2 désigne la variance des résidus du modèle et où

$$p_{h-1} = \Phi \left[\left(c - \widehat{y}_{t+h-1|t}^{NFE} \right) / \sigma_{h-1} \right] \quad (3.45)$$

désigne la probabilité estimée que l'observation y_{t+h} appartienne au régime un.

Considérons enfin la prévision ponctuelle par la méthode d'Estimation Directe (DE). Dans ce cas il convient au préalable d'estimer une relation entre la valeur y_{t+h} et y_t . Considérons le cas d'une prévision à l'ordre $h = 2$. On sait que :

$$\begin{aligned} y_{t+2} = & \phi_{1,0} + \phi_{1,1} y_{t+1} + \left(\phi_{2,0} + \phi_{2,1} y_{t+1} \right) \times \mathbb{I}_{y_{t+1} > c} + \varepsilon_{t+2} \\ = & \phi_{1,0} + \phi_{1,1} \left[\phi_{1,0} + \phi_{1,1} y_t + \left(\phi_{2,0} + \phi_{2,1} y_t \right) \times \mathbb{I}_{y_t > c} + \varepsilon_{t+1} \right] \\ & + \left\{ \phi_{2,0} + \phi_{2,1} \left[\phi_{1,0} + \phi_{1,1} y_t + \left(\phi_{2,0} + \phi_{2,1} y_t \right) \times \mathbb{I}_{y_t > c} + \varepsilon_{t+1} \right] \right\} \times \mathbb{I}_{y_{t+1} > c} + \varepsilon_{t+2} \end{aligned} \quad (3.46)$$

En négligeant les termes multipliés par l'indicatrice $\mathbb{I}_{y_{t+1} > c}$, on se ramène ainsi à une relation non linéaire entre y_{t+2} et y_t de la forme :

$$y_{t+2} = \psi_{0,2} + \psi_{1,2} y_t + \mathbb{I}_{y_t > \theta_2} \left(\psi_{2,2} + \psi_{3,2} y_t \right) + \mu_{t+2} \quad (3.47)$$

⁶La valeur de y_{t+h-1} étant inconnue à la date t .

où les paramètres $\psi_{j,2}$, $j = 0, \dots, 3$ et le seuil θ_2 sont des combinaisons non linéaires des paramètres initiaux. Reste alors à estimer *in sample* cette relation en utilisant la même méthode de Moindres Carrés Non Linéaires que celle utilisée pour l'estimation des paramètres d'un modèle SETAR ou STAR usuel. Compte tenu des paramètres estimés $\widehat{\psi}_j$ et $\widehat{\theta}$, la prévision optimale, notée $\widehat{y}_{t+2|t}^{DE}$, s'écrit :

$$\widehat{y}_{t+2|t}^{DE} = \widehat{\psi}_0 + \widehat{\psi}_1 y_t + \mathbb{I}_{y_t > \theta} \left(\widehat{\psi}_2 + \widehat{\psi}_3 y_t \right) \quad (3.48)$$

De la même façon, la généralisation pour une prévision à un ordre h conduit à une relation de la forme :

$$\widehat{y}_{t+h|t}^{DE} = \widehat{\psi}_{0,h} + \widehat{\psi}_{1,h} y_t + \mathbb{I}_{y_t > \theta_h} \left(\widehat{\psi}_{2,h} + \widehat{\psi}_{3,h} y_t \right) \quad (3.49)$$

où $\widehat{\psi}_{j,h}$, $j = 0, \dots, 3$ et $\widehat{\theta}_h$ désignent des estimateurs de combinaisons (non linéaires) des paramètres initiaux obtenus sur la période d'échantillon par la projection (non linéaire) de y_{t+h} sur y_t .

Lin et Granger (1994) ou Clements et Smith (1997) ont ainsi comparé ces différentes méthodes d'estimation ponctuelles dans le cas des modèles STAR et SETAR. Leur principale conclusion est que les méthodes de simulation par Monte Carlo ou Bootstrap donnent de bons résultats comparativement aux autres méthodes et présentent en outre l'avantage d'une grande simplicité dans leur mise en oeuvre.

3.2 Méthodes de Prévision par Intervalle de Confiance et par Densité

Un avantage des méthodes ponctuelles de prévision par simulation (Bootstrap ou Monte-Carlo) réside dans le fait qu'elles permettent d'obtenir en sous produit un ensemble de répliques qui autorise à sous tour l'estimation de la distribution conditionnelle $g(y_{t+h}|\Omega_t)$ et/ou la construction d'un intervalle de confiance éventuellement non continu et non symétrique.

En effet, comme nous l'avons mentionné précédemment les méthodes de prévision par Monte Carlo ou par Bootstrap requièrent le calcul d'un ensemble de N simulations des prévisions conditionnelles notées $\hat{y}_{t+h|t}^{(i)}$, $i = 1, \dots, N$ avec

$$\hat{y}_{t+h|t}^{(i)} = F\left(\hat{y}_{t+h-1|t}^{(i)}; \theta\right) + \varepsilon_{t+h-1}^{(i)} \quad (3.50)$$

dans le cas des prévisions par Monte-Carlo et

$$\hat{y}_{t+h|t}^j = F\left(\hat{y}_{t+h-1|t}^{(i)}; \theta\right) + \hat{\varepsilon}_{t+h-1}^{(i)} \quad (3.51)$$

dans le cas des prévisions par Bootstrap. Il est naturel de penser que, dans ce contexte, la prévision ponctuelle apparaît réductrice, dans la mesure où elle ne conduit qu'à considérer la moyenne empirique de ces simulations avec $\hat{y}_{t+h|t}^{MC} = \sum_{i=1}^N (1/N) \hat{y}_{t+h|t}^{(i)}$ ou $\hat{y}_{t+h|t}^B = \sum_{i=1}^N (1/N) \hat{y}_{t+h|t}^{(i)}$. Au contraire, on peut envisager à partir de la distribution empirique des N simulations $\hat{y}_{t+h|t}^{(i)}$ de construire (i) soit un estimateur de la distribution conditionnelle théorique $g(y_{t+h} | \Omega_t)$, (ii) soit directement un intervalle de confiance.

3.2.1 Prévision par Densité

La prévision par densité fournit bien plus d'information au décideur économique que la seule espérance de cette distribution qui reportée dans le cas de la prévision ponctuelle. En particulier, elle permet (i) d'appréhender les éventuelles asymétries de la distribution conditionnelle de la prévision et (ii) l'éventuelle existence de plusieurs modes. En effet, contrairement au cas d'un modèle linéaire, un modèle non linéaire peut engendrer une distribution de prévisions asymétrique, et cela même dans le cas d'une distribution symétrique des résidus. Or, comme le note Teräsvirta (2005), cette information est essentielle : si par exemple, dans le cas d'une prévision d'inflation, la densité de la prévision présente une skewness positive, cela implique que les erreurs associées à la prévision ponctuelle ont une plus forte probabilité d'être positives que négatives. Ce qui peut nécessiter des réponses de politique économique différentes. L'examen des densités de

prévision permet en outre de mettre en évidence l'existence d'éventuelles caractéristiques bi-modales ou multi-modales, même si ce phénomène est a priori relativement rare sur séries macro-économiques.

Techniquement, les simulations Monte Carlo ou Bootstrap $\hat{y}_{t+h|t}^{(i)}$ peuvent être considérées comme des réalisations de la distribution conditionnelle théorique $g(y_{t+h}|\Omega_t)$ dont l'espérance correspond à la prévision optimale ponctuelle $\hat{y}_{t+h|t} = E(y_{t+h}|\Omega_t)$. A partir de ces N réalisations, on peut dès lors construire un estimateur à noyau de la densité conditionnelle par les techniques usuelles de l'économétrie non paramétriques. L'estimateur à noyau de la fonction de densité $g(y_{t+h}|\Omega_t)$ évaluée au point $\hat{y}_{t+h|t}^{(i)}$, correspondant à la $i^{\text{ème}}$ simulation, noté $\hat{g}(\hat{y}_{t+h|t}^{(i)}|\Omega_t)$, est défini par :

$$\hat{g}(\hat{y}_{t+h|t}^{(i)}|\Omega_t) = \frac{1}{N} \sum_{j=1}^N K\left(\frac{\hat{y}_{t+h|t}^{(i)} - \hat{y}_{t+h|t}^{(j)}}{\lambda}\right) \quad (3.52)$$

où $K(\cdot)$ désigne une fonction kernel et λ est un paramètre de lissage (*bandwidth parameter*). Différentes fonctions kernel peuvent être utilisées dans ce cas (gaussienne, triangulaire, quartic, Epanechnikov, quadratic spectral etc.), puisqu'il est reconnu de façon générale qu'à l'exception d'un kernel uniforme, le choix de cette fonction n'a que peu d'incidence sur le biais d'estimation de la fonction de densité⁷. Quant au choix du paramètre de lissage λ , il détermine le poids de chaque simulation $\hat{y}_{t+h|t}^{(j)}$ dans la construction de l'estimateur de la densité au point particulier $\hat{y}_{t+h|t}^{(i)}$. Ce choix est primordial non seulement en ce qui concerne le biais de l'estimateur à noyau de $g(\hat{y}_{t+h|t}^{(i)}|\Omega_t)$, mais aussi sa variance. On peut alors choisir la valeur de λ de sorte à minimiser un critère de qualité d'ajustement⁸ de l'estimateur tel que l'*AMISE* (*Approximate Mean Integrated Squared Error*) ou la *GCV* (*General Cross-Validation*).

⁷Voir Yatchew (2003) pour plus de détails sur les procédures non paramétriques d'estimation de fonction de densité.

⁸Voir Yatchew (2003) pour plus de détails.

3.2.2 Prévision par Intervalle de confiance

Une autre façon de rendre compte de l'incertitude autour de la prévision ponctuelle consiste naturellement à reporter un intervalle de confiance sur cette prévision. Dans le cadre d'un modèle linéaire, l'intervalle de confiance sur la prévision est construit comme un intervalle symétrique autour de la prévision ponctuelle. Par exemple, sous l'hypothèse de normalité des résidus du modèle, un intervalle de confiance pour un seuil de $\alpha\%$ correspond au segment continu défini par les bornes $\hat{y}_{t+h|t} - \Phi^{-1}(1 - \alpha/2) \hat{\sigma}_{h|t}$ et $\hat{y}_{t+h|t} + \Phi^{-1}(1 - \alpha/2) \hat{\sigma}_{h|t}$ où $\hat{\sigma}_{h|t}^2$ désigne la variance de l'erreur de prévision $y_{t+h} - \hat{y}_{t+h|t}$. A l'inverse dans le cas d'un modèle non linéaire, l'intervalle de confiance sur la prévision peut être (i) symétrique ou non symétrique et (ii) continu ou non continu suivant la forme de la densité conditionnelle $g(y_{t+h} | \Omega_t)$. Hyndman (1996) propose ainsi trois principales méthodes pour calculer un intervalle de confiance dans le cas d'un modèle non linéaire.

- La première méthode consiste à construire un intervalle de confiance *symétrique et continu* autour de $\hat{y}_{t+h|t}$ tel que :

$$P[\hat{y}_{t+h|t} - \omega \leq y_{t+h} \leq \hat{y}_{t+h|t} + \omega | \Omega_t] = 1 - \alpha \quad (3.53)$$

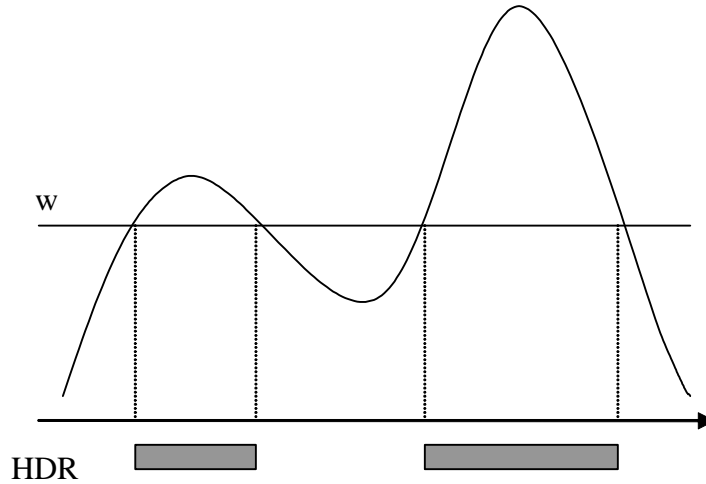
- La seconde méthode consiste à considérer un intervalle de confiance *continu (pas nécessairement symétrique)* entre les quantiles $g^{-1}(\alpha/2)$ et $g^{-1}(1 - \alpha/2)$ de la distribution conditionnelle $g(y_{t+h} | \Omega_t)$ (éventuellement approximés par les quantiles de la distribution empirique) :

$$P[g^{-1}(\alpha/2) \leq y_{t+h} \leq g^{-1}(1 - \alpha/2) | \Omega_t] = 1 - \alpha \quad (3.54)$$

- La troisième méthode revient à envisager un intervalle de confiance *non nécessairement symétrique et non nécessairement continu*. Formellement, une telle approche à définir la région de plus haute probabilité ou *HDR (High Density Region)* telle que :

$$HDR_\alpha = \{y_{t+h} | g(y_{t+h} | \Omega_t) \geq w\} \quad (3.55)$$

Fig. 3.1 Définition de la High Density Region



$$\Pr(y_{t+h} \in HDR_\alpha | \Omega_t) = 1 - \alpha \quad (3.56)$$

Sur le graphique (3.1) est représenté un exemple d'*HDR* obtenu pour une distribution conditionnelle bi-modale. Il s'agit alors de déterminer un seuil w tel que la probabilité que la série y_{t+h} appartienne à cet intervalle disjoint soit égale à $1 - \alpha\%$.

Lorsque la densité conditionnelle de la prévision $g(y_{t+h} | \Omega_t)$ est symétrique et unimodale, les trois intervalles de confiance coïncident. Seule la troisième approche fondée sur la *HDR* permet à la fois de révéler l'asymétrie et l'aspect multi-modal de la distribution de la prévision. C'est une des raisons pour lesquelles, cette méthode de construction de l'intervalle de confiance est généralement préconisée (Hyndman, 1996). De plus, on peut montrer que cette méthode conduit toujours à l'intervalle de confiance présentant la plus faible amplitude (ou une amplitude identique à celles des autres intervalles dans le cas d'une distribution continue et symétrique). Enfin, cette méthode est très facile à mettre en oeuvre à la suite des simulations de Monte-Carlo ou Bootstrap.

Chapter 4

Comment Evaluer une Prévision dans un Modèle non Linéaire ?

Généralement, les économistes s'intéressent principalement à la production et par conséquent à l'évaluation de prévisions ponctuelles, accompagnée éventuellement d'un minimum d'informations (sous la forme par exemple d'un écart type de l'erreur de prévision ou d'un intervalle de confiance) concernant le degré d'incertitude de autour de cette prévision. Mais, le débat sur l'apport des modèles non linéaires en matière de prévision a conduit à un profond renouvellement des méthodes d'évaluation de ces prévisions.

