UNIVERSITÉ DE LA MÉDITERRANÉE (AIX-MARSEILLE II)
FACULTÉ DES SCIENCES ECONOMIQUES ET DE GESTION

ECOLE DOCTORALE DE SCIENCES ECONOMIQUES ET DE GESTION D'AIX-MARSEILLE
N°372

Année 2010

Numéro attribué par la bibliothèque

LLLLLLLLLLLL

Thèse pour le Doctorat ès Sciences Economiques

présentée par

Philippe CHARLOT

Structures Markoviennes cachées et modèles à corrélations conditionnelles dynamiques : extensions et applications aux corrélations d'actifs financiers

---

soutenue le 25 novembre 2010 devant le jury composé de

René GARCIA	Professeur à l'EDHEC	Rapporteur
Christophe HURLIN	Professeur à l'Université d'Orléans	Rapporteur
Roselyne JOYEUX	Professeur à Macquarie University	Examinateur
Anne PEGUIN-FEISSOLLE	Directrice de Recherche CNRS, GREQAM	Examinateur
Vêlayoudom MARIMOUTOU	Professeur, Directeur de l'Institut Français de Pondichéry	Directeur

L'Université de la Méditerranée n'entend ni approuver, ni désapprouver les opinions particulières du candidat : ces opinions doivent être considérées comme propres à leur auteur.

Résumé

L'objectif de cette thèse est d'étudier le problème de la modélisation des changements de régime dans les modèles à corrélations conditionnelles dynamiques en nous intéressant plus particulièrement à l'approche Markov-switching. A la différence de l'approche standard basée sur le modèle à chaîne de Markov caché (HMM) de base, nous utilisons des extensions du modèle HMM provenant des modèles graphiques probabilistes. Cette discipline a en effet proposé de nombreuses dérivations du modèle de base permettant de modéliser des structures complexes. Cette thèse se situe donc à l'interface de deux disciplines : l'économétrie financière et les modèles graphiques probabilistes.

Le premier essai présente un modèle construit à partir d'une structure hiérarchique cachée markovienne qui permet de définir différents niveaux de granularité pour les régimes. Il peut être vu comme un cas particulier du modèle RSDC de Pelletier (2006). Basé sur le HMM hiérarchique de Fine, Singer, and Tishby (1998) notre modèle permet de capter des nuances de régimes qui sont ignorées par l'approche Markov-Switching classique.

La seconde contribution propose une version Markov-switching du modèle de Engle and Sheppard (2001) construite à partir du modèle HMM factorisé de Ghahramani and Jordan (1997). Alors que l'approche Markov-switching classique suppose que les tous les éléments de la matrice de corrélation suivent la même dynamique, notre modèle permet à tous les éléments de la matrice de corrélation d'avoir leur propre dynamique de saut.

Dans la dernière contribution, nous proposons un modèle DCC construit à partir d'un arbre de décision. L'objectif de cet arbre est de relier le niveau des volatilités individuelles avec le niveau des corrélations. Pour cela, nous utilisons un arbre de décision Markovien caché, qui est une extension de HMM développée par Jordan, Ghahramani, and Saul (1997).

Mots clés : Modèle GARCH multivariés ; corrélations conditionnelles dynamiques ; modèle de Markov caché ; modèle de Markov caché hiérarchique ; modèle de Markov caché factorisé ; arbre de décision Markovien caché.

Classification JEL : C32, C51, G1, G0.

Abstract

The objective of this thesis is to study the modelling of change in regime in the dynamic conditional correlation models. We focus particularly on the Markov-switching approach. Unlike the standard approach based on the Hidden Markov Model (HMM), we use extensions of HMM coming from probabilistic graphical models theory. This discipline has in fact proposed many derivations of the basic model to model complex structures. Thus, this thesis can be view at the interface of two disciplines : financial econometrics and probabilistic graphical models.
The first essay presents a model constructed from a hierarchical hidden Markov which allows to increase the granularity of the regimes. It can be seen as a special case of RSDC model of Pelletier (2006). Based on the hierarchical Fine, Singer, and Tishby (1998), our model can capture nuances of regimes that are ignored by the classical Markov-Switching approach.
The second contribution proposes a Markov-switching version of the DCC model of Engle and Sheppard (2001). It is built from the factorial HMM of Ghahramani and Jordan (1997). While the classical Markov-switching approach assumes that all elements of the correlation matrix follow the same switching dynamic, our model allows all elements of the correlation matrix to have their own switching dynamic.
In the final contribution, we propose a model DCC constructed based on a decision tree. The objective of this tree is to link the level of volatility with the level of individual correlations. For this, we use a hidden Markov decision tree, which is an extension of HMM developed by Jordan, Ghahramani, and Saul (1997).

Keywords : Multivariate GARCH ; Dynamic conditional correlations ; Regime switching ; Hidden Markov model ; Hierarchical hidden Markov model ; Factorial Hidden Markov models ; Hidden Markov decision tree.
JEL Classification : C32, C51, G1, G0.

Remerciements

Cette thèse est l'aboutissement d'un long travail durant lequel de nombreuses personnes m'ont, de près où de loin, aidé, soutenu et encouragé. Je profite de cette page pour les en remercier.

Je tiens tout d'abord à remercier chaleureusement Vêlayoudom Marimoutou pour sa disponibilité et son soutien constant au cours de ces années de thèse. Par son enthousiasme jamais pris en défaut et ses patientes réponses à mes sollicitations économétriques, ce travail lui doit beaucoup.