Globalement, cette évolution est allée dans le sens d'une plus grande prise en compte de l'incertitude des prévisions dans la phase d'évaluation. Sont ainsi apparues des approches visant tout d'abord à évaluer la validité des intervalles de confiance associés aux prévisions avec les travaux pionniers de Christoffersen (1998). Puis toujours dans le sens d'une plus grande prise en compte de l'incertitude, les procédures d'évaluation se sont progressivement orientées vers la l'évaluation de la densité conditionnelle complète des prévisions. En effet, on conçoit que tester la validité d'une prévision ne peut se résumer à tester la validité de la prévision moyenne. Il est certes intéressant de tester la validité de la prévision moyenne de l'inflation (2% par exemple), mais il est évident que si le modèle utilisé conduit à une prévision symétrique autour de cette valeur, alors que dans les faits, il y a deux fois plus de chances que l'inflation soit supérieure à cette valeur de 2% qu'elle lui soit inférieure, on peut sans conteste rejeter la va-

lité de ce modèle de prévision qui ne peut conduire qu'à une appréciation erronée du phénomène prévu.

Dans le cadre de cette évolution des méthodes de validation des prévisions, plusieurs dimensions de l'analyse doivent être prise en compte de façon simultanée. La première dimension du débat oppose d'une part les techniques d'évaluation des prévisions ponctuelles et par intervalle de confiance et d'autre part les méthodes d'évaluation des densités de prévision. Dans le cas des premières, les tests de validation sont généralement fondés sur des fonctions de pertes associées aux erreurs de prévisions définies par rapport aux prévisions ponctuelles ou aux séquences de violation dans le cas de prévision par intervalle de confiance. On définit une violation comme une situation dans laquelle la valeur *ex-post* de la variable prévue n'est pas comprise au sein de l'intervalle construit *ex-ante*. Dans le cas de l'évaluation des densités de prévisions, les tests de validation ne portent plus sur l'erreur de prévision, mais visent généralement à tester l'adéquation entre la densité conditionnelle de prévision et la densité du processus générateur des données (*DGP*) telle que révélée par les réalisations de la variable à prévoir. Un modèle de prévision est alors jugé valide si la densité conditionnelle de prévision est "proche" du *DGP*.

La deuxième dimension du débat oppose l'évaluation des prévisions sur la période d'estimation du modèle (*in sample*) à l'évaluation hors période d'estimation (*out of sample*). Ce choix est classique en économétrie et n'est pas spécifique en économétrie non linéaire. L'évaluation *out of sample* requiert une estimation du modèle de type récurrente (*recursive estimate*) ou de type glissante (*rolling estimate*) et donc un ajustement des formules des statistiques de tests et de leur distribution. Dans cette partie toutefois, nous ne présenterons que les formules pour le cas d'une évaluation *in sample*¹.

La troisième dimension du débat oppose les méthodes fondées sur une

¹Voir Corradi et Swanson (2006) pour le passage à une évaluation *out of sample* de ces techniques.

évaluation absolue aux méthodes fondées sur une évaluation relative des prévisions. En effet, quelle que soit la forme de la prévision (ponctuelle ou densité), il existe deux grandes classes de tests de validation. La première classe regroupe tous les tests qui s'attachent à tester la validité des prévisions issues d'un modèle en particulier et qui admettent pour hypothèse nulle la validité des prévisions issues de ce modèle. A l'inverse, d'autres tests reposent au contraire sur une évaluation relative des prévisions issues de deux modèles ou plus généralement d'un nombre fini de modèles. Un exemple est le test de Diebold et Mariano (1995) qui permet d'évaluer si les prévisions ponctuelles issues de deux modèles alternatifs sont équivalentes au regard d'une fonction de perte calculée à partir des erreurs de prévisions. Dans ce cas, le rejet de l'hypothèse nulle n'implique que la prééminence, mesurée en termes de qualité prédictive, d'un modèle par rapport à l'autre, mais ce rejet n'implique en rien que la prévision du meilleur modèle soit correctement spécifiée.

Dans ce contexte, certains tests d'évaluation des densités de prévisions permettent de comparer un nombre finis de modèles alternatifs potentiellement tous mal spécifiés par rapport à un modèle de référence qui lui-même peut être mal spécifié. Ces tests permettent ainsi de choisir parmi un ensemble de modèle le moins mauvais (ou le meilleur) par rapport à un modèle de référence, mais en aucun cas ils ne garantissent la validité (au sens de la distance avec la densité du *DGP*) de la densité de prévision obtenue à partir du meilleur modèle testé.

Nous allons à présent successivement présenter les différentes méthodes d'évaluation des prévisions propres aux prévisions ponctuelles, aux prévisions par intervalle de confiance et par densité.

4.1 Evaluation des Prévisions Ponctuelles

L'évaluation des prévisions ponctuelles se fait traditionnellement sur la base de critères construits à partir d'une séquence d'erreurs de prévisions. Le critère le plus communément utilisé correspond à l'estimateur de la variance de l'erreur de prévision, noté² *MSFE* (*Mean Squared Forecast Error*) ou de son écart type, noté *RMSFE* (*Root Mean Squared Forecast Error*). Cet estimateur est généralement construit à partir de m expériences successives³ de prévisions à l'horizon h menées conditionnellement à l'information disponible aux dates $T, T + 1, \dots, T + m$:

$$MSFE = \frac{1}{m} \sum_{j=1}^m (\hat{y}_{T+h+j|T+j} - y_{T+h+j})^2 = \frac{1}{m} \sum_{j=1}^m \hat{\varepsilon}_{T+h+j|T+j}^2 \quad (4.1)$$

où $\hat{y}_{T+h+j|t+j}$ désigne la prévision à l'horizon h réalisée conditionnellement à l'information disponible en $T + j$.

Une mesure alternative de la qualité des prévisions ponctuelles est le critère de la moyenne des erreurs absolues, généralement notée *MAFE* (*Mean Absolute Forecast Error*) ou *MAE* (*Mean Absolute Error*) :

$$MAFE = \frac{1}{m} \sum_{j=1}^m |\hat{y}_{T+h+j|T+j} - y_{T+h+j}| = \frac{1}{m} \sum_{j=1}^m |\hat{\varepsilon}_{T+h+j|T+j}| \quad (4.2)$$

Naturellement, le modèle qui possède le plus critère *MSFE* et/ou *MAFE* le plus faible est celui qui présente les meilleures propriétés prédictives. Dès lors, la question se pose de pouvoir comparer ces critères pour

²On trouve aussi les appellations *MSPE* (*Mean Squared Prediction Error*) et *RMSPE* (*Root Mean Squared Prediction Error*).

³Les paramètres du modèle de prévision peuvent être estimés de façon glissante ou de façon récurrente. Une alternative consiste à calculer la variance de l'erreur de prévision à une date de référence donnée, à partir des N prévisions simulées par la méthode de Monte-Carlo ou de Bootstrap. :

$$MSFE_t = \frac{1}{N} \sum_{i=1}^N (\hat{y}_{t+h|t}^{(i)} - y_{t+h})^2$$

où $\hat{y}_{t+h|t}^{(i)}$ est une simulation de prévision (Monte Carlo ou Bootstrap) à l'horizon h conditionnellement à l'information disponible à la date t .

deux modèles A et B concurrents. C'est précisément afin de répondre à cette question que différents tests ont été proposés depuis les travaux pionniers de Diebold et Mariano (1995). Nous présenterons successivement ce test puis une extension proposée par Harvey, Leybourne et Newbold (1997) présentant de meilleures propriétés sur petits échantillons.

Dans un second temps, nous présenterons une autre approche de l'évaluation des prévisions ponctuelles qui est spécifique aux modèles non linéaires. Il s'agit dans ce cas, non plus de mesurer directement une erreur de prévision sur l'endogène (ou la variance de cette erreur), mais d'évaluer la capacité du modèle à bien prévoir les régimes. Dans le cas où le mécanisme de transition entre régimes est fondé sur les valeurs passées de l'endogène (SETAR par exemple), les deux approches doivent fournir des évaluations compatibles.

4.1.1 Les Tests de Diebold et Mariano (1995)

Dans cette perspective de comparaison des prévisions ponctuelles, Diebold et Mariano (1995) ont proposé trois statistiques permettant de tester l'hypothèse nulle selon laquelle deux modèles concurrents présentent les mêmes qualités prédictives. La grande force de leur démarche repose sur le fait que ces statistiques peuvent être appliquées à partir d'un grand nombre de critères (*MSFE*, *MAE*, etc.) ou de fonctions de perte et que leur distribution asymptotique est établie sous des hypothèses très générales. Ainsi, avec ces tests il est possible de comparer (sur la base de n'importe quel critère) deux modèles présentant des erreurs de prévisions éventuellement non normales, de moyenne non nulle, éventuellement auto-corrélées et même corrélées entre elles.

On considère deux modèles *A* et *B* (par exemple un modèle linéaire et un modèle SETAR) fournissant *m* prévisions successives à l'horizon *h* notées respectivement $g(\hat{\varepsilon}_{t+h|t,A})$ et $g(\hat{\varepsilon}_{t+h|t,B})$ pour $t = 1, \dots, m$. Soit $g(\hat{\varepsilon}_{t+h})$ une fonction de perte associée à l'erreur de prévision $\hat{\varepsilon}_{t+h}$. Cette fonction de

perte peut être différente des critères d'évaluation usuels (*MSFE*, *MAE*) et refléter par exemple la sensibilité du décideur économique à l'asymétrie de la distribution des erreurs. Quelle que soit la fonction de perte ou le critère utilisé, l'hypothèse nulle du test de Diebold et Mariano (1995) est que les deux modèles présentent en moyenne des qualités prédictives identiques. Formellement, l'hypothèse nulle s'écrit sous la forme :

$$H_0 : E [g(\hat{\varepsilon}_{t+h|t,A})] = E [g(\hat{\varepsilon}_{t+h|t,B})] \quad (4.3)$$

ou alternativement sous la forme :

$$H_0 : E(d_t) = 0 \quad (4.4)$$

où d_t désigne l'écart des valeurs de la fonction de perte associées aux prévisions des modèles A et B , *i.e.* le différentiel de pertes, tel que :

$$d_t = g(\hat{\varepsilon}_{t+h|t,A}) - g(\hat{\varepsilon}_{t+h|t,B}) \quad (4.5)$$

Si l'on souhaite par exemple comparer les deux modèles sur la base du critère de la *MSFE* ou de la *MAE*, il suffit respectivement de poser $g(\hat{\varepsilon}_t) = \hat{\varepsilon}_t^2$ ou $g(\hat{\varepsilon}_t) = |\hat{\varepsilon}_t|$.

Diebold et Mariano proposent alors une première statistique, notée *DM*, fondée sur la moyenne empirique des différentiels de perte $\bar{d} = (1/m) \sum_{j=1}^m d_j$. Sous l'hypothèse nulle, cette moyenne converge vers une loi normale centrée et de variance égale à $\sigma_{\bar{d},0}^2/m$ où $\sigma_{\bar{d},0}^2$ désigne la variance de long terme du différentiel de perte d_t :

$$DM = \frac{\bar{d}}{\sqrt{\text{var}(\bar{d})}} = \frac{\bar{d}}{\sigma_{\bar{d},0}/\sqrt{m}} \xrightarrow{m \rightarrow \infty} N(0, 1) \quad (4.6)$$

La variance de long terme $\sigma_{\bar{d},0}^2$ peut être estimée de façon standard par un estimateur à noyau qui correspond, à un coefficient multiplicatif près, à une somme pondérée des auto-covariances avancées et retardées du processus d_j . Etant donné que pour une prévision optimale à l'horizon h , les erreurs de prévisions peuvent être au plus dépendantes à l'ordre $h - 1$,

Diebold et Mariano proposent de considérer ce résultat comme approximativement valide pour tout modèle de prévision et de ne retenir des poids strictement positifs que pour les autocovariances d'ordre inférieur ou égal à $h - 1$. Plusieurs fonctions kernel satisfont cette propriété, mais Diebold et Mariano retiennent une fonction kernel uniforme. Sous cette hypothèse, l'estimateur de la variance de long terme des différentiels de perte s'écrit tout simplement :

$$\hat{\sigma}_{\bar{d},0}^2 = \sum_{j=-(h-1)}^{h-1} \hat{\gamma}_d(j) = \hat{\gamma}_d(0) + 2 \sum_{j=1}^{h-1} \hat{\gamma}_d(j) \quad (4.7)$$

où $\hat{\gamma}_d(j) = (1/m) \sum_{i=j+1}^m (d_i - \bar{d})(d_{i-j} - \bar{d})$ désigne l'autocovariance empirique d'ordre j du processus d_t . Notons au passage que si l'on considère une prévision à l'ordre $h = 1$, un estimateur de la variance de long terme $\hat{\sigma}_{\bar{d},0}^2$ se ramène tout simplement à la variance empirique $\hat{\gamma}_0(j)$ des différentiels de pertes d_1, \dots, d_m . Dans le cas général, $\hat{\sigma}_{\bar{d},0}^2$ est un estimateur convergent de $\sigma_{\bar{d},0}^2$ et la statistique DM fondée sur $\hat{\sigma}_{\bar{d},0}^2$ converge asymptotiquement approximativement vers une loi normale centrée réduite.

Notons que ce test ne peut pas être appliqué lorsque les modèles A et B sont emboîtés (*nested models*). L'intuition est la suivante : supposons que les paramètres des deux modèles soient connus et que les données soient générées par le modèle le plus restreint, les prévisions fournies par les deux modèles sont alors strictement identiques et la distance d_t est nulle. Dans ce cas, la distribution asymptotique de la statistique DM ne peut plus être établie sous l'hypothèse nulle. Naturellement, cette dégénérescence n'existe plus lorsque les paramètres des modèles emboîtés sont estimés. Clark et McCracken (2001) proposent un test alternatif applicable aux modèles emboîtés ou non.