Je tiens également à remercier René Garcia et Christophe Hurlin d'avoir accepté d'être rapporteurs de cette thèse. Les conseils et remarques de leurs rapports me permettent de mieux cerner la direction à prendre pour continuer la suite des travaux présentés dans cette thèse. Je remercie également Roselyne Joyeux pour sa présence dans le jury et de faire un long voyage pour assister à la soutenance. Je remercie également Anne Péguin-Feissolle d'avoir accepté d'être dans le jury, mais aussi pour sa grande gentillesse, ses encouragements, sa disponibilité et pour avoir constamment fait preuve d'intérêt pour mon travail. Je tiens également à témoigner ma gratitude à Marcel Aloy pour son écoute et ses explications techniques qui m'ont permit d'avancer dans le processus de cette thèse.

Ce travail doit également beaucoup au personnel du GREQAM, notamment Bernadette, Corinne, Isabelle et Lydie. Je les remercie pour avoir su gérer les diverses angoisses administratives qui ont jalonnées cette thèse. J'exprime également ma reconnaissance à l'Ecole Doctorale 372 ainsi qu'au GREQAM pour m'avoir apporté un soutien financier me permettant de présenter mes travaux dans des conférences.

Mes remerciements vont également aux thésards du GREQAM qui j'ai croisé tout au long de ses années : Andreea, Adriana, Benoit, Gwenola, Elvira, Elsa, Jimmy, Leila, Luis, Mathieu, Renaud, Paul, Mandy. Je pense également à Gérald et Aroune. Je remercie tout particulièrement Zakaria, compagnon de route de cette thèse. Mes discussion avec lui auront joué un rôle prépondérant dans mes recherches. J'en profite pour faire également un clin d'oeil à Mignon. J'ai également une pensée émue pour mes camarades ñass de la vénérable Большая семья, particulièrement Alonz, Gigi, Monta et Priax, qui me soutiennent depuis toujours.

Cette thèse n'aurait pas été possible sans le soutien de mon père et d'Isabelle. Ce soutien indéfectible, moral mais aussi financier, au cours de ma chaotique scolarité m'a permis de mener ce travail à son terme. Je conclurai en remerciant de tout cœur Tatiana. Son soutien sans limite, ses encouragements, son aide dans les mauvais moments, bref son accompagnement bienveillant font que sans elle, tout cela serait resté sans saveur.

Table des matières

Table des matières xi

Table des figures xv

Liste des tableaux xvii

1 Introduction générale 1
- 1.1 Origines des modèles à corrélations conditionnelles . . . 2
  - 1.1.1 Séries financières et faits stylisés . . . 2
  - 1.1.2 Modélisation de la volatilité . . . 3
    - Modèles à volatilité conditionnellement hétéroscédastique . . . 4
    - Modèles à volatilité stochastique . . . 4
    - Modèles à volatilité multifractale . . . 4
- 1.2 Cadre général des modèles à corrélations dynamiques. . . . 5
  - 1.2.1 Origines . . . 5
    - Modélisation directe de H_t . . . 5
    - Modèles à corrélations conditionnelles . . . 6
  - 1.2.2 Modèles à corrélations dynamiques conditionnelles . . . 7
    - Ecritures vectorielle ou matricielles contraintes. . . . 8
    - Simplifications condensant l'information . . . 9
    - Prise en compte de l'asymétrie . . . 9
    - Semi/non paramétrique DCC . . . 9
    - Changements de régimes . . . 9
    - Grandes matrices de corrélations conditionnelles . . . 11
    - Structures par arbre de décision . . . 12
- 1.3 Cadre général des modèles graphiques probabilistes. . . . 14
  - 1.3.1 Origines et formalisme . . . 14
  - 1.3.2 Modèle de Markov caché . . . 14
    - Origines . . . 14

Chaîne de Markov . . . 15
Modèle HMM . . . 16
1.3.3 Extensions . . . 17
1.4 Constributions de cette thèse . . . 19

2 The Hierarchical RSDC Model 21
2.1 Introduction . . . 22
2.2 DCC-MGARCH and Regime Switching models . . . 22
2.2.1 (S)TAR approach . . . 23
Theorical aspects . . . 23
STCC model . . . 25
2.2.2 Markov-Switching approach . . . 26
Theorical aspects . . . 26
RSDC model . . . 27
2.3 Hierarchical Hidden Markov Structure for Dynamic Correlations . . . 28
2.3.1 Structure . . . 28
The Hierarchical Hidden Markov Model . . . 28
The Hierarchical RSDC model . . . 30
2.3.2 Estimation . . . 31
Estimation by EM algorithm . . . 33
Forward-Backward algorithm . . . 34
Reestimation formulas . . . 35
Estimation Gradient methods . . . 35
Smoothed probabilities . . . 36
2.4 Applications . . . 36
2.4.1 Simulated data . . . 36
2.4.2 Correlations between S&P500 index futures and 10-year bond futures . . . 39
2.4.3 Correlations between exchange rate data . . . 41
2.5 Conclusion . . . 43

3 A DCC with factorial HMM representation 45
3.1 Introduction . . . 46
3.2 The Model . . . 46
3.2.1 Conditional Correlations Models . . . 47
3.2.2 The main approach . . . 48
3.2.3 Definition of the Model . . . 49
3.2.4 Estimation via equivalent HMM representation . . . 51
3.2.5 Smoothing . . . 52
3.2.6 Multi-step-ahead correlation matrix forecast . . . 53
3.3 Applications . . . 53
3.4 Conclusion . . . 61

4 A hidden Markov decision tree-based DCC 63
4.1 Introduction . . . 64
4.2 The Model . . . 65
4.2.1 What is a Hidden Markov Decision Tree ? . . . 65
A Hierarchical Mixtures-of-Experts with Markov dependencies . . . 65
A factorial and coupled HMM . . . 67
4.2.2 Hidden Markov decision tree for correlations . . . 69
Starting point . . . 69

Hidden tree structure . . . 69
Specification for univarite volatilities . . . 71
Specification for correlations . . . 71
4.2.3 Estimation . . . 72
4.3 Application . . . 73
4.3.1 Results of the model . . . 74
4.3.2 Deepening of the relationship between first and second level . . . 77
4.4 Conclusion . . . 80

5 Conclusions et perspectives **81**

A **85**
A.1 An example of HHMM . . . 86
A.2 EM algorithm . . . 86
A.3 Scaling factor . . . 87
A.4 Exchange rate data . . . 88
A.5 S&P500 index futures and 10-year bond futures . . . 88
A.6 Estimated parameters of the HRSDC for the real database applications . . . 88

B **91**
B.1 Stability of stochastic matrix under kronecker product . . . 92
B.2 Smoothed probabilities . . . 92
B.3 Constant matrix construction . . . 94

Bibliographie **97**

Table des figures

1.1 Cartographie des différentes axes de développements des modèles DCC. 13
1.2 Représentation des probabilités de transition d'une chaîne de Markov à trois états. 16
1.3 Représentation d'un HMM sous forme de DBN. 17
1.4 Quatre extensions de HMM utilisés dans cette thèse. 18
(a) Factorial HMM. 18
(b) Coupled HMM. 18
(c) Hidden Markov Decision Tree. 18
(d) Hierarchical HMM. 18

2.1 Examples of logistic and exponential functions (resp. subfigures 2.1(a) and 2.1(b)) with $c = 0$ and different values of γ. 24
(a) $G_t(m_t;\gamma,c) = \frac{1}{1+\exp(-\gamma(m_t-c))}$ 24
(b) $F_t(m_t;\gamma,c) = 1-\exp(-\gamma(m_t-c))$ 24
2.2 Basic structure of a HHMM with three emitting states. 29
2.3 Hierarchical Hidden structure of the HRSDC. 30
2.4 The increase of the granularity in the HRSDC model. 32
2.5 Simulated Correlation Processes. The last two graphs are an example of the estimated correlations of the two models for the DGP1. 37
2.6 Estimated correlations of the HRSDC, DCC_{ES} and DSTCC with the sample of Colacito and Engle (2006). 39
2.7 Smoothed probabilities of the HRSDC. Regimes 1.1 and 1.2 correspond to the emitting states $\{s_1^2, s_2^2\}$ and regimes 2.1 and 2.2 to $\{s_3^2, s_4^2\}$. 40
2.8 Smooth probabilities of the regimes defined by the internal states i_1^1 and i_2^1 (first application). 41
2.9 Estimated correlations for the HRSDC (in black) and DSTCC (in red). 42
2.10 Smoothed probabilities for the HRSDC (exchange rate data). 43

3.1 The Markov-switching model views as a dynamic Bayesian networks. 48
3.2 fHMM view as DBN. 50

3.3 Canadian dollar, Yen and Pound from march 1999 to july 2009. 54
3.4 Estimated correlations of diagonal DCC. 56
3.5 Estimated correlations and smoothed probabilities of the multivariate Markov-switching DCC. 57
(a) Estimated correlations. 57
(b) Smoothed probabilities. 57
3.6 Estimated correlations and smoothed probabilities of the multivariate Markov-switching DCC with FHMM specification. 59
(a) Estimated correlations. 59
(b) Smoothed probabilities. 59

4.1 The hidden Markov model and the hierarchical mixture of experts as graphical models. 66
(a) Hidden Markov model. 66
(b) Hierarchical mixture of experts. 66
4.2 Factorial and Coupled HMM as graphical models. 67
(a) Factorial hidden Markov model. 67
(b) Coupled hidden Markov model. 67
4.3 The hidden Markov decision tree model views as a directed acyclic graph. 68
4.4 Futures prices of 10-year bond and S&P500 1994 to july 2003. 73
4.5 Structure of the decision process for a bivariate dataset. 74
4.6 Estimated correlations and smoothed probabilities. 76
(a) Estimated correlations. 76
(b) Smoothed probabilities. 76
4.7 Structure of the decision process for a bivariate dataset of the extended version. . . 77
4.8 Estimated correlations and smoothed probabilities. 78
(a) Estimated correlations. 78
(b) Smoothed probabilities. 78

A.1 An example of Hierarchical Hidden structure. 86
A.2 Exchange rate database. 88
A.3 S&P500 index futures and 10-year bond futures. 89

Liste des tableaux

2.1 Performance measures results. 38

3.1 Descriptive statistics, GARCH(1,1) estimates, DCC and MS-DCC estimated parameters. 55
(a) Descriptive statistics for the Canadian dollar, Yen and Pound. 55
(b) GARCH(1,1) estimates. 55
(c) Diag-DCC estimates. 55
(d) MS-DCC. 55
3.2 FHMM-DCC estimated parameters. 60

4.1 Descriptive statistics for the 10 year treasury bonds and the S&P 500 futures prices. 74
4.2 Model estimates. 75
4.3 Model estimates. 79