Diebold et Mariano (1995) ont en outre proposé dans leur étude deux autres statistiques de test qui présentent quant à elle des distributions exactes sur petits échantillons. Cette discussion sur les propriétés à distance

finie des tests n'est pas sans intérêt en pratique, car dans de nombreux cas le nombre de prévisions m disponibles est relativement faible. Ainsi, le second test proposé par Diebold et Mariano (1995) est un test de signe : l'hypothèse nulle considérée est que la médiane des différentiels de pertes est nulle

$$H_0 : med [g(\widehat{\varepsilon}_{t+h|t,A}) - g(\widehat{\varepsilon}_{t+h|t,B})] \quad (4.8)$$

Cette hypothèse nulle ne traduit pas exactement la même notion d'équivalence des qualités prédictives des modèles A et B que celle testée par la statistique DM , sauf bien évidemment, dans le cas de distributions de pertes symétriques où l'espérance et la médiane sont confondues. Pour autant, il s'agit d'une notion a priori intuitive qui implique notamment que sous H_0 la probabilité d'obtenir une perte plus forte avec le modèle A (comparativement au modèle B) doit être identique à la probabilité d'obtenir une perte plus forte avec le modèle B (comparativement au modèle A), c'est-à-dire :

$$\Pr [g(\widehat{\varepsilon}_{t+h|t,A}) > g(\widehat{\varepsilon}_{t+h|t,B})] = \Pr [g(\widehat{\varepsilon}_{t+h|t,A}) < g(\widehat{\varepsilon}_{t+h|t,B})] \quad (4.9)$$

Si l'on suppose que les différentiels de pertes sont *i.i.d.*, le nombre de différentiels positifs suit une loi binomiale de paramètres m et 0.5 sous l'hypothèse nulle. Soit DM_2 la statistique de test associée, cette dernière vérifie⁴ :

$$DM_2 = \sum_{t=1}^m \mathbb{I}_{(d_t > 0)} \underset{H_0}{\sim} B\left(m, \frac{1}{2}\right) \quad (4.11)$$

où $\mathbb{I}_{(d_t > 0)}$ désigne la fonction indicatrice valant 1 si $d_t > 0$ et 0 sinon. Dans la même perspective, Diebold et Mariano (1995) construisent un troisième test de Wilcoxon fondé sur les rangs et les signes des différentiels de perte.

⁴Sur grands échantillons, la version de Student de cette statistique est asymptotiquement distribuée selon une loi normale centrée réduite :

$$\frac{DM_2 - m/2}{\sqrt{m/4}} \xrightarrow[m \rightarrow \infty]{d} N(0, 1) \quad (4.10)$$

Dans le cas d'échantillons de différentiels non *i.i.d.*, Diebold et Mariano (1995) proposent une stratégie de test alternative.

Dans ce cas, les auteurs supposent que la distribution des différentiels est symétrique, centrée et *i.i.d.* La statistique de test, notée DM_3 , correspond à la somme des rangs des valeurs absolues des observations positives :

$$DM_3 = \sum_{t=1}^m \mathbb{I}_{(d_t > 0)} \text{rang}(|d_t|) \quad (4.12)$$

Cette statistique admet une distribution exacte et invariante à la distribution des différentiels de pertes qui est tabulée⁵.

4.1.2 Les Tests de Harvey, Leybourne et Newbold (1997)

Parmi les trois approches présentées précédemment, c'est sans conteste le test asymptotique fondé sur la statistique DM qui est le plus souvent retenu dans les applications empiriques des tests de Diebold et Mariano (1995). Pourtant, Harvey, Leybourne et Newbold (1997) montrent que cette statistique a tendance sur petits échantillons à présenter une taille empirique largement supérieure à la taille nominale. Ces rejets à tort sont encore plus fréquents lorsque l'horizon h de la prévision augmente. C'est pourquoi les auteurs ont proposé⁶ en conséquence une version modifiée de cette statistique qui permet de corriger ce défaut. Pour le reste, la logique du test est exactement le même que celle du test de Diebold et Mariano (1995).

La correction proposée par Harvey, Leybourne et Newbold est liée tout d'abord au fait qu'ils considèrent l'expression exacte de la variance de long terme de la moyenne empirique des différentiels $\text{var}(\bar{d})$ et non la variance asymptotique de cette moyenne empirique comme chez Diebold et Mariano

⁵La version de Student de cette statistique vérifie :

$$\frac{DM_3 - m(m+1)/4}{\sqrt{m(m+1)(2m+1)/24}} \xrightarrow{m \rightarrow \infty} N(0,1) \quad (4.13)$$

⁶Plus précisément, les auteurs proposent deux tests concurrents : l'un fondé sur une statistique de Diebold Mariano modifiée et l'autre sur une statistique de Morgan-Granger-Newbold (Granger et Newbold, 1977) modifiée. Cette dernière est fondée sur l'étude de la corrélation de $x = \widehat{\varepsilon}_{t+h|t,A} - \widehat{\varepsilon}_{t+h|t,B}$ et de $z = \widehat{\varepsilon}_{t+h|t,A} + \widehat{\varepsilon}_{t+h|t,B}$. Ils recommandent finalement l'utilisation de la statistique de Diebold Mariano modifiée, notamment en raison du manque de robustesse du second type de test en présence de queues de distribution épaisse dans la distribution des différentiels d'erreurs de prévisions.

(1995). La seconde différence tient au fait qu'ils estiment cette variance de long terme à partir d'estimateurs des autocovariances $\hat{\gamma}_d$ obtenus à partir d'une correction pour petit échantillon. Ainsi, leur estimateur de la variance de long terme s'écrit au final :

$$var_{est}^*(\bar{d}) = \frac{1}{m} \left[\hat{\gamma}_d^*(0) + \frac{2}{m} \sum_{j=1}^{h-1} (m-j) \hat{\gamma}_d^*(j) \right] \quad (4.14)$$

avec $\hat{\gamma}_d^*(j) = (m-j)^{-1} \sum_{i=j+1}^m (d_i - \bar{d})(d_{i-j} - \bar{d})$. En supposant que l'ordre j des autocovariances ($j \leq h-1$) est relativement petit comparativement au nombre de prévisions disponibles m , les auteurs montrent que cette estimateur est biaisé sur petit échantillon (mais asymptotiquement non biaisé) puisque pour un horizon h donné :

$$E[var_{est}^*(\bar{d})] \simeq \left[\frac{m+1-2h}{m} + \frac{h(h-1)}{m^2} \right] \times var(\bar{d}) \quad (4.15)$$

Harvey, Leybourne et Newbold proposent dès lors d'utiliser un estimateur non biaisé de la variance dans la construction d'une statistique de Diebold et Mariano modifiée, noté DM^* , telle que :

$$DM^* = \frac{\bar{d}}{\sqrt{var_{est}^*(\bar{d})}} \simeq \left[\frac{m+1-2h}{m} + \frac{h(h-1)}{m^2} \right]^{\frac{1}{2}} DM \quad (4.16)$$

où DM désigne la statistique de Diebold et Mariano définie par l'équation (4.6) et où $\bar{d} = (1/m) \sum_{j=1}^m d_j$ avec $d_j = g(\hat{\varepsilon}_{j+h|j,A}) - g(\hat{\varepsilon}_{j+h|j,B})$. Les auteurs montrent que dans différentes expériences, cette correction permet de réduire très sensiblement la fréquence de rejet à tort sous l'hypothèse nulle d'égalité des modèles, même si elle ne permet pas totalement de corriger ce problème. La statistique DM^* a souvent une taille empirique supérieure à la taille nominale désirée (*oversized test*).

4.1.3 Les Tests Non-Paramétriques de Prévisions de Régime : Pesaran et Timmerman (1992)

Une façon alternative d'évaluer la qualité d'un modèle de prévision à changement de régimes consiste précisément à étudier sa capacité à prévoir

les régimes futurs, qui dans le cas de modèle STAR dépendent du niveau anticipé de l'endogène. Cette idée a été reprise notamment par Pesaran et Timmermann (1992) qui proposent un test visant à évaluer la validité des changements de direction prévus (*predicted directions of change*) pour une variable donnée.

Considérons un modèle de type STAR à deux régimes admettant pour fonction de transition $G(s_t; \gamma, c)$, où s_t désigne une variable de transition. On pose :

$$\widehat{r}_{T+j+h|T+j} = \begin{cases} 1 & \text{si } G(\widehat{s}_{T+j+h|T+j}; \gamma, c) > 0.5 \\ -1 & \text{si } G(\widehat{s}_{T+j+h|T+j}; \gamma, c) \leq 0.5 \end{cases} \quad (4.17)$$

où $\widehat{s}_{T+j+h|T+j}$ désigne la prévision de la variable de transition à l'horizon h réalisée conditionnellement à l'information disponible à la date $t+h$. A l'instar de ce que l'on obtient dans un modèle dichotomique (Logit, Probit, etc.), la variable $\widehat{r}_{T+j+h|T+j}$ peut être interprétée comme la prévision du régime à l'horizon h obtenue à partir de la prévision de la variable de transition $\widehat{s}_{T+j+h|T+j}$. On note $r_{T+j+h|T+j}$ la variable dichotomique définie comme dans l'équation (4.17), à l'exception près que cette prévision de régime est établie à partir de l'observation de la variable de transition s_{T+j+h} et non de sa prévision comme dans le cas précédent. On définit alors un ratio de succès⁷, noté RS , par la fréquence d'observations d'une prévision $\widehat{r}_{T+j+h|T+j}$ conforme à la prévision $r_{T+j+h|T+j}$, obtenue sur m périodes successives :

$$RS = \frac{1}{m} \sum_{j=1}^m \mathbb{I}(\widehat{r}_{T+j+h|T+j} \times r_{T+j+h|T+j} > 0) \quad (4.18)$$

où $\mathbb{I}(\cdot)$ désigne la fonction indicatrice.

L'idée du test consiste alors à mesurer la distance entre le ratio de succès d'un modèle et le ratio de succès que l'on obtiendrait dans le cas où les

⁷Ce critère correspond à une variante près au critère de la méthode d'estimation non paramétrique des modèles dichotomiques *LAD* (*Least Absolute Deviation*) proposée par Powell (1984).

prévisions de régimes établies respectivement sur la base de l'observation de la variable de transition et de sa prévision seraient indépendamment distribuées. Il s'agit donc de déterminer si le modèle considéré (plus précisément la prévision de la variable de transition) conduit à faire mieux en termes de prévisions de régimes qu'un modèle dans lequel la prévision de la variable de transition serait totalement indépendante de la réalisation *ex-post*; modèle qui convenons-le, constituerait le comble des constructions possibles que pourrait réaliser un économètre. Pour ce faire, il convient de calculer ce que serait le ratio de succès dans une telle configuration. On montre aisément que sous l'hypothèse d'indépendance, le ratio de succès vérifie :

$$RS_{IND} = P \hat{P} + (1 - P) (1 - \hat{P}) \quad (4.19)$$

où les probabilités \hat{P} et P sont définies par :

$$\hat{P} = \frac{1}{m} \sum_{j=1}^m \mathbb{I}(\hat{r}_{T+j+h|T+j} = 1) \quad P = \frac{1}{m} \sum_{j=1}^m \mathbb{I}(r_{T+j+h|T+j} = 1) \quad (4.20)$$

La variance du ratio de succès est alors égale à :

$$\begin{aligned} var(RS_{IND}) &= \frac{1}{m} (2\hat{P} - 1)^2 P(1 - P) \\ &\quad + \frac{1}{m} (2P - 1)^2 \hat{P} (1 - \hat{P}) + \frac{4}{m^2} P\hat{P}(1 - P) (1 - \hat{P}) \end{aligned} \quad (21)$$

L'hypothèse nulle du test est que les prévisions de régime $\hat{r}_{T+j+h|T+j}$ et $r_{T+j+h|T+j}$ sont indépendamment distribuées :

$$H_0 : RS = RS_{IND} \quad (4.22)$$

Sachant que, dans le cas général, la variance du ratio de succès est égale à $var(RS) = RS_{IND}(1 - RS_{IND})/m$, on peut alors construire un test d'égalité des ratios de succès, noté TR , de la façon suivante :

$$TR = \frac{RS - RS_{IND}}{\sqrt{var(RS) - var(RS_{IND})}} \quad (4.23)$$

Sous l'hypothèse nulle, Pesaran et Timmermann (1992) montrent que cette statistique est asymptotiquement distribuée selon une loi normale centrée réduite. Une façon similaire de tester la qualité des prévisions, *via* la prévision des régimes, consiste à construire un tableau de contingence des régimes prévus et réalisés et de réaliser tout simplement un test d'indépendance du Chi-deux. Ce test peut alors être étendu naturellement à des modèles à plus de deux régimes (Pesaran et Timmermann, 1994).

4.1.4 Tests de Prévisions Enveloppées (*Forecast Encompassing*) : Harvey, Leybourne et Newbold (1998)

Une autre façon de comparer les prévisions ponctuelles de deux modèles consiste à retenir une approche de type "prévisions enveloppées" (*forecast encompassing*). De façon générale, la méthode dite de l'enveloppement (Mizon et Richard, 1986) consiste à évaluer la capacité d'un modèle à expliquer ce que l'autre modèle teste. Ainsi, le modèle A enveloppe le modèle B si les faits du modèle B peuvent être expliqués par le modèle A mais pas l'inverse. Dans le cas d'hypothèses emboîtées, l'hypothèse nulle peut s'écrire comme une restriction sur H_1 . La méthode d'enveloppement consiste alors à créer une sorte d'emboîtement artificiel des deux modèles et à tester un ensemble de restrictions sur cet emboîtement.

Dans le cas des prévisions, la logique est la même. Soient deux modèles concurrents A et B . On cherche à tester si les prévisions du modèle A enveloppent celles du modèle B , *i.e.* si les prévisions du modèles B n'apportent aucune information supplémentaire sur la formation des prévisions du modèle A . On note $\hat{y}_{t+h|t,A}$ et $\hat{\varepsilon}_{t+h|t,A}$ (respectivement $\hat{y}_{t+h|t,B}$ et $\hat{\varepsilon}_{t+h|t,B}$) les prévisions et erreurs de prévision à un ordre h associées au modèle A (respectivement B). Les tests d'enveloppement sont fondés sur une régression linéaire du type :

$$y_{t+h} - \hat{y}_{t+h|t,B} = \alpha (y_{t+h} - \hat{y}_{t+h|t,A}) + v_t \quad t = 1, \dots, m \quad (4.24)$$

où v_t est *i.i.d.* Sous l'hypothèse nulle que les prévisions du modèle A

enveloppent les prévisions du modèle B , le coefficient α doit être nul. Autrement dit, l'erreur de prévision du modèle A ne peut pas expliquer une partie de celle du modèle B , l'inverse n'étant pas vrai. Il suffit donc de tester l'hypothèse nulle $H_0 : \alpha = 0$ pour tester l'enveloppement de B par A .

Généralement, les tests d'enveloppement sont construits à partir de la forme équivalente suivante :

$$y_{t+h} = (1 + \alpha) \hat{y}_{t+h|t,A} + \alpha \hat{y}_{t+h|t,B} + v_t \quad (4.25)$$

Dans les deux cas, la statistique de test se ramène à une statistique de Student associée à l'hypothèse nulle $\alpha = 0$. Ces tests d'enveloppement ont été appliqués dans le cas de modèles linéaires macro-économiques notamment par Chong et Hendry (1986). Harvey, Leybourne et Newbold (1998) ont montré que ces tests pouvaient être largement *oversized* sur petits échantillons notamment lorsque les erreurs de prévision ne sont pas normalement distribuées. Aussi, ont-ils proposé un certain nombre de statistiques alternatives parmi lesquelles figure la statistique de Student modifiée t^* suivante :

$$t^* = \frac{\hat{\alpha}}{\tilde{\sigma} \sqrt{m}} \sum_{j=1}^m (\hat{\varepsilon}_{t+h|t,A} - \hat{\varepsilon}_{t+h|t,B})^2 \quad (4.26)$$

où $\hat{\alpha}$ désigne l'estimateur *MCO* du paramètre α et où $\tilde{\sigma}^2$ correspond à la variance estimée des différentiels d'erreurs de prévisions. Mais contrairement au cas standard, ces différentiels sont pondérés par les résidus estimés \hat{v}_t de l'équation (4.24) ou de l'équation (4.25) suivant le modèle retenu :

$$\tilde{\sigma}^2 = \frac{1}{m} \sum_{j=1}^m (\hat{\varepsilon}_{t+h|t,A} - \hat{\varepsilon}_{t+h|t,B})^2 \hat{v}_t^2 \quad (4.27)$$

Cette statistique suit asymptotiquement une distribution normale centrée réduite. Harvey, Leybourne et Newbold (1998) montrent en outre que le test d'égalité des *MSFE* de Diebold et Mariano (1995) peut aussi s'interpréter comme un test d'enveloppement.

4.1.5 *Le Test d'Échec de Prévision : Clements et Hendry (2002)*

En parallèle de tous ces tests de validation qui permettent de comparer deux modèles de prévisions, Clements et Hendry (2002) proposent un test permettant d'évaluer les échecs de prévision (*forecast failure*) d'un modèle particulier. L'idée consiste à postuler qu'un modèle est mauvais en termes prédictifs si la variance de l'erreur de prévision hors période d'estimation (*out of sample*) est statistiquement supérieure à la variance des erreurs du modèle sur la période d'estimation (*in sample*). Ainsi, Clements et Hendry proposent de construire une statistique d'échec de prévision définie par :

$$Q = \frac{\widehat{\varepsilon}_{T+1|T}^2}{\widehat{\sigma}_T^2} \left(\frac{T}{T-k} \right) \quad (4.28)$$

où $\widehat{\varepsilon}_{T+1|T}$ désigne l'erreur de prévision à l'horizon $h = 1$. Cette prévision est établie conditionnellement à l'information disponible à la date T sur la base d'un modèle estimé sur les observations disponibles aux dates $t = 1, \dots, T$. Le terme $\widehat{\sigma}_T^2$ désigne alors l'estimateur usuel de la variance des erreurs sur la période d'estimation (*in sample*). Enfin, le terme k désigne le nombre de paramètres du modèle. Sous l'hypothèse nulle que le modèle est un bon modèle prédictif, la variance de l'erreur de prévision hors période d'estimation doit être égale à la variance des résidus sur la période d'estimation, et dès lors :

$$Q \xrightarrow[m \rightarrow \infty]{d} F(1, T - k - 1) \quad (4.29)$$

Notons que la logique du test de Clements et Hendry (2002) est sensiblement différente de la logique des tests présentés précédemment puisqu'il ne s'agit pas ici de comparer deux modèles concurrents en termes de prévision, mais plutôt de tester si un modèle particulier peut être jugé "*bon*" en termes de prévisions.

4.2 Evaluation des Intervalles de Confiance

Dans de nombreuses applications, l'évaluation des qualités prédictives d'un modèle ne peut se limiter au seul examen de l'espérance conditionnelle et même de la variance conditionnelle. C'est particulièrement vrai en finance comme par exemple dans le cas de la prévision de la *Value-at-Risk*, mais aussi dans de nombreuses autres problématiques. De plus, il apparaît évident que lorsque l'on dispose d'un estimateur de la densité de prévision (comme c'est le cas par exemple à la suite d'une prévision de type Monte Carlo ou Bootstrap), il est totalement réducteur d'évaluer les capacités prédictives d'un modèle sur la base d'un seul moment conditionnel. La première tentative de dépasser cette contradiction consiste à évaluer un modèle en utilisant des intervalles de confiance associés aux prévisions hors période d'estimation (*out of sample*). Cette approche a notamment été initiée par les travaux pionniers de Christoffersen (1998).

Christoffersen part du constat que s'il existe des critères d'évaluation des prévisions ponctuelles, il n'existe en revanche pas de critère ou de test permettant d'évaluer la qualité des intervalles de confiance associés à ces prévisions. C'est pourquoi il propose un ensemble de définitions relatives à la validité des intervalles de confiance et en déduit différents tests. Il convient de noter que tester la validité d'un intervalle de confiance associé à une prévision est une approche qui se situe à mi-chemin entre l'évaluation des prévisions ponctuelles et de l'évaluation de la densité des prévisions. On tient compte de l'incertitude autour de la prévision ponctuelle dans un certain intervalle, mais pour autant on ne va pas jusqu'à évaluer l'ensemble de la densité de la prévision.

Sur le plan technique, la démarche de Christoffersen s'inscrit exactement dans la même logique que celle de Diebold et Mariano (1995) dans le sens où il propose un ensemble de tests qui ne dépendent pas du modèle utilisé pour faire la prévision (*model free*) et de la manière dont sont construits⁸

⁸Pour une synthèse sur les différentes manières de construire un intervalle de confiance

les intervalles de confiance associés aux prévisions. Mais une différence notable réside dans le fait que Diebold et Mariano proposent des tests de comparaison de modèles fondés sur les prévisions ponctuelles, alors que Christoffersen propose différents tests de l'hypothèse nulle de validité d'un intervalle de confiance de prévision construit à partir d'un modèle donné.

Comment juger de la validité d'un intervalle de confiance ? L'idée de base consiste à évaluer si les valeurs *ex-post* de la variable prévues sont comprises au sein des intervalles construits *ex-ante* dans une proportion égale au niveau de risque nominal considéré. Cette notion de couverture avait déjà été proposée par McNees (1995), mais aussi dans les travaux pionniers de Kupiec (1995) sur les méthodes de la validation de la Value-at-Risk (*VaR*). La *VaR* est une mesure du risque extrême qui est en effet définie comme le fractile de la distribution des rendements d'un actif financiers. En ce sens, la *VaR* n'est rien d'autre que la borne supérieure d'un intervalle de confiance défini sur les rendements de l'actif. Mais le principal apport de Christoffersen (1998) par rapport à ces travaux originaux est qu'il étend cette notion de couverture à un cadre conditionnel.

Afin de mieux appréhender l'apport de Christoffersen, considérons les cas d'un intervalle de prévision associé à un modèle de type ARMA avec une spécification de type GARCH des résidus. On sait que dans ce cas, l'intervalle de confiance sur les prévisions doit être construit à partir d'un estimateur de la variance conditionnelle des résidus et non plus à partir d'un estimateur de la variance non conditionnelle comme dans le cas standard. L'amplitude de cet intervalle est alors variable dans le temps : dans les périodes de faible volatilité (conditionnelle) les prévisions sont précises et l'amplitude de l'intervalle est faible, mais dans les périodes agitées, la volatilité anticipée est forte et l'amplitude de l'intervalle de confiance s'accroît. Pour ce même processus, si l'on considère au contraire une séquence d'intervalles de confiance de même amplitude (fondés par exemple sur la variance non

voir Chatfield (1993).

conditionnelle), il se peut qu'en moyenne l'hypothèse de couverture non conditionnelle soit valide : la proportion d'observations de la série au sein de cet intervalle constant peut être égale a niveau de risque nominal. En revanche, conditionnellement à l'information passée, la couverture dite conditionnelle ne peut pas être réalisée : on devrait au contraire observer des paquets ou successions (*clusters*) de violations de cet intervalle de confiance, notamment dans les périodes de forte volatilité.

4.2.1 Définition de la couverture conditionnelle

Soit une séquence d'observations $\{y_t\}_{t=1}^T$ d'un processus y_t . On considère une séquence d'intervalles de confiance, pour un niveau de risque de $\alpha\%$, associées aux prévisions hors période d'estimation (*out of sample*) de y_t réalisées conditionnellement à l'information disponible à la date $t - 1$. On note respectivement $L_{t|t-1}(\alpha)$ et $U_{t|t-1}(\alpha)$ les bornes inférieures et supérieures de ces intervalles de confiance successifs pour $t = 1, \dots, T$. Soit I_t (ou *hit function*) la variable dichotomique associée à la violation de cet intervalle : cette dernière prend la valeur 1 lorsque la réalisation *ex-post* de y_t n'est pas comprise dans l'intervalle de confiance prévu *ex-ante* conditionnellement à l'information disponible en $t - 1$, et 0 sinon :

$$I_t = \begin{cases} 1 & \text{si } y_t \notin [L_{t|t-1}(\alpha), U_{t|t-1}(\alpha)] \\ 0 & \text{si } y_t \in [L_{t|t-1}(\alpha), U_{t|t-1}(\alpha)] \end{cases} \quad (4.30)$$

Christoffersen propose alors la définition suivante : la séquence des intervalles de confiance $\{L_{t|t-1}(\alpha), U_{t|t-1}(\alpha)\}_{t=1}^T$ satisfait la propriété de *couverture conditionnelle* par rapport à l'ensemble d'information Ω_{t-1} si et seulement si :

$$E[I_t | \Omega_{t-1}] = \alpha \quad \forall t = 1, \dots, T \quad (4.31)$$

Naturellement, on peut déduire de cette définition en posant $\Omega_{t-1} = \emptyset$, une condition de *couverture non conditionnelle*, similaire à celle utilisée par McNees (1995) ou Kupiec (1995), qui revient à évaluer si $E[I_t] = \alpha$, $\forall t$. Un test simple de l'hypothèse de couverture non conditionnelle peut alors

être construit à partir de l'estimateur de l'espérance non conditionnelle $\sum_{t=1}^T I_t/T$ (Kupiec, 1995).

De plus lorsque l'on restreint l'ensemble d'information Ω_{t-1} à l'ensemble des valeurs passées de la *hit function* $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, \dots, I_1\}$, Christoffersen montre que le test de couverture conditionnel $E[I_t | \Omega_{t-1}] = \alpha$ est équivalent au test de l'hypothèse selon laquelle les processus de violations $\{I_t\}_{t=1}^T$ sont identiquement et indépendamment distribués selon une loi de Bernoulli de paramètre p :

$$\{I_t\}_{t=1}^T \underset{H_0}{\sim} \text{Bernoulli}(p) \quad (4.32)$$

4.2.2 Tests de Ratios de Vraisemblance : Christoffersen (1998)

Christoffersen (1998) propose trois tests de ratio de vraisemblance : un test de couverture non conditionnelle, un test d'indépendance des violations et un test de couverture conditionnelle.

(a) Test de couverture non conditionnelle

Au test de l'hypothèse nulle de couverture non conditionnelle (Kupiec, 1995) $H_0 : E[I_t] = \alpha$ correspond une statistique LR_{UC} vérifiant :

$$LR_{UC} = -2 [\log L(\alpha, I_t) - \log L(\hat{\pi}, I_t)] \xrightarrow[T \rightarrow \infty]{d} \chi^2(1) \quad (4.33)$$

où $L(\alpha, I_t)$ et $L(\pi, I_t)$ désignent respectivement la vraisemblance de la séquence $\{I_t\}_{t=1}^T$ sous l'hypothèse nulle et sous l'hypothèse alternative $E[I_t] = \pi \neq \alpha$:

$$L(\alpha, I_t) = (1 - \alpha)^{n_0} \alpha^{n_1} \quad (4.34)$$

$$L(\hat{\pi}, I_t) = (1 - \hat{\pi})^{n_0} \hat{\pi}^{n_1} \quad (4.35)$$

avec $\hat{\pi} = n_1/(n_0 + n_1)$ l'estimateur du maximum de vraisemblance de π avec $n_1 = \sum_{t=1}^T I_t$ et $n_0 = T - n_1$.

(b) Test d'indépendance des violations

De la même façon, Christoffersen (1998) propose un test d'indépendance des violations de l'intervalle de confiance contre une alternative très spécifique de chaîne de Markov de premier ordre. On suppose ici que les violations I_t satisfont une représentation de type chaîne de Markov avec deux états (1 ou 0) et pour matrice de transition :

$$\Pi_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix} \quad (4.36)$$

où $\pi_{ij} = \Pr(I_t = j | I_{t-1} = i)$. Sous l'hypothèse nulle que les violations I_t sont indépendamment distribuées dans le temps, cette matrice devient :

$$\Pi_0 = \begin{bmatrix} 1 - \pi_2 & \pi_2 \\ 1 - \pi_2 & \pi_2 \end{bmatrix} \quad (4.37)$$

Ainsi quel que soit l'état initial à la date $t - 1$ du système (violation ou non de l'intervalle de confiance), la réalisation de la variable t exactement la même probabilité π_2 d'appartenir à l'intervalle de confiance à la période t . Dès lors, une statistique de ratio de vraisemblance, notée LR_{IND} , permet très simplement de tester l'hypothèse nulle d'indépendance des violations.

$$LR_{IND} = -2 \left[\log L \left(\widehat{\Pi}_0, I_t \right) - \log L \left(\widehat{\Pi}_1, I_t \right) \right] \xrightarrow[T \rightarrow \infty]{d} \chi^2(1) \quad (4.38)$$

Si l'on note n_{ij} le nombre d'observations du couple $I_t = j$ et $I_{t-1} = i$, on montre que la vraisemblance sous l'hypothèse nulle s'écrit :

$$\log L \left(\widehat{\Pi}_0, I_t \right) = (1 - \widehat{\pi}_2)^{(n_{00} + n_{10})} \widehat{\pi}_2^{(n_{01} + n_{11})} \quad (4.39)$$

où $\widehat{\pi}_2 = (n_{01} + n_{11}) / (n_{00} + n_{01} + n_{10} + n_{11})$ désigne l'estimateur du maximum de vraisemblance de π_2 . De la même façon sous l'hypothèse alternative, la vraisemblance de la séquences des violations $\{I_t\}_{t=1}^T$ s'écrit :

$$\log L \left(\widehat{\Pi}_1, I_t \right) = (1 - \widehat{\pi}_{01})^{n_{00}} \widehat{\pi}_{01}^{n_{01}} (1 - \widehat{\pi}_{11})^{n_{10}} \widehat{\pi}_{11}^{n_{11}} \quad (4.40)$$

avec :

$$\widehat{\Pi}_1 = \begin{bmatrix} 1 - \widehat{\pi}_{01} & \widehat{\pi}_{01} \\ 1 - \widehat{\pi}_{11} & \widehat{\pi}_{11} \end{bmatrix} = \begin{bmatrix} \frac{n_{00}}{n_{00} + n_{01}} & \frac{n_{01}}{n_{00} + n_{01}} \\ \frac{n_{10}}{n_{10} + n_{11}} & \frac{n_{11}}{n_{10} + n_{11}} \end{bmatrix} \quad (4.41)$$

(c) Test de couverture conditionnelle

Enfin, Christoffersen (1998) propose un test de l'hypothèse nulle de couverture conditionnelle. Il s'agit alors tout simplement d'un test joint de l'hypothèse de couverture non conditionnelle et d'indépendance des violations. La statistique de ratio de vraisemblance, notée LR_{CC} , s'écrit alors :

$$LR_{CC} = -2 \left[\log L(\alpha, I_t) - \log L(\hat{\Pi}_1, I_t) \right] \xrightarrow{T \rightarrow \infty} \chi^2(2) \quad (4.42)$$

Par construction, on vérifie bien évidemment que $LR_{CC} = LR_{UC} + LR_{IND}$.