A.1 Estimated parameters for the correlations of the second real data application. . . . 89
A.2 Estimated parameters for the correlations of the second real data application. . . . 90

CHAPITRE

1

Introduction générale

L'objet de cette thèse est d'étudier le problème des changements de régime dans les modèles à corrélations conditionnelles dynamiques en proposant des modélisations issues des modèles graphiques probabilistes. Après avoir effectué une brève revue des modèles utilisés traditionnellement en économétrie financière pour modéliser la volatilité, nous présentons plus spécifiquement les modèles GARCH multivarié à corrélations conditionnelles dynamiques. Un bref état de l'art permettra ensuite de pouvoir cerner les avancées récentes dans ce domaine. La troisième partie de cette introduction apporte quelques éléments de base concernant les modèles graphiques probabilistes et présente succinctement quelques-uns des modèles que nous utiliserons par la suite. La dernière partie de l'introduction offre un résumé des contributions de cette thèse.

1.1 Origines des modèles à corrélations conditionnelles

1.1.1 Séries financières et faits stylisés

L'idée de représenter l'évolution d'une série financière par une loi de probabilité apparaît pour la première fois dans la thèse de Bachelier (1900). La détermination de la nature de la loi sous jacente a depuis focalisée une attention croissante des chercheurs en économie et statistique.

Les variations d'une série financière à la hausse ou à la baisse avec une amplitude plus ou moins forte caractérisent la *volatilité*. L'étude de la volatilité d'une série financière revient donc à étudier son rendement, c'est-à-dire l'étude de la trajectoire de ses variations. En notant S_t le cours d'une série financière à la date t, le rendement r_t consiste en une transformation logarithmique du cours :

$$r_t = \log S_t - \log S_{t-1} \tag{1.1}$$

$$\simeq \frac{S_t - S_{t-1}}{S_{t-1}} \tag{1.2}$$

Le rendement r_t peut ainsi être interprété comme la variation relative du cours d'une série financière. Cette transformation permet ainsi de comparer plusieurs séries.

La recherche des spécifications relatives à la distribution des rendements a été influencée par des études mettant en évidence certaines caractéristiques et régularités statistiques propres aux séries financières. Ces faits stylisés communs aux séries financières ont été et sont encore largement discutés dans la littérature. Nous exposons brièvement ces régularités statistiques :

- non stationnarité : le cours d'une série financière n'est généralement pas stationnaire au second ordre.
- absence d'autocorrélation des variations de prix : les autocorrélations du cours sont généralement très faibles. Ce fait renvoie à l'hypothèse de bruit blanc du cours.
- autocorrélation de carrés des variations de prix : les carrés des rendements présentent généralement de fortes autocorrélations.
- clustering : de fortes variations des rendements sont suivies de fortes variations. Ce regroupement des extrêmes est évoqué dans l'article précurseur de Mandelbrot (1963b)[1].
- queues de distribution épaisses : les cours ainsi que les rendements sont leptokurtiques. Cette caractéristique, qui tend à rejeter le paradigme gaussien, a été notamment popularisée par les articles de Mandelbrot (1963a) et Fama (1965).

[1] "Large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes."

– effets de levier : cet effet fait référence à la réponse asymétrique des valeurs passées positives et négatives sur la volatilité. Concrètement, les valeurs négatives ont tendance à provoquer une augmentation de la volatilité supérieure à celle induite par une hausse du cours d'une ampleur similaire. Black (1976) montre que l'effet de levier a tendance à s'accompagner d'un effet d'asymétrie alors que la réciproque n'est pas vérifiée.
– saisonnalité : la volatilité tend à augmenter lorsque les marchés ne fonctionnent pas. De nombreux effets de saisonnalités apparaissent en fonction de la fréquence des observations. Parmi ceux-ci, citons le cas de l'effet week-end (voir French (1980)) et l'effet janvier (voir Keim (1983)).

Le lecteur intéressé par un exposé plus complet de ces faits stylisés pourra consulter le premier chapitre de la monographie de Francq and Zakoian (2009) ainsi que l'article de Cont (2001).

Les méthodes classiques d'analyse des séries temporelles comme les modèles de type ARMA ne sont globalement pas adaptées pour reproduire de telles propriétés statistiques dans la mesure où une formulation simple de type ARMA ne permet pas de traduire une relation non linéaire. Rappelons qu'un processus ARMA(p,q) est de la forme :

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_t \varepsilon_{t-1} \forall t \text{ avec } \phi_i, \theta_i \in \mathbb{R} \text{ et } \in_t \rightsquigarrow WN(0,\sigma) \tag{1.3}$$

La valeur présente des séries est écrite comme une fonction linéaire des valeurs passées et de la valeur présente d'un bruit, qui est l'innovation de la série. Cette formulation présente l'intérêt d'être simple à mettre en œuvre, mais elle est cependant beaucoup trop restrictive. La linéarité des modèles ARMA s'exprime avec le théorème de Wold. C'est-à-dire que tout processus centré et stationnaire peut être représenté sous une forme MA infini :

$$X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + \eta_t \text{ avec } \eta_t \rightsquigarrow WN(0,\sigma) \text{ et} \sum_j |\theta_j| < \infty \tag{1.4}$$

Autrement dit, tout processus faiblement stationnaire peut être reformulé sous la forme d'une combinaison linéaire d'une suite de variables aléatoires non corrélées dans le temps. Cette hypothèse forte de linéarité ne permet donc pas de prendre en compte des comportements d'asymétrie, de rupture ou l'existence de cluster de volatilité.

1.1.2 Modélisation de la volatilité

Pour tenter de reproduire les propriétés statistiques évoquées précédemment, il a donc fallu s'orienter vers des spécifications non-linéaires. Dans ce cadre, les modèles économétriques sont généralement construits à partir de la représentation multiplicative suivante :

$$r_t = h_t^{1/2} \eta_t \tag{1.5}$$

où η_t est une suite de variables aléatoires indépendantes et identiquement distribuées selon une loi de probabilité discrète, de moyenne nulle et de variance unité. Le terme h_t représente la variance conditionnelle du processus r_t. Dans cette formulation, il est donc possible de construire un processus permettant de reproduire certains des faits stylisés en jouant sur le choix de distribution de η_t mais aussi en spécifiant la formulation de h_t. Plusieurs classes de modèles ont émergé dans la littérature quant à la spécification de la variance conditionnelle. Nous en présentons les principales.

Modèles à volatilité conditionnellement hétéroscédastique Introduits par Engle (1982), les modèles de type ARCH (Auto Regressive Conditional Heteroskedasticity) sont construits à partir de l'hypothèse que la variance des perturbations évolue avec le temps :

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 \tag{1.6}$$

Ainsi, dans cette configuration le carré des perturbations suit un processus autorégressif d'ordre p. Ce premier modèle à variance conditionnellement hétéroscédastique a été rapidement généralisé par Bollerslev (1986) en introduisant une composante moyenne mobile pour aboutir au modèle GARCH (Generalized Auto Regressive Conditional Heteroskedasticity) :

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} \tag{1.7}$$

Les paramètres à estimer du modèles sont alors $\{\omega, \alpha_i, \beta_j\}$, $i = 1, ..., p$ et $j = 1, ..., q$. Cette classe de modèle a rapidement été très populaire en raison de sa facilité à être estimée.

Modèles à volatilité stochastique Une autre classe de modèle est celle composée des modèles à volatilité stochastique. Cette approche proposée par Taylor (1982) est définie comme :

$$\log h_t = \omega + \phi \log h_{t-1} + v_t \tag{1.8}$$

où les innovations v_t sont indépendantes de η_t et avec $v_t \overset{iid}{\sim} \mathcal{N}(0, \sigma)$. Les paramètres à estimer sont $\{\omega, \eta, \phi\}$. Cette approche possède donc deux sources d'aléa : v_t correspond à l'impact du flux d'information tandis que η_t représente la nature de l'information (bonne ou mauvaise). A la différence des modèles GARCH, où la variance est spécifiée directement à partir des valeurs observées des rendements, les modèles à volatilité stochastique spécifient la volatilité comme un processus latent. Cette particularité rend l'estimation de ces modèles plus compliquée que la classe précédente.

Modèles à volatilité multifractale Cette approche trouve ses origines dans les travaux de Mandelbrot (1967, 1974, 1982, 1988). Un premier modèle, nommé MMAR (Multifractal Model of Asset Returns) basée sur cette approche prenant en compte des relations d'échelle a été proposée par Calvet, Fisher, and Mandelbrot (1997). Ce modèle a ensuite été enrichi pour aboutir au modèle MSM (Markov-Switching Multifractal) par Calvet and Fisher (2001, 2002) où la volatilité est déterminée par le produit de $\bar{k}$ composantes aléatoires $M_{k,t}$:

$$h_t = \bar{\sigma} \left(\prod_{i=1}^{\bar{k}} M_{k,t} \right)^{1/2} \tag{1.9}$$

où $\bar{\sigma}$ est un paramètre à estimer. A chaque instant les composantes de volatilités suivent une dynamique markovienne au premier ordre avec une probabilité de transition spécifiée comme :

$$\gamma_k = 1 - (1 - \gamma_1)^{(b^{k-1})} \tag{1.10}$$

avec b et γ_k des paramètres devant être estimés.

Parce qu'ils ne reproduisent que partiellement les régularités statistiques évoquées précédemment, ces modèles pionniers ont connu de nombreux raffinements. En particulier de nombreuses spécifications non linéaires ont été proposées. Cette sophistication des modèles de base est à l'origine

d'une vaste littérature que nous ne pouvons résumer dans cette thèse[2]. Par ailleurs, ces différents axes de recherches illustrent la difficulté de modélisation de la volatilité des marchés financiers. Les origines des régularités statistiques observées sur les séries financières ne sont pas encore parfaitement comprises aujourd'hui. Leur modélisation statistique pose des problèmes théoriques complexes. A ce jour, aucune des approches présentées ne s'est réellement imposée et l'économétrie financière reste un domaine en constante évolution.

1.2 Cadre général des modèles à corrélations dynamiques.

L'extension des modèles GARCH univariés au cadre multivarié est motivée par l'intérêt de décrire la volatilité de plusieurs séries temporelles conjointement, afin d'exploiter autant que possible les liens qui peuvent exister entre ces séries[3]. Dans ce cadre, la volatilité conditionnelle d'une série vectorielle de rendements r_t, de dimension $K \times 1$ s'écrit à partir de la formulation :

$$r_t = H_t^{1/2} \eta_t \tag{1.11}$$

où $H_t = [h_{ijt}]$ est la matrice de covariance des rendements de taille $K \times K$ et η_t un processus de dimension indépendant et identiquement distribué de moyenne nulle et tel que $\mathbb{E}(\eta_t \eta_t) = \mathbb{I}$ où $\mathbb{I}$ est une matrice identité de taille K.