Les trois tests de Christoffersen (1998) permettent ainsi de porter un diagnostic fin quant à la validité des intervalles de confiance associés aux prévisions en répondant finalement à deux questions essentielles :

- les valeurs réalisées de la variable ont-elles été observées (en moyenne) à l'intérieur de mes différents intervalles de confiance dans une proportion égale à α ?
- les apparitions de « mauvaises » prévisions (c'est-à-dire hors intervalle de confiance) sont-elles liées dans le temps ?

Pour autant, ces tests ne portent pas sur la distribution complète de densité des prévisions. En cela ils s'opposent aux tests d'évaluation par densité.

4.3 Evaluation des Densités de Prévisions

La présentation de cette section reprend le plan de la présentation des méthodes d'évaluation des densités de prévision proposé par Corradi et Swanson (2006) dans le *Handbook of Economic Forecasting*.

4.3.1 Tests de Spécification Correcte

Les premiers tests d'évaluation des densités de prévisions se sont attachés à déterminer si les réalisations des prévisions étaient issues d'une distribution identique à la "vraie" distribution des données, c'est-à-dire à la densité issue du processus générateur des données (DGP). Ce sont donc avant tout des tests de spécification "correcte" de la densité de prévision.

Le test pionnier dans cette perspective est sans conteste celui de Diebold, Gunther et Tay (1998) qui repose sur la transformation probabiliste (*probability integral transform*) de Rosenblatt (1952). Rappelons le principe de cette transformation. Soit $f_t(y|\Omega_{t-1}, \theta_0)$ la densité conditionnelle des prévisions (à l'horizon $h = 1$) de la variable y_t obtenue dans le modèle de référence (AR, SETAR, ou autre) conditionnellement à l'ensemble d'information Ω_{t-1} disponible à la date $t-1$ et où θ_0 désigne un ensemble de paramètres connus. Soit $\{y_t\}_{t=1}^T$ la séquence des réalisations de la variable y sur la période d'évaluation des prévisions du modèle. Sous l'hypothèse que la distribution conditionnelle des prévisions associée à ce modèle corresponde effectivement au processus générateur des données, alors les variables transformées :

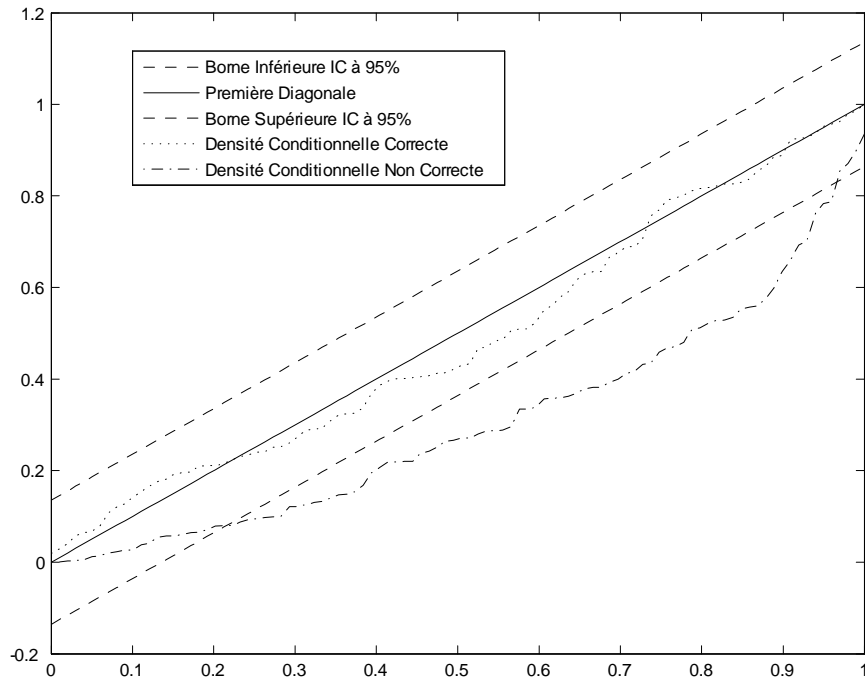
$$z_t = F_t(y_t|\Omega_{t-1}, \theta_0) = \int_{-\infty}^{y_t} f_t(u|\Omega_{t-1}, \theta_0) du \quad t = 1, \dots, T \quad (4.43)$$

sont identiquement et indépendamment distribuées selon une loi uniforme sur $[0, 1]$. Autrement dit, lorsque à chaque date t , la densité conditionnelle des prévisions associée au modèle testé correspond à la "vraie" distribution conditionnelle associée au processus générateur de données, la séquence des variables transformées $\{z_t\}_{t=1}^T$ est *i.i.d.* $U_{[0,1]}$. Par conséquent, dans cette perspective, une manière évidente d'évaluer si la spécification de la densité conditionnelle des prévisions est correcte revient à tester l'adéquation de la distribution des variables transformées z_t à une loi uniforme.

Diebold, Gunther et Tay (1998) proposent ainsi de comparer sur un même graphique la fonction de répartition empirique des variables trans-

formées et la fonction de répartition théorique de la loi uniforme, *i.e.* la droite à 45°. Comme cela est illustré sur la Figure (4.1), si le modèle utilisé pour effectuer la prévision ne permet pas d'obtenir une spécification correcte de la densité conditionnelle des prévisions, on doit alors observer un écart important entre les deux fonctions de répartition.

Fig. 4.1 Fonctions de Répartition Empirique des Variables Transformées z_t associées à deux Modèles de Prévision



Cette démarche a notamment été utilisée par Diebold, Gunther et Tay (1998) dans le domaine du management des risques financiers, par Diebold, Tay et Wallis (1998) pour les prévisions d'inflation, par Clements et Smith (2000) pour les prévisions sur la production et le chômage ainsi que par Clements et Smith (2001) pour les prévisions sur les taux de change. Ainsi,

en utilisant une approche graphique de ce type, Clements et Smith (2001) évaluent la capacité respective d'un modèle de type marche aléatoire et d'un modèle de type SETAR à générer une densité de prévision conditionnelle correcte pour le taux de change DM - Dollar et pour le taux de change yen-Dollar. Ils montrent ainsi que la fonction de répartition empirique associée aux variables transformées construites à partir des prévisions obtenues par un modèle de type marche aléatoire diffère sensiblement d'une droite à 45° . En revanche, il n'existe pratiquement pas de différence avec cette droite dans le cas de prévisions générées par un modèle SETAR. Ces observations illustrent le fait qu'il est difficile de rejeter l'hypothèse selon laquelle la densité de prévision (horizon $h = 1$ et/ou $h = 2$) associée au modèle SETAR est correcte, alors que cette hypothèse est clairement violée dans le cas d'un modèle de type marche aléatoire. Ces résultats contrastent avec ceux généralement obtenus dans le cadre du débat sur l'utilisation de modèles linéaires / non linéaires pour la prévision des taux de change, puisqu'il est aujourd'hui convenu que l'utilisation de modèles non linéaires n'apporte pas grand chose par rapport aux modèles linéaires si l'on s'en réfère aux seuls critères d'évaluation des prévisions ponctuelles (*MSFE*, *MAFE*, etc.).

Une première façon de tester si la densité de prévision est correcte consiste à construire un intervalle de confiance à $1 - \alpha\%$ autour de la fonction de répartition théorique de la loi uniforme. Clements et Smith (2001) utilisent pour cela les valeurs critiques exactes de la statistique de Kolmogorov pour des petits échantillons de taille T . Pour un intervalle de confiance à 95% ($\alpha = 0.025$), cet intervalle est donné⁹ par la droite à $45^\circ \mp \sqrt{\ln(1/\alpha)/(2T)}$.

Une seconde façon consiste à utiliser directement un test non paramétrique d'adéquation de loi, comme par exemple le test de Kolmogorov-Smirnov. Le principe est alors le même : il s'agit alors de tester l'adéquation entre la répartition $F_t(y \mid \Omega_{t-1}, \theta_0)$ et la fonction de répartition d'une loi uniforme sur $[0, 1]$. Mais deux problèmes se posent à ce niveau

⁹Pour plus de détails, voir Miller (1956).

là. Tout d'abord, les paramètres θ_0 du modèle utilisé pour effectuer la prévision ne sont généralement pas connus mais estimés. L'erreur d'estimation de ces paramètres vient alors "polluer" la distribution conditionnelle des prévisions, ce qui peut notamment induire un rejet à tort de l'hypothèse nulle de spécification correcte. Sur le plan technique, la distribution asymptotique des tests usuels d'adéquation de loi (Kolmogorov-Smirnov) dépend alors de paramètres de nuisance. De plus, l'ensemble d'information Ω_{t-1} n'est généralement pas observé dans son intégralité, seul un sous ensemble tronqué est observé. C'est pourquoi, Bai (2003) propose un test général de spécification correcte d'une distribution conditionnelle qui permet en particulier de répondre à ces deux problèmes et qui peut être appliqué dans le contexte des densités conditionnelles de prévision.

4.3.1.1 Le Test de Bai (2003)

De façon générale, l'hypothèse nulle du test de Bai (2003) est la suivante : la distribution cumulative conditionnelle de la variable y_t est dans la famille paramétrique $F_t(u|\Omega_{t-1}, \theta_0)$. Le principe du test de Bai (2003) reste alors similaire au test de Kolmogorov : il s'agit de vérifier l'adéquation de la fonction de répartition de la séquence des processus $z_t = F_t(y_t|\Omega_{t-1}, \theta_0)$ avec celle d'une loi uniforme sur $[0, 1]$ pour $t = 1, \dots, T$. Mais contrairement au cas standard, les paramètres θ_0 ne sont pas connus mais seulement estimés. De plus, Bai suppose que seul un sous ensemble de l'ensemble d'information est observable. Soit $\tilde{\Omega}_{t-1} = \{y_{t-1}, \dots, y_1, z_{t-1}, \dots, z_1\}$ un sous ensemble observable et tronqué de Ω_{t-1} . On note $\hat{\theta}$ l'estimateur de θ_0 . Les variables transformées établies à partir des paramètres estimés $\hat{\theta}$ sont notées \hat{z}_t et vérifient :

$$\hat{z}_t = F_t\left(y_t|\tilde{\Omega}_{t-1}, \hat{\theta}\right) \quad (4.44)$$

Le problème est que les erreurs d'estimation et l'éventuelle troncature de l'ensemble d'information viennent en quelque sorte "polluer" le test d'adéquation à loi uniforme que l'on pourrait directement réaliser à partir de la séquence des processus \hat{z}_t . Plus précisément, la distribution asymp-

totique des statistiques de tests d'adéquation fondés sur ces processus \widehat{z}_t dépend alors de paramètres de nuisance. C'est pourquoi, à partir de ces processus transformés \widehat{z}_t , Bai (1993) propose un test de type Kolmogorov couplé à une transformation martingale de Khmaladze (1981). Le principal avantage de cette statistique est que sa distribution asymptotique n'est pas affectée par les paramètres de nuisance. De plus, les valeurs critiques de ce test sont faciles à calculer et ne nécessitent pas de recours à des techniques de simulation.

Soit $\widehat{V}_T(r)$ un processus empirique défini à partir de la séquence des variables transformées $\{\widehat{z}_t\}_{t=1}^T$ tel que :

$$\widehat{V}_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^T [\mathbb{I}(\widehat{z}_t \leq r) - r] \quad (4.45)$$

Soit $V_T(r)$ le processus équivalent défini quant à lui, à partir des "vraies" transformées probabilistes¹⁰ z_t établies sur la base de la "vraie" distribution conditionnelle, *i.e.* $z_t = F_t(y_t | \Omega_{t-1}, \theta_0)$ et tel que :

$$V_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^T [\mathbb{I}(z_t \leq r) - r] \quad (4.46)$$

Bai montre que le processus $\widehat{V}_T(r)$ admet la représentation asymptotique suivante :

$$\widehat{V}_T(r) = V_T(r) - \bar{g}(r)' \sqrt{T} (\widehat{\theta} - \theta_0) + o_p(1) \quad (4.47)$$

avec

$$\bar{g}(r) = \text{plim} \frac{1}{T} \sum_{t=1}^T \frac{\partial F_t(x | \Omega_{t-1}, \theta_0)}{\partial \theta} \Big|_{x=F_t^{-1}(r | \Omega_{t-1}, \theta_0)} \quad (4.48)$$

Le deuxième terme de droite de la relation (4.47) met en évidence le fait que le processus $\widehat{V}_T(r)$ dépend asymptotiquement de la vraie distribution $F_t(x | \Omega_{t-1}, \theta_0)$ et donc des vrais paramètres θ_0 . On retrouve donc l'idée qu'effectuer un test d'évaluation des prévisions établies à partir d'un modèle

¹⁰Rappelons que sous l'hypothèse nulle de spécification correcte, ces transformées sont *i.i.d.* $U[0, 1]$.

(AR, SETAR ou autre) dont les paramètres sont estimés est rendu difficile par le fait que la distribution asymptotique des statistiques ainsi construites dépend des vrais paramètres (paramètres de nuisance) de la distribution conditionnelle des prévisions. L'erreur d'estimation venant en quelque sorte "polluer" le test d'adéquation à loi uniforme que l'on pourrait directement réaliser à partir de la séquence $\widehat{V}_T(r)$.

Toute l'astuce du test de Bai consiste alors à appliquer une transformation qui permet de retirer le terme $\bar{g}(r)' \sqrt{T} (\hat{\theta} - \theta_0)$ et d'obtenir un processus transformé admettant une distribution asymptotique correspondant à un mouvement Brownien. Cette transformation est la suivante :

$$\widehat{W}_T(r) = \widehat{V}_T(r) - \int_0^1 \left[\dot{g}(s)' C^{-1}(s) \int_s^1 \dot{g}(\tau) d\widehat{V}_T(\tau) \right] ds \quad (4.49)$$

où $g(r) = (r, \bar{g}(r)')'$ et $\dot{g}(r) = \left(1, \dot{\bar{g}}(r)'\right)'$. Bai montre que cette statistique converge vers un mouvement Brownien standard, c'est à dire vers une distribution totalement indépendante du modèle considéré et de la vraie valeur des paramètres de ce modèle. Il propose de considérer la statistique suivante :

$$T_T = \sup_{0 \leq r \leq 1} \left| \widehat{W}_T(r) \right| \xrightarrow[T \rightarrow \infty]{d} \max_{0 \leq r \leq 1} |W(r)| \quad (4.50)$$

où $W(\cdot)$ est un mouvement Brownien standard. Bien évidemment, les valeurs critiques de cette statistique sont obtenues par simulation, mais ceci est fait une fois pour toutes. Bai reporte ainsi les valeurs critiques à 10%, 5% et 1% : 1.94, 2.22 et 2.80. La dernière question qui se pose est celle de la mise en oeuvre pratique de ce test.

4.3.1.2 Application au modèle SETAR

Reprenons l'exemple d'une densité de prévision obtenue à partir d'un modèle SETAR (1, 1, 1) :

$$y_t = h(y_{t-1}, \beta) + \varepsilon_t$$

$$h(y_{t-1}, \beta) = \beta_{1,0} + \beta_{1,1}y_{t-1} + (\beta_{2,0} + \beta_{2,1}y_{t-1}) \times \mathbb{I}(y_{t-1} > c) \quad (4.51)$$

avec $\beta = (\beta_{1,0} \ \beta_{1,1} \ \beta_{2,0} \ \beta_{2,1})'$ et où $\mathbb{I}(\cdot)$ désigne la fonction indicatrice. Dans cet exemple, la connaissance des valeurs passées de y_t suffit à caractériser l'ensemble d'information Ω_{t-1} . Dès lors, le problème de la distinction entre les ensembles d'information Ω_{t-1} et $\tilde{\Omega}_{t-1}$ ne se pose pas a priori. On suppose que les résidus ε_t sont identiquement et indépendamment distribués selon une loi normale de densité $f(x, \lambda)$ et de répartition $F(x, \lambda)$, où λ est un vecteur de paramètres inconnus.

L'hypothèse nulle du test de spécification correcte (*in sample*) de la densité conditionnelle de prévision du modèle SETAR s'écrit alors sous la forme :

$$H_0 : \Pr(y_t \leq r \mid \Omega_{t-1}, \theta) = F[r - h(y_{t-1}, \beta), \lambda] \quad (4.52)$$

où $\theta = (\beta', \lambda')$ désigne le vecteur des paramètres du modèle. Pour des valeurs connues des paramètres β et λ du modèle SETAR, on peut construire les variables transformées z_t telles que :

$$z_t = F_t(y_t \mid \Omega_{t-1}, \theta) = F[y_t - h(y_{t-1}, \beta), \lambda] \quad (4.53)$$

et appliquer ensuite les tests de Diebold, Gunther et Tay (1998). Sous l'hypothèse nulle de spécification correcte, ces variables z_t sont *i.i.d.* et distribuées selon une loi uniforme sur $[0, 1]$. Mais dans la pratique, les paramètres β du modèle SETAR et les paramètres λ de la distribution des résidus ne sont pas connus, mais seulement estimés. Soient $\hat{\beta}$ et $\hat{\lambda}$ deux estimateurs convergents de β et de σ . Dès lors, toute la démarche du test de Bai repose sur l'utilisation des variables transformées suivantes :

$$\hat{z}_t = F_t(y_t \mid \Omega_{t-1}, \hat{\theta}) = F[y_t - h(y_{t-1}, \hat{\beta}), \hat{\lambda}] \quad (4.54)$$

Soient $V_T(r)$ et $\hat{V}_T(r)$, les processus associés respectivement à z_t et \hat{z}_t définis par les équations (4.46) et (4.45). De façon générale, Bai montre (Théorème 2, Section 4) que pour tout modèle non linéaire de type *location-*

scale, le processus $\widehat{V}_T(r)$ peut se décomposer¹¹ sous la forme :

$$\widehat{V}_T(r) = V_T(r) - f(F^{-1}(r)) a_T + \frac{\partial F(F^{-1}(r))}{\partial \lambda} b_T + o_p(1) \quad (4.57)$$

où a_t et b_t désignent deux paramètres indépendants de r . Ainsi, pour cette classe de modèle, les vecteurs $g(r)$ et $\dot{g}(r)$ comportent toujours au maximum trois éléments et s'écrivent respectivement sous la forme¹² :

$$g(r) = \left(r, f(F^{-1}(r)), \frac{\partial F(F^{-1}(r))'}{\partial \lambda} \right)' \quad (4.59)$$

$$\dot{g}(r) = \left(1, \frac{\dot{f}(F^{-1}(r))}{f(F^{-1}(r))}, \frac{\partial f(F^{-1}(r))' / \partial \lambda}{f(F^{-1}(r))} \right)' \quad (4.60)$$

A partir de cette expression de $\dot{g}(r)$, on peut alors construire la séquence des variables martingales transformées $\widehat{W}_T(r)$ à partir de l'équation (4.49).

Toutefois, la transformation martingale utilisée par Bai (2003) requiert l'évaluation d'une intégrale $\int_0^1 \left[\dot{g}(s)' C^{-1}(s) \int_s^1 \dot{g}(\tau) d\widehat{V}_T(\tau) \right] ds$. Bai propose alors une méthode numérique permettant d'évaluer cette intégrale et d'obtenir directement une valeur pour la statistique de test T_T qui correspond à la valeur supérieure des valeurs $\widehat{W}_T(r)$ (équation 4.50). Soient v_1, \dots, v_T les valeurs ordonnées de $\widehat{z}_1, \dots, \widehat{z}_T$, la statistique du test de Bai peut

¹¹Dans le cas où l'on considère des densités paramétriques de la forme $f(x, \lambda)$ où le paramètre λ est un paramètre d'échelle tel que $f(x) = f(x/\lambda)/\lambda$, cette formule se simplifie de la façon suivante :

$$\widehat{V}_T(r) = V_T(r) - f(F^{-1}(r)) a_T + f(F^{-1}(r)) F^{-1}(r) b_T + o_p(1) \quad (4.55)$$

et l'on a alors :

$$g(r) = (r, f(F^{-1}(r)), f(F^{-1}(r)) F^{-1}(r))' \quad (4.56)$$

¹²Par exemple, dans le cas où la fonction de répartition $F(r)$ correspond à celle d'une loi normale centrée réduite, notée $\Phi(r)$, on a :

$$\dot{g}(r) = (1, -\Phi^{-1}(r), 1 - \Phi^{-1}(r)^2)' \quad (4.58)$$

s'écrire sous la forme suivante :

$$T_T = \sup_{0 \leq j \leq T} \left| \widehat{W}_T(r) \right| = \max_{0 \leq j \leq T} \sqrt{T} \left| \frac{j}{T} - \frac{1}{T} \sum_{k=1}^j \dot{g}(v_k)' C_k^{-1} D_k (v_k - v_{k-1}) \right| \quad (4.61)$$

avec $D_k = \sum_{i=k}^n \dot{g}(v_i)$ et $C_k = \sum_{i=k}^n \dot{g}(v_i) \dot{g}(v_i)' (v_{i+1} - v_i)$. Sous l'hypothèse nulle de spécification correcte, cette statistique admet une distribution asymptotique libre de tout paramètre de nuisance. Il suffit de comparer la réalisation de cette statistique aux seuils critiques reportés par Bai : si la réalisation de T_T est supérieure à 2.22, on est alors conduit à rejeter l'hypothèse de spécification correcte de la densité conditionnelle de prévision pour un risque de première espèce de 5%.

De façon générale, sous l'hypothèse nulle de spécification dynamique correcte, les variables transformées $z_t = F_t(y_t \mid \Omega_{t-1}, \theta)$ sont (i) indépendantes et (ii) identiquement distribuées selon une loi uniforme sur $[0, 1]$. L'hypothèse alternative de mauvaise spécification peut dès lors se traduire par une violation de l'une et/ou de l'autre hypothèse. Or, le test de Bai (2003) n'a une puissance asymptotique unitaire uniquement que dans le cas d'alternatives dans lesquelles il y a une violation de l'hypothèse d'uniformité. En revanche, ce test est peu puissant contre des alternatives pour lesquelles il y a violation de l'indépendance. C'est pourquoi, Hong et Li (2004) ont proposé un test puissant pour des alternatives dans lesquelles il y a à la fois violation de l'indépendance et de l'uniformité. Ce test est fondé sur la comparaison de la densité non paramétrique jointe (z_t, z_{t-j}) à la densité jointe de deux variables uniformes sur $[0, 1]$. Pour cela, ils introduisent un estimateur kernel modifié qui permet d'obtenir un "bon" estimateur non paramétrique y compris sur les bords du domaine de définition, c'est à dire aux alentours de 0 et de 1. Tout comme la statistique de Bai, celle proposée par Hong et Li présente en outre l'avantage d'avoir une distribution asymptotique standard et libre de tout paramètre de nuisance. L'inconvénient principal de cette approche réside dans le fait qu'elle implique le choix d'un paramètre de lissage dans la phase d'estimation non

paramétrique. Corradi et Swanson (2005b) proposent un test présentant des propriétés similaires en terme de puissance au test de Hong et Li, mais qui ne nécessite pas le choix d'un paramètre de lissage et dont la statistique converge à un taux paramétrique. L'inconvénient de leur test est que la distribution asymptotique n'est pas libre de paramètre de nuisance et que les valeurs critiques associées doivent donc être simulées selon des procédures assez sophistiquées de Bootstrap.

4.3.2 *Tests de Comparaison de Densités Conditionnelles Mal Spécifiées*

Les tests précédents admettent tous pour hypothèse nulle la spécification (dynamique) correcte de la distribution conditionnelle des densités de prévision. Autrement dit, ils conduisent à tester si un modèle particulier (AR, SETAR etc.) conduit à une densité conditionnelle de prévision (*in sample* ou *out of sample*) correctement spécifiée au sens où cette dernière correspond parfaitement à la densité conditionnelle qui peut être dérivée du DGP des données.

Une approche alternative consiste à comparer un ensemble de modèles de prévisions potentiellement mal spécifiés. Il s'agit alors de repérer parmi un ensemble de modèles concurrents, le modèle qui donne la densité de prévision la plus proche possible de la "vraie" densité de prévision. La différence fondamentale avec l'approche par les tests de spécification correcte, c'est que dans le cadre de l'approche par comparaison on ne suppose pas sous l'hypothèse nulle qu'il existe parmi les différents modèles testés, un modèle bien spécifié. Ainsi, dans le pire des cas, ces tests de comparaison peuvent être appliquées pour discriminer des densités conditionnelles de prévision qui seraient toutes mal spécifiées, c'est-à-dire pour lesquelles les tests de spécification correcte conduiraient au rejet de l'hypothèse nulle.

Cette idée de comparer les qualités prédictives de deux modèles concurrents potentiellement mal spécifiés se trouvait déjà dans les tests de

comparaisons de prévisions ponctuelles proposés par Diebold et Mariano (1995). Mais il existe deux différences fondamentales entre les tests de comparaison à la Diebold et Mariano et les tests de comparaison que nous allons présenter dans cette section. La première différence c'est que les tests fondés sur des prévisions ponctuelles permettent de comparer uniquement des modèles pris deux à deux. Rappelons que dans le cas du test de Diebold et Mariano, l'hypothèse nulle revient à tester l'égalité entre deux fonctions de pertes associées aux prévisions de deux modèles concurrents. A l'inverse les tests de comparaison fondés sur les densités de prévision permettent de comparer un ensemble fini de $N \geq 1$ modèle(s) à un modèle pris comme référence. La deuxième différence plus fondamentale repose sur l'objet de la comparaison. Dans le cas des tests de Diebold et Mariano, la comparaison est établie sur la base d'une fonction de perte calculée à partir des erreurs de prévisions ponctuelles. A l'inverse, dans le cas des tests de comparaison de densités, la comparaison s'établit sur la base de l'ensemble de la distribution conditionnelle des prévisions. On retrouve donc tous les avantages de l'utilisation d'une densité de prévision en lieu et place d'une prévision ponctuelle que nous avons déjà évoqués dans le cadre de la présentation des tests de spécifications correctes. Ces tests, relativement récents, autorisent ainsi une sorte de synthèse entre la littérature sur les tests de comparaison des prévisions ponctuelles à la Diebold et Mariano (1995) et la littérature sur les tests de spécification correcte des densités conditionnelles de prévision.

Corradi et Swanson (2006) opposent deux façons de comparer les densités conditionnelles : soit l'on recourt à un critère d'information de type Kullback-Leibler permettant de classer les modèles, soit l'on se fonde sur un test statistique (et non un simple critère) de comparaison des densités de prévision. Mais cette distinction pose problème dans la mesure où des tests de comparaison de densités conditionnelles de prévision ont été récemment proposés à partir des critères d'information de type Kullback-Leibler. C'est pourquoi, il existe aujourd'hui deux grandes approches de tests qui

reposit soit sur l'utilisation de mesure de distance à la Kullback-Leibler, soit sur l'étude directe des fonctions de répartition empiriques. Nous ne présenterons qu'un test issu de la première approche : le test de Bao, Lee et Saltoglu (2004).

4.3.2.1 *Le test de Bao, Lee et Saltoglu (2004)*

Le test de Bao, Lee et Saltoglu (2004) permet de comparer des densités de prévision potentiellement mal spécifiées. Ce test est fondé sur le critère d'information de Kullback-Leibler (1951) qui permet de mesurer la distance entre deux densités conditionnelles.

De la même façon que précédemment, on note $f_t(y_t, \theta) = f_t(y | \Omega_{t-1}, \theta)$ la densité conditionnelle des prévisions à l'horizon $h = 1$ de la variable y_t obtenue dans le modèle (AR, SETAR, ou autre) conditionnellement à l'ensemble d'information Ω_{t-1} disponible à la date $t - 1$ et où θ désigne un ensemble des paramètres de ce modèle. Soit $\{y_t\}_{t=1}^T$ la séquence des réalisations de la variable y sur la période d'évaluation des prévisions du modèle. On suppose que les observations y_t ont été générées selon un DGP associé à une densité conditionnelle, inconnue, notée $\varphi_t(y_t) = \varphi_t(y | \Omega_{t-1})$. Dans ce contexte, on dit que la densité de prévision est correctement spécifiée si et seulement si pour une valeur θ_0 des paramètres on a :

$$f_t(y_t, \theta_0) = \varphi_t(y_t) \quad \forall y_t \quad (4.62)$$

Dans ce cas, cette densité de prévision associée au modèle de référence est optimale dans le sens elle domine les densités de prévision de tout autre modèle pour n'importe quelle fonction de perte (Diebold, Gunther et Tay, 1998).

Dès lors, dans ce contexte, la mesure de la distance entre la vraie distribution $\varphi_t(y_t)$ et la densité conditionnelle de prévision du modèle $f_t(y_t, \theta_0)$ doit permettre de "classer" les modèles du plus proche au plus éloigné du DGP et donc de choisir un modèle optimal. Pour cela, les auteurs proposent

de retenir un critère d'information de Kullback-Leibler :

$$I(\varphi : f, \theta) = E[\ln \varphi_t(y_t) - \ln f_t(y_t, \theta)] \quad (4.63)$$

où l'espérance est définie par rapport à la vraie loi de y_t . Plus précisément, la distance entre un modèle et la vraie densité est définie comme le minimum en θ du critère d'information de Kullback-Leibler :

$$I(\varphi : f, \theta^*) = E[\ln \varphi_t(y_t) - \ln f_t(y_t, \theta^*)] \leq I(\varphi : f, \theta) \quad \forall \theta \quad (4.64)$$

où θ^* désigne la pseudo vraie valeur de θ qui minimise $I(\varphi : f, \theta)$. On admet¹³ que ce critère peut être estimé par la quantité suivante :

$$I(\varphi : f, \hat{\theta}) = \frac{1}{T} \sum_{t=1}^T [\ln \varphi_t(y_t) - \ln f_t(y_t, \hat{\theta})] \quad (4.65)$$

où $\hat{\theta}$ désigne l'estimateur du paramètre θ^* qui maximise la log-vraisemblance du modèle, c'est à dire $(1/T) \sum_{t=1}^T \ln f_t(y_t, \hat{\theta})$. On note $\hat{f}_t(y_t) = f_t(y_t, \hat{\theta})$.