1.2.1 Origines

L'extension des modèles GARCH au cadre multivarié n'est pas immédiate et fait apparaître plusieurs difficultés. Une généralisation directe du modèle univarié entraînerait une inflation du nombre de paramètres à estimer conduisant très rapidement à des problèmes numériques. Les développements de modèles GARCH multivariés doivent donc concilier une spécification plus ou moins parcimonieuse en paramètres tout en préservant une certaine flexibilité pour ne pas perdre de leur pouvoir explicatif. Deux classes de modèles ordonnent ces extensions. Dans la première, la matrice des covariances conditionnelles est modélisée directement. Dans la seconde, cette matrice est obtenue en intégrant les corrélations conditionnelles.

Modélisation directe de H_t Dans cette première classe de modèles, la matrice de covariance conditionnelle est modélisée directement et adopte donc la formulation :

$$H_t = f(H_{t-1}, H_{t-2}, ..., r_{t-1}, r_{t-2}, ...) \tag{1.12}$$

où f est une fonction de transformation mesurable par rapport à l'ensemble d'informations à la date $t-1$, noté $\mathscr{F}_{t-1}$. Une des premières contributions est celle de Bollerslev, Engle, and Wooldridge (1988) avec le modèle VEC. Ce modèle est une généralisation du modèle de base univarié. Cette première formulation, si elle présente un intérêt au niveau théorique n'est pas satisfaisante

[2] L'ouvrage de Engle (1995) regroupe les articles fondateurs des modèles GARCH. La synthèse de Silvennoinen and Teräsvirta (2009) présente un panorama des modèles GARCH univariés. On pourra également se référer au travail de Bollerslev (2008) qui recense de façon exhaustive les différentes déclinaisons ainsi qu'à la monographie de Francq and Zakoian (2009) qui présente un état récent de la recherche théorique de cette classe de modèle. Les fondements des modèles à volatilité stochastiques sont présentés dans le recueil d'articles de Shephard (2005) ainsi que dans l'article de Shephard and Andersen (2009). Les modèles à volatilité multifractale sont l'objet d'un ouvrage de synthèse de Calvet and Fisher (2008).

[3] D'un point de vue concret, une formulation multivariée peut alors apporter une information importante afin de construire un portefeuille composé d'actifs très variés. Ainsi, les covariances entre les actifs constituent un outil de décision crucial dans le CAPM.

car trop générale. Une formulation contrainte est également proposée avec le modèle diagonal-VECH, impliquant que les éléments de la matrice H_t dépendent du carré des résidus passés et d'un terme autorégressif. Les covariances sont alors une combinaison linéaire du produit croisé des résidus passés ainsi que d'un terme autorégressif. Cependant, cette configuration, même contrainte demeure très lourde au niveau numérique. Dans la continuité du modèle VEC, Engle and Kroner (1995) proposent la représentation BEKK[4], imposant des contraintes moins fortes sur les paramètres et permettant des conditions plus faibles pour respecter la stationnarité du processus. Par ailleurs, l'avantage considérable de cette formulation réside dans le fait que la matrice H_t est écrite sous une forme quadratique. Il en résulte qu'elle est donc toujours définie positive, allégeant par là même les difficultés numériques de l'estimation. Une solution alternative à l'écriture quadratique est proposée par Kawakatsu (2003), qui consiste à construire la matrice de covariance à partir d'une représentation de Choleski.

Pour limiter l'inflation de paramètre, plusieurs chercheurs ont proposé des modèles basés sur des facteurs. Ainsi, Engle, Ng, and Rothschild (1990) exploitent l'idée que les éléments dépendant du temps dans la matrice de variance-covariance sont définis par une fonction de combinaisons linéaires de variables aléatoires. L'approche factorielle est également exploitée par Alexander (2002), van der Weide (2002) ainsi que Vrontos, Dellaportas, and Politis (2003) qui proposent un modèle GARCH multivarié basé sur l'analyse en composantes principales[5]. Lanne and Saikkonen (2007) réalisent une spécification autorisant un nombre réduit de facteurs conditionnellement hétéroscédastiques (modèle GOF GARCH), contrairement aux approches orthogonales précédentes. Enfin, le modèle présenté par García-Ferrer, González-Prieto, and Peña (2008) peut être considéré comme une raffinement du GOF-GARCH dans lequel la sélection des facteurs est réalisée par une analyse en composantes indépendantes au lieu d'une analyse en composantes principales (modèle GICA GARCH).

Des spécifications semi/non paramétriques ont également été proposées. Hafner and Rombouts (2007) établissent un modèle semi-paramétrique dans lequel la matrice de covariance est modélisée par une spécification paramétrique (comme un modèle VEC ou BEKK) tandis que l'erreur de distribution est estimée non paramétriquement. Une approche similaire, combinant une composante paramétrique et une composante non paramétrique, est proposée par Long and Ullah (2005).