Sur la base du critère $I(\varphi : f, \theta^*)$, ils serait donc possible théoriquement de comparer différents modèles et de choisir le modèle le plus proche de la vraie distribution conditionnelle des données. Naturellement, le problème c'est que l'on observe pas cette vraie densité $\varphi_t(y_t)$. Bao, Lee et Saltoglu (2004) proposent donc d'utiliser l'inverse de la transformation probabiliste *PIT* (*probability integral transform*) de Rosenblatt (1952). L'équivalence entre $\ln \left[\frac{\varphi_t(y_t)}{\hat{f}_t(y_t)} \right]$ et la log-vraisemblance des variables transformées permet alors d'estimer de façon convergente la quantité $I(\varphi : f, \theta^*)$.

Plus précisément, Bao, Lee et Saltoglu (2004) reprennent un résultat de Berkowitz (2001) qui montre que :

$$\ln \left[\frac{\varphi_t(y_t)}{\hat{f}_t(y_t)} \right] = \ln \left[\frac{p_t(x_t)}{\phi(y_t)} \right] \quad (4.66)$$

¹³En effet, cette assertion n'est pas triviale dans le sens où l'espérance est définie par rapport à la vraie densité inconnue.

où $p_t(\cdot)$ désigne la densité conditionnelle des fractiles de la loi normale associées aux variables PIT et $\phi(\cdot)$ désigne la densité d'une loi normale centrée réduite. Formellement, comme nous l'avons vu dans la section précédente, les variables transformées probabilistes associées aux observations y_t et à la densité conditionnelle du modèle de prévision, s'écrivent :

$$u_t = \int_{-\infty}^{y_t} \hat{f}_t(u, \hat{\theta}) du = \int_{-\infty}^{y_t} \hat{f}_t(u) du \quad (4.67)$$

Rappelons que si le modèle est correcte, les variables $\{u_t\}_{t=1}^T$ sont identiquement distribuées selon une loi uniforme $U_{[0,1]}$. On note alors x_t le fractile normal associé à u_t , soit :

$$x_t = \Phi^{-1}(u_t) \quad (4.68)$$

où $\Phi(\cdot)$ désigne la fonction de répartition de la loi normale. Si le modèle est correctement spécifié, les variables $\{x_t\}_{t=1}^T$ sont identiquement distribuées selon une loi $N(0, 1)$.

Toute l'idée du test de Bao, Lee et Saltoglu consiste alors à substituer à la mesure de la distance entre $\varphi_t(y_t)$ (inconnue) et $\hat{f}_t(y_t)$, la mesure de la distance entre la densité $p_t(x_t)$ et celle de loi normale centrée réduite. Plus la densité conditionnelle de prévision est correctement spécifiée, plus la distribution conditionnelle $p_t(\cdot)$ des variables x_t sera proche d'une loi normale. Ils proposent ainsi un critère de distance défini par :

$$\tilde{I}(\varphi : f, \hat{\theta}) = \frac{1}{T} \sum_{t=1}^T [\ln p_t(x_t) - \ln \phi(x_t)] \quad (4.69)$$

La difficulté consiste alors à postuler une distribution conditionnelle $p_t(\cdot)$ assez générale pour les variables x_t . Tout comme Berkowitz (2001), les auteurs considèrent un processus $AR(L)$ tel que :

$$x_t = \rho' X_{t-1} + \sigma \eta_t \quad (4.70)$$

où $X_{t-1} = (1, x_{t-1}, \dots, x_{t-L})'$, ρ est un vecteur de dimension $(L+1)$ de paramètres et η_t un résidu *i.i.d.* Berkowitz (2001) suppose que les résidus

η_t sont distribués selon une loi normale, ce qui implique que les variables x_t sont conditionnellement normales. Cette hypothèse a été notamment critiquée par Clements et Smith (2000) et Noceti, Smith et Hodges (2003). Noceti, Smith et Hodges avancent en effet l'idée qu'un simple test de normalité des variables x_t pourrait être plus puissant que le test de spécification ainsi construit. C'est pourquoi Bao, Lee et Saltoglu (2004) proposent d'utiliser une loi plus flexible, notée $p(\eta_t, \zeta)$ où ζ désigne un vecteur de paramètre. Sous l'hypothèse $\zeta = \zeta^*$, la distribution se ramène à une loi $N(0, 1)$. Ils retiennent une distribution de type seminonparamétrique (*SNP*) d'ordre K proposée par Gallant et Nychka (1987) telle que :

$$p(\eta_t, \zeta) = \frac{\left(\sum_{k=0}^K r_k \eta_t^k\right)^2 \phi(\eta_t)}{\int_{-\infty}^{\infty} \left(\sum_{k=0}^K r_k u^k\right)^2 \phi(u) du} \quad (4.71)$$

avec $\zeta = (r_1, \dots, r_K)'$, $r_0 = 1$. Si $r_k = 0, \forall k > 0$, alors $p(\eta_t) = \phi(\eta_t)$. Ainsi, la densité de x_t s'écrit :

$$p_t(x_t) = \frac{1}{\sigma} p\left(\frac{x_t - \rho' X_{t-1}}{\sigma}, \zeta\right) \quad (4.72)$$

qui dégénère en une loi $N(0, 1)$ si et seulement si $\rho = 0, \sigma = 1$ et $\zeta = 0'_K$. Sous ces hypothèses, le critère d'information se ramène à l'expression suivante¹⁴ :

$$\tilde{I}(\varphi : f, \hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \left\{ \ln \left[\frac{1}{\sigma} p\left(\frac{x_t - \rho' X_{t-1}}{\sigma}, \zeta\right) \right] - \ln \phi(x_t) \right\} \quad (4.73)$$

Il suffit alors d'estimer les paramètres ζ, ρ et σ par maximisation de la vraisemblance $(1/T) \sum_{t=1}^T \ln p(x_t, \zeta)$. Soient $\hat{\zeta}, \hat{\rho}$ et $\hat{\sigma}$ les estimateurs correspondants. On a finalement :

$$\tilde{I}(\varphi : f, \hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \left\{ \ln \left[\frac{1}{\hat{\sigma}} p\left(\frac{x_t - \hat{\rho}' X_{t-1}}{\hat{\sigma}}, \hat{\zeta}\right) \right] - \ln \phi(x_t) \right\} \quad (4.74)$$

Bao, Lee et Saltoglu (2004) proposent enfin une adaptation de ce critère qui permet de ne considérer que la comparaison des densités de prévision

¹⁴On retrouve ici la formulation qui conduit au test LR de Berkowitz (2001) en posant $p(\cdot) = \phi(\cdot)$.

sur un certain intervalle (et non sur le support complet) afin de proposer une application aux prévisions de *Value-at-Risk*. Enfin, ils proposent en outre une modification de ce critère permettant de réaliser une évaluation des densités de prévision conditionnelle obtenue sur des exercices *out of sample*.

Chapter 5

Conclusion

Ce rapport propose une synthèse de la littérature récente consacrée aux modèles économétriques non linéaires et à leurs apports dans le domaine de la prévision. Il s'inscrit dans le débat académique qui porte sur la question de savoir s'il convient de prendre en compte la non linéarité dans le cadre des prévisions des agrégats macroéconomiques et financiers.

En effet, il est largement reconnu que la plupart, si ce n'est la très quasi-totalité, des séries économiques et financières sont issues de processus générateur de données présentant de fortes non linéarités. Que ce soit sous la forme d'asymétries, ou plus rarement sous la forme d'aspect multimodaux, ces non linéarités traduisent des phénomènes essentiels que l'on ne peut négliger qu'au prix d'un appauvrissement très important de l'analyse économique. L'exemple typique est celui des séries d'emploi ou de chômage, sur lesquelles différentes études fondées sur des méthodes non paramétriques, comme celle proposée notamment par Neftçi (1984), mettent en évidence d'importants phénomènes d'asymétrie selon le cycle économique, les hausses du chômage étant généralement plus brutales que les baisses.

Pour autant, ce n'est pas parce qu'il existe des asymétries et plus généralement des non linéarités dans le processus générateur des données économiques et financières, que le recours à des modèles non linéaires à des

fins de prévision se trouve justifié. En effet, tout dépend de la forme de non linéarité utilisée (choix du modèle), de la forme de la prévision retenue (ponctuelle ou par densité) et de la manière dont on cherche à évaluer cette prévision. Ainsi par exemple, il est aujourd'hui reconnu depuis les travaux systématiques de Stock et Watson (1999), que l'utilisation de modèles à changements de régime et plus spécifiquement de modèles à seuil de type STAR n'apporte pas grand-chose en termes de prévision par rapport aux modèles non linéaires de type ARIMA, lorsque l'on retient comme critère de comparaison les seules prévisions ponctuelles. Il y a alors une distinction très nette entre d'une part l'apport indéniable de ces modèles en matière d'ajustement des données *in sample* et d'autre leur apport plutôt faible en terme de prévision ponctuelle *out of sample*.

Mais cet apport quant à la prévision dépend du critère d'évaluation retenu : si la prévision est envisagée sous la forme d'une HDR ou sous la forme d'une densité, les prévisions issues des modèles à seuil permettent alors de rendre compte de l'asymétrie autour des prévisions ponctuelles, ce que ne peuvent faire les modèles linéaires. En ce sens, leur apport devient essentiel.

C'est pourquoi finalement l'évaluation de l'apport des modèles non linéaires à la prévision nécessite qu'au préalable soient clairement définies les différentes formes de modélisations non linéaires mobilisables pour la prévision, les différentes formes de prévisions que peuvent offrir ces modèles aux économistes et enfin les différentes méthodes d'évaluation qui leur sont associées. C'est précisément ce que nous avons souhaité réaliser dans ce rapport.

Appendix A

Bibliographie

Amendola, A. et Niglio, M., (2003), "The exact multi-step ahead predictor of SETARMA models", *Working Paper* 3.130, Di.S.E.S., Università degli Studi di Salerno.

Anderson, H. M., (1997), "Transaction costs and non-linear adjustment towards equilibrium in the US Treasury bill market", *Oxford Bulletin of Economics and Statistics* 59, pp. 465-484.

Bai, J., (2003), "Testing Parametric Conditional Distributions of Dynamic Models", *Review of Economics and Statistics*, 85, pp. 531-549.

Bao Y., Lee, T.H et Saltoglu B. (2004), "A test for density forecast comparison with applications to risk management", *Working Paper*.

Brännäs, K., De Gooijer, J. G. et Teräsvirta, T., (1998), "Testing linearity against nonlinear moving average models", *Communications in Statistics: Theory and Methods* 27, pp. 2025-2035.

Bec, F., Ben Salem, M. et Carrasco, M., (2004), "Tests of unit-root versus threshold specification with an application to the PPP", *Journal of Business and Economic Statistics* 22, pp. 382-395.

Bec, F., Rahbek, A. et Shephard, N., (2005), "The autoregressive conditional root (ACR) model", *Working Paper* n° 2005-26, INSEE-CREST.

Berkowitz, J. (2001), "Testing Density Forecasts With Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, pp. 465-474.

Berlinet, A. et Francq, C., (1990), "Stationarité et identification d'un processus bilinéaire strictement superdiagonal", *Statistique et Analyse des Données* 15, pp. 1-24.

Bhaskara Rao, M., Subba Rao, T. et Walker, A. M., (1983), "On the existence of some bilinear time series models", *Journal of Time Series Analysis* 4, pp. 95-110.

Bibi, A., (2003), "On the covariance structure of the time-varying bilinear models", *Stochastic Analysis and Applications* 21, pp. 25-60.

Bibi, A. et Gautier, A., (2005), "Stationnarité et inférence asymptotique de modèles bilinéaires périodiques", *C. R. Acad. Sci. Paris, Ser. I* 341, 679-682.

Bibi, A. et Oyet, A. J., (2002), "A note on the properties of some time varying bilinear models", *Statistics & Probability Letters* 58, pp. 399-411.

Bohl, M.T. et Siklos, P.L., (2004), "The present value model of US stock prices redux: a new testing strategy and some evidence", *Quarterly Review of Economics and Finance* 44, pp. 208-223.

Brockett, R. W., (1976), "Volterra series and geometric control theory", *Automatica* 12, pp. 167-176.

Brockwell, P., Liu, J. et Tweedie, R. L., (1992), "On the existence of stationarity threshold autoregressive moving-average processes", *Journal of Time Series Analysis* 13, pp. 95-107.

Brown, B. W. et Mariano, R. S. (1984), "Residual-Based Procedures for Prediction and Estimation in a Nonlinear Simultaneous System", *Econometrica*, 52, pp. 321-343.

Caner, M. et Hansen, B. E., (2001), "Threshold autoregression with a unit root", *Econometrica* 69, pp. 1555-1596.

Chan, F. et McAleer, M., (2002), "Maximum likelihood estimation of STAR and STAR-Garch Models: Theory and Monte Carlo Evidence", *Journal of Applied Econometrics* 17(5), pp. 590-534.

Charemza, W. W., Lifshits, M. et Makarova, S., (2002), "Conditional testing for unit-root bilinearity in financial time series: some theoretical and empirical results", *Journal of Economic Dynamics and Control*, forthcoming.

Chan, K. S., (1993), "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model", *Annals of Statistics* 21, pp. 520-533.

Chan, K.S et TONG, H., (1986), "On estimating thresholds in autoregressive models", *Journal of Time Series Analysis* 7, pp. 178-190.

Chatfield, C. ,(1993), "Calculating Interval Forecasts", *Journal of Business and Economic Statistics*, 11, pp. 121-135.

Chong, Y.Y. et D.F. Hendry (1986), "Econometric Evaluation of Linear Macro-Economic Models", *Review of Economic Studies*, 53, pp. 671-690.

Christoffersen, P.F.,(1998), "Evaluating Interval Forecasts", *International Economic Review*, 39, pp. 841-862.

Clark, T.E. et McCracken M.W., (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models", *Journal of Econometrics*, 105, pp. 85-110.

Clements, M. P., Franses, P. H. et Swanson, N. R.(2004), "Forecasting economic and financial time-series with non-linear models", *International Journal of Forecasting*, 20, pp. 169-183.

Clements, M. P., Franses, P. H., Smith J. et Van Dijk, D., (2003), "On SETAR Non-Linearity and Forecasting", *Journal of Forecasting*, 22, pp.359-375.

Clements M.P. et Hendry DF. (2002), "Modelling Methodology and Forecast Failure", *The Econometrics Journal*, 5, pp. 319-344.

Clements M.P. et Smith J. (1996), "Multi-step estimation for forecasting", *Oxford Bulletin of Economics and Statistics*, 58, pp. 657-684.