Modèles à corrélations conditionnelles Les modèles à corrélations conditionnelles diffèrent de la formulation explicitée dans le paragraphe dans le sens où la matrice de covariance conditionnelle incorpore une matrice de corrélation conditionnelle. Initiée par Bollerslev (1990) avec le modèle à corrélations conditionnelles constantes (CCC), cette classe repose sur la décomposition suivante :

$$H_t = D_t R D_t \tag{1.13}$$

où R est, dans l'article pionnier de Bollerslev (1990), une matrice de corrélation conditionnelle constante dans le temps de taille $K \times K$. Le terme D_t est une matrice diagonale de taille $K \times K$ dont les éléments contiennent les écarts types conditionnels de chaque série :

$$D_t = \text{diag}\{h_{i,t}^{1/2}\},\ i = 1, ..., K \tag{1.14}$$

Cette information est extraite grâce à un modèle univarié à variance conditionnelle de la famille GARCH.

La représentation définie par l'équation (1.13) implique donc une modélisation de la matrice de covariance conditionnelle en deux étapes. Dans un premier temps, l'information concernant la volatilité individuelle est extraite avec une filtration de type GARCH. Cette filtration[6] permet

[4]Le nom de cette formulation provient d'une synthèse des analyses réalisées concernant les représentations multivariées par Robert F. Engle, Kenneth F. Kroner, Yoshihisa Baba et Dennis F. Kraft.

[5]Modèles Orthtogonal GARCH, Generalized Orthogonal GARCH et Full Factor GARCH.

[6]L'étape d'extraction de la volatilité individuelle est appélée par Engle (2009) *degarching*.

d'obtenir les résidus standardisés de la série vectorielle de rendements :

$$\varepsilon_t = D_t^{-1} r_t \tag{1.15}$$

Les corrélations conditionnelles correspondent alors à l'espérance des résidus standardisés :

$$\mathbb{E}_{t-1}[\varepsilon_t \varepsilon_t'] = D_t^{-1} H_t D_t^{-1} = R \tag{1.16}$$

avec R une matrice carrée symétrique définie positive. La matrice de covariance conditionnelle est ainsi une combinaison des variances conditionnelles individuelles et de la variance conditionnelle des résidus standardisés. D'un point de vue pratique, cette formulation présente un avantage numérique indéniable. En effet, la première étape relative à la filtration GARCH des séries individuelles correspond à une somme de problèmes numériques faciles à traiter tandis que l'estimation de la variance des résidus standardisés peut s'effectuer dans un second temps. Cette estimation en plusieurs étapes décompose donc un problème numérique complexe en une somme de problèmes plus faciles à traiter.
Néanmoins, si le modèle CCC constitue une avancée théorique importante, il apparaît comme trop restrictif pour une utilisation concrète. L'hypothèse de corrélations constantes sur les marchés financiers n'est en effet pas vérifiée empiriquement[7]. Les recherches se sont alors focalisées sur l'introduction de corrélations variant dans le temps.

1.2.2 Modèles à corrélations dynamiques conditionnelles

L'introduction de corrélations conditionnelles dynamiques consiste à donner un caractère temporel à la matrice de corrélation, l'équation (1.13) étant alors reformulée comme :

$$H_t = D_t R_t D_t \tag{1.17}$$

Deux spécifications dynamiques ont inauguré presque simultanément cette extension du modèle de base. Tse and Tsui (2002) proposent ainsi d'écrire les corrélations conditionnelles à partir d'un processus autorégressif (modèle DCC_T) :

$$R_t = (1-a-b)\overline{R} + aR_{t-1} + b\Psi_{t-1} \tag{1.18}$$

où $\overline{R}$ est la matrice des corrélations non conditionnelles, Ψ_{t-1} une matrice de corrélation empirique des résidus standardisés et $a+b \leq 1$. Ce modèle est donc basé sur l'hypothèse que les corrélations sont une fonction des corrélations passées ainsi que des corrélations empiriques.
Parallèlement à Tse and Tsui (2002), Engle and Sheppard (2001) proposent également une écriture autorégressive des corrélations conditionnelles dynamiques (modèle DCC_{ES}). Cependant, leur approche diffère de la précédente car les corrélations ne sont pas formulées comme une fonction des corrélations retardées. Ainsi, au lieu de modéliser directement les corrélations, Engle and Sheppard (2001) imposent une dynamique de type BEKK au processus des covariances des résidus standardisés :

$$Q_t = \Omega + a\varepsilon_{t-1}\varepsilon_{t-1}' + bQ_{t-1} \tag{1.19}$$

avec a et b deux paramètres vérifiant $a+b<1$. Les corrélations sont alors obtenues pour une combinaison des covariances :

$$R_t = \mathrm{diag}\{Q_t\}^{-1/2} Q_t \mathrm{diag}\{Q_t\}^{-1/2} \tag{1.20}$$

[7] L'hypothèse de corrélations constante est d'abord rejetée dans une analyse de Longin and Solnik (1995). L'étude de Longin and Solnik (2001) montre que les corrélations ont tendance à augmenter davantage dans une situation de forte volatilité à la baisse que lors d'une forte volatilité à la hausse. Ce caractère asymétrique du mouvement des corrélations est confirmé dans les articles de Ang and Chen (2002) et Campbell, Koedijk, and Kofman (2002).

La constante ω est estimée indépendamment par *variance targeting*. Ainsi, les auteurs contournent le problème délicat de l'estimation de la constante en ayant uniquement recours à un estimateur des moments. Cette constante est alors estimée par :

$$\hat{\Omega} = (1 - a - b)\overline{Q} \tag{1.21}$$

avec

$$\overline{Q} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t \tag{1.22}$$

Dans la formulation de Engle and Sheppard (2001), le processus suivi par les corrélations ne nécéssite que deux paramètres. Les modèles de type DCC ont rapidement connu un grand succès. Outre le fait d'incorporer les corrélations dans l'écriture des covariances conditionnelles, c'est leur facilité d'estimation qui rend ce modèle attractif pour le praticien[8].
Néanmoins, cette spécification parcimonieuse en paramètres impose une dynamique des corrélations très restrictive. Diverses extensions ont alors été proposées afin d'apporter davantage de flexibilité et enrichir le processus suivi par les corrélations. Nous allons tenter d'en exposer les principales.