Clements, M.P. et Smith J., (1997), "The performance of alternative forecasting methods for SETAR models", *International Journal of Forecasting* 13, pp. 463–475.

Clements, M.P. et Smith J., (2000), "Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment", *Journal of Forecasting*, 19, pp. 255-276.

Clements, M.P. et Smith J., (2001), "Evaluating Forecasts from SETAR Models of Exchange Rates", *Journal of International Money and Finance*, 20, pp. 133-148.

Corradi V. et Swansson N.R. (2005), "A Test for Comparing Multiple Misspecified Conditional Distributions", à paraître dans *Econometric Theory*.

Corradi, V. and Swanson N.R., (2005b), "Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification", à paraître dans le *Journal of Econometrics*.

Corradi V. et Swansson N.R. (2006), "Predictive Density Evaluation", *Handbook of Economic Forecasting*.

Dampser, A. P., Laird, N. M. et Rubin, D. B., (1977), "Maximum likelihood estimation from incomplete data via the EM algorithm", *Journal of The Royal Statistical Society* 39, Series B, pp. 1-38.

De Gooijer, J., (1998), "On threshold moving-average models", *Journal of Time Series Analysis* 19, pp. 1-18.

De Gooijer, J., (2001), "Cross-validation criteria for SETAR model selection", *Journal of Time Series Analysis* 22, pp. 267-281.

De Gooijer, J. et De Bruin, P., (1998), "On SETAR forecasting", *Statistics & Probability Letters*, 37, pp. 7-14.

Diebold, F.X., T. Gunther et A.S. Tay, (1998), "Evaluating Density Forecasts with Applications to Finance and Management", *International Economic Review*, 39, pp. 863-883.

Diebold, F. X. et Inoue, A., (1999), "Long memory and structural change", *PIER Working paper* N° 01-006.

Diebold, F. X., Lee, J. H. et Weinbach, G. C., (1994), "Regime switching with time-varying transition probabilities", in *Non Stationary Time Series and Cointegration*, Hargreaves, C. (ed.), pp 283-302, Oxford University Press.

Diebold, F. X. et Mariano, R. S. (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, pp. 253-263.

Diebold, F. X., Rudebush, G. D. et Sichel, D. E., (1993), "Further evidence on business cycle duration dependence", in Stock, J. et Watson, M. (eds.), *Business Cycles, Indicators, and Forecasting*, pp. 255-280, University of Chicago Press and NBER.

Diebold, F.X., A.S. Tay et K.D. Wallis, (1998), "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters", in *Festschrift in Honor of C.W.J. Granger*, eds. R.F. Engle and H. White, Oxford University Press, Oxford.

van Dijk, D. et Franses, P. H., (1999), "Modelling multiple regimes in the business cycle", *Macroeconomic Dynamics* 3, pp. 311-340.

van Dijk, D., Franses, P. H. et Paap, (2002), "A nonlinear long memory model for US unemployment", *Journal of Econometrics* 102, pp. 135-165.

van Dijk, D., Teräsvirta, T. et Franses, P. H., (2000), "Smooth Transition Autoregressive Models: A Survey of Recent Developments", *Econometric Institute Research Report* EI2000-23/A.

van Dijk, D. Teräsvirta, T. et Franses, P.H. (2002), "Smooth Transition Autoregressive Models: A Survey of Recent Developments", *Econometric Reviews*, 21, pp. 1-47.

Douc, R., Moulines, E. et Ryden, T., (2004), "Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regimes", *Annals of Statistics* 32, pp. 2254-2304.

Enders, W. et Granger, C.W.J., (1998), "Unit root test and asymmetric adjustment with an example using the term structure of interest rates", *Journal of Business and economic Statistics* 16, pp. 304-311.

Enders, W. et Siklos, P. L., (2001), "Cointegration and threshold adjustment", *Journal of Business and economic Statistics* 19, pp. 166-176.

Engle, R. F., Granger, C. W. J., Rice, J. et Weiss, A., (1986), "Semi-parametric estimates of the relation between weather and electricity sales", *Journal of the American Statistical Association* 81, pp. 310-320.

Galeano, P. et Peña, D. (2004), "Model selection criteria and quadratic discrimination in ARMA and SETAR time series models", *Working Paper* 04-14, Statistics and Econometrics Series 05, Universidad Carlos III, Madrid.

Gallant, A.R. et Nychka, D.W. (1987), "Semi-nonparametric maximum likelihood estimation", *Econometrica* 55, pp.3693-390.

Gourieroux, C. et Robert, C. Y., (2001), "Tails and extremal behaviour of stochastic unit root models", *Working Paper*, CREST, CEPREMAP, University of Toronto and University Paris VII.

Granger, C. W. J. et Andersen, A. P., (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprech, Gottinguen.

Granger, C. W. J. et Newbold P., (1977), *Forecasting Economic Time Series*, New York, Academic Press.

Granger, C.W. et Terasvirta, T. (1993), *Modelling NonLinear Economic Relationships*, Oxford University Press.

Guégan, D. (2005), "La persistance dans les marchés financiers", *Note de Recherche* IDHE-MORA N°02-2005.

Guégan, D. et Pham, D. T., (1987), "Minimalité et inversibilité des modèles bilinéaires à temps discrets", *C.R. Acad. Sci. Paris, Série I*, 448, pp. 159-162.

- Guégan, D. et Rioublanc, S., (2005), "Regime switching models: real or spurious long memory?", *Note de recherche IDHE-MORA n°02-2005*.
- Hall, S. G., Psaradakis, Z. et Sola M., (1999), "Detecting periodically collapsing bubbles: a Markov-switching unit root test", *Journal of applied Econometrics* 14, pp. 143-154.
- Hamilton, J. D., (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica* 57, pp. 357-384.
- Hamilton, J. D., (1990), "Analysis of time series subject to change in regime", *Journal of Econometrics* 45, pp. 39-70.
- Hansen, B.E. (1996), "Inference when a Nuisance Parameter is not identified under the Null Hypothesis", *Econometrica*, 64, pp. 413-430.
- Harvey, D., Leybourne, S. et Newbold, P.,(1997), "Testing the Equality of Prediction Mean Squared Errors", *International Journal of Forecasting*, 13, pp. 281-291.
- Harvey, D., Leybourne, S. et Newbold, P.,(1998), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models", *Journal of Business and Economic Statistics*, 16, pp. 254-259.
- Haug, A. et Siklos, P. L., (2005), The behavior of short-term interest rates: international evidence of non-linear adjustment, *Studies in Nonlinear Dynamics and Econometrics* (à paraître).
- Hong, Y.M. et Li, H.T. (2004), "Nonparametric Specification Testing for Continuous Time Models with Applications to Term Structure Interest Rates", *Review of Financial Studies*, 18, pp. 37-84.
- Hyndman, R. J. (1996), "Computing and graphing highest density regions", *The American Statistician* 50, pp. 120-126.
- Jansen, E. S. et Teräsvirta, T., (1996), "Testing parameter constancy and super exogeneity in econometric equations", *Oxford Bulletin of Economics and Statistics* 58, pp. 735-768.

- Johansen, S., (1988), "Statistical analysis of cointegration vectors", *Journal of Dynamics and Control* 12, pp. 231-254.
- Johansen, S., (1991), "Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models", *Econometrica* 59, pp. 1551-1580.
- Johansen, S., (1996), *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford, Oxford University Press.
- Kapetanios, G., (2004), "Nonlinear autoregressive models and long memory", *Working Paper* n°516, Queen Mary, University of London.
- Kapetanios, G. et Shin, Y., (2006), "Unit root tests in three-regime SETAR models", *Econometrics Journal* 9, pp. 252-278.
- Kristensen, D., (2005), "On stationarity and ergodicity of the bilinear model with application to GARCH models", *Working Paper*, University of Wisconsin-Madison.
- Krolzig, H.-M., (1997), *Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*, Springer, Berlin.
- Kupiec, P. (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models", *Journal of Derivatives*, 3, pp.73-84.
- Liu, J., (1990), "A note on causality and invertibility of a general bilinear time series model", *Advances in Applied Probability* 22, pp. 247-250.
- Liu, J., (1992), "On stationarity and asymptotic inference of bilinear time series models", *Statistica Sinica* 2, pp. 479-494.
- Liu, J. et Brockwell, P. J., (1988), "On the general bilinear time series model", *Journal of Applied probability* 25, pp. 553-564.
- Liu, J. et Susko, E., (1992), "On strict stationarity and ergodicity of a nonlinear ARMA model", *Journal of Applied Probability* 29, pp. 363-373.
- Lin, J.-L. et Granger, C. W. J.(1994), "Forecasting from Non-Linear Models in Practice", *Journal of Forecasting*, 13, pp. 1-9.

Ling, S. et Tong, H., (2005), "Testing a linear MA model against threshold MA models", *The Annals of Statistics* 33, 6, pp. 2529-2552.

Ling, S., Tong, H. et Li, D., (2005), "Ergodicity and invertibility of threshold MA(1) models", *Research Reports 122*, London School of Economics.

Lundbergh et Teräsvirta, (2002), "Forecasting with smooth transition autoregressive models", Chapitre 23 in Clements, M. et Hendry, D. F. (eds.), *A companion to Economic Forecasting*, Blackwells, Oxford.

Luukkonen, R., Saikkonen, P. et Teräsvirta, T. (1988), "Testing Linearity against Smooth Transition Autoregressive Models", *Biometrika*, 75, pp. 491-499.

McNees S. (1995), "Forecast Uncertainty: Can it be Measured?", *Federal Reserve Bank of Boston*.

Miller, J.M., (1956), "Table of Percentage Points of Kolmogorov Statistics", *Journal of the American Statistical Association*, 51, pp.111-121

Mizon G. et Richard J. (1986), "The Encompassing Principle and its Application to Testing Nonnested Models", *Econometrica*, 54, pp. 657-678.

Neftçi, S.N. (1984), "Are economic time series asymmetric over the business cycle?", *Journal of Political Economy* 92, 307-328.

Noceti P., Smith J. et Hodges S. (2003), "An evaluation of tests of distributional forecasts", *Journal of Forecasting* 22, pp. 447-455.

Ozaki, T., (1985), "Non-linear time series models and dynamical systems", in Hamman, E. J., Krishnaiah, P. R. et Rao, M. M. (eds.), *Handbook of statistics*, Vol. 5, Elsevier Science Publish, Amsterdam, New York.

Pell, D. et Davidson, J., (1998), "A non-linear error correction mechanism based on the bilinear model", *Economics letters* 58, pp. 165-170.

Pemberton, J., (1987), "Exact least squares multi-step prediction from non-linear autoregressive models", *Journal of Time Series Analysis* 8, pp. 443-448.

- Pesaran, M. H. et Timmermann, A.G., (1992). "A Simple, Non-Parametric Test Of Predictive Performance", *Journal of Business and Economic Statistics*, 4, pp.561-565.
- Pesaran, M. H. et Timmermann, A.G., (1994). "A generalization of the non-parametric Henriksson-Merton test of market timing," *Economics Letters*, vol. 44(1-2), pp. 1-7.
- Pham, D. T., (1985), "Bilinear markovian representation and bilinear models", *Stochastic Processes and Applications* 20, pp. 295-306.
- Potter S. (1995), "A nonlinear approach to US GNP", *Journal of Applied Econometrics*, 10, pp. 109-125.
- Priestley, M. B., (1991), *Non-Linear and Non-Stationarity Time Series Analysis*, Harcourt Brace Jovanovich, London.
- Rahbek, A. et Shephard, N., (2002), "Inference and ergodicity in the autoregressive conditional root model", *Working Paper*, University of Copenhagen et University of Oxford.
- Robinson, P. M., (1983), "Non-parametric estimators for time series", *Journal of Time Series Analysis* 4, pp. 185-207.
- Rosenblatt, M., (1952), "Remarks on a Multivariate Transformation", *Annals of Mathematical Statistics*, 23, pp. 470-472.
- Seo, M., (2005), "Unit root test in a threshold autoregression: asymptotic theory and residual-based block bootstrap", *London School of Economics Discussion Paper* 484.
- Skalin, J. et Terasvirta, T.(2002), "Modeling asymmetries and moving equilibria in unemployment rates", *Macroeconomic Dynamics* 6, pp.202-241.
- Smallwood, A. D., (2005), "Join tests for non-linearity and long memory: the case of purchasing power parity", *Studies in Nonlinear Dynamics & Econometrics* 9, 2, Article 7.
- Subba Rao, T., (1981), "On the theory of Bilinear Time series Models", *Journal of the Royal Statistical Society B* 43, N°2, pp. 244-255.

Subba Rao, T., (1997), "Statistical analysis of nonlinear and nongaussian time series models", in Csiszar, I., Michaeltzky, G. (eds.), *Stochastic Differential and Difference Equations*, Birkhäuser, Boston, pp. 285-298.

Subba Rao, T. et Gabr, M. M., (1984), "An Introduction to Bispectral Analysis and Bilinear Time Series Models", *Lecture Notes in Statistics*, 24, Springer-Verlag, New-York.

Subba Rao, T., (1981), "On the theory of Bilinear Time series Models", *Journal of the Royal Statistical Society B* 43, 2, pp. 244-255.

Taylor, M. P. et Peel, D. A., (2000), "Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals", *Journal of International Money and Finance* 19, pp. 33-53.

Taylor, M. P., Peel, D. A. et Sarno, L., (2001), "Nonlinear mean-reversion in real exchange rates: towards a solution to the Purchasing Power Parity", *International Economic Review* 42, pp. 1015-1042.

Teräsvirta, T., (1994), "Specification, estimation and evaluation of smooth transition autoregressive models", *Journal of the American Statistical Association* 89, pp. 208-218.

Teräsvirta, T., (1998), "Modelling economic relationships with smooth transition regressions", in Ullah, A. et Giles, D.E.A. (eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507-552.

Teräsvirta, T., (2005), "Forecasting economic variables with non linear models", *Working Paper Series in Economics and Finance*, N° 598, Stockholm School of Economics.

Tiao G.C et Tsay, R.S, (1994), "Some advances in non-linear and adaptative modelling in times-series", *Journal of Forecasting*, 13, pp. 109-131.

Tong, H., (1978), "On a treshold model", in C. H. Chen (ed.), *Pattern recognition and Signal Processing*, Sijhoff & Noordhoff, Amsterdam.

Tong, H., (1983), *Threshold Models in Nonlinear Time Series Analysis*, Lecture Notes in Statistics, Springer-Verlag, New-York.

Tong, H., (1990), *Non-Linear Time Series: A Dynamical System Approach*, Clarendon Press, Oxford.

Tong, H. et Lim, K. S., (1980), "Threshold autoregression, limit cycle and cyclical data", *Journal of The Royal Statistical Society*, B 42, pp. 245-292.

Tsay, R. S., (1986), "Nonlinearity tests for time series", *Biometrika* 73(2), pp. 461-466.

Wecker, W. E., (1981), "Asymmetric time series", *Journal of the American Statistical Association* 76, pp. 16-21.

Yatchew, A. (2003), *Semi-Parametric Regression for the Applied Econometrician*, Cambridge.