Ecritures vectorielle ou matricielles contraintes. L'écriture scalaire de l'équation (1.19) étant très réductrice, Engle (2002) propose une écriture matricielle permettant d'avoir des coéfficients de covariations spécifiques (modèle G-DCC). L'équation (1.19) se reformule alors comme :

$$Q_t = [\overline{Q} - A'\overline{Q}A - B'\overline{Q}B] + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'Q_{t-1}B \tag{1.23}$$

dans laquelle A et B sont deux matrices de taille $K \times K$. Cette formulation très générale impose cependant des conditions de stationnarité délicates à respecter puisque que $\overline{Q} - A'\overline{Q}A - B'\overline{Q}B$ doit être une matrice semi-définie. Dans le but de réduire le nombre de paramètres, une écriture vectorisée peut être une solution efficace. En remplaçant les matrice A et B par des vecteurs l'équation (1.23) s'écrit :

$$Q_t = [\iota\iota' - A - B] \circ \overline{Q} + A \circ \varepsilon_{t-1}\varepsilon'_{t-1} + B \circ Q_{t-1} \tag{1.24}$$

Néanmoins, si cette formulation diminue effectivement le nombre de paramètres tout en préservant une certaine généralité, les conditions pour que la constante soit semi-définie positive restent difficiles à satisfaire. Hafner and Franses (2009) proposent une version parcimonieuse de l'équation précédente en écrivant :

$$Q_t = [1 - \sum_{i=1}^{n} \alpha_i - \beta]\overline{V} + \alpha\alpha' \circ \varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1} \tag{1.25}$$

où α un vecteur de longueur K et β un scalaire.
Une autre solution pour économiser des paramètres consiste à attribuer la même dynamique aux séries similaires. Billio, Caporin, and Gobbo (2006) proposent un DCC vectorisé :

$$Q_t = cc' + \alpha\alpha' \circ \varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1} \tag{1.26}$$

où a, b, et c sont des vecteurs partitionnés. Ce modèle est généralisé par Billio and Caporin (2009) en substituant aux vecteurs partionnés des matrices par blocs.

[8] En effet, ce modèle s'estime en plusieurs étapes, là où les modèles à covariances conditionelles s'estimaient en une étape impliquant un problème numérique complexe. Dans le modèle Engle and Sheppard (2001), l'estimation peut se décomposer en trois étapes : (i) *degarching*, i.e. extraction des volatilités individuelles avec un processus de type GARCH, (ii) estimation de la constante Ω par *variance targeting*, (iii) estimation des paramètres a et b avec la méthode du maximum de vraisemblance. Une explication plus approfondie de l'estimation de ce modèle se trouve dans les chapitres suivants de cette thèse.

Simplifications condensant l'information Afin de limiter le nombre de paramètres tout en gardant une certaine généralité dans la dynamique des corrélations, une possibilité consiste à regrouper les séries qui ont les mêmes dynamiques en groupe. Cette spécification avec des vecteurs partitionnés ou des matrices par blocs a été explorée notamment par Billio, Caporin, and Gobbo (2006) et Billio and Caporin (2009). Néanmoins, dans ces deux approches, l'allocation des vecteurs/matrices s'effectue manuellement, au choix du modélisateur. Un axe de recherche récent de la littérature concerne le partitionnement de la matrice de corrélation selon des techniques de *clustering*. Zhou and Chan (2008) comparent ainsi différentes méthodes de partitionnement et concluent à l'efficacité du clustering hiérarchique (modèle C-DCC1). So and Yip (2009) proposent également une approche par clusters (modèle C-DCC2) reposant sur une approche Bayésienne.
Cependant, les méthodes de partitionnement de type clusters présentent l'inconvénient d'être difficilement interprétables. Une approche alternative, suggérée par Caporin and Paruolo (2009) propose d'utiliser des méthodes de partitionnement issues de l'économétrie spatiale. Dans ce cadre, les paramètres du modèle définie par l'équation (1.23) sont déterminés à partir d'une matrice de proximité indiquant l'intensité de la dépendance d'une série par rapport à une autre. Cette approche plug-in présente l'avantage d'être interprétable tout en relâchant l'hypothèse d'un partitionnement a priori.

Prise en compte de l'asymétrie La prise en compte des effets d'asymétrie consiste à introduire un paramètre attribuant un poids aux résidus négatifs. Une extension asymétrique de l'équation (1.23) est formulée par Cappiello, Engle, and Sheppard (2006) :

$$Q_t = (\bar{V} - A'\bar{V}A - B'\bar{V}B - G'\bar{N}G) + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'Q_{t-1}B + G'n_{t-1}n'_{t-1}G \tag{1.27}$$

avec A, B et G des matrices diagonales. L'effet d'asymétrie est capté par la variable n_t, qui vérifie : $n_t = \mathbb{1}_{\{\varepsilon_t<0\}} \circ \varepsilon_t$ et permet de calculer $\bar{N} = T^{-1}\sum_{t=1}^{T} n_t n'_t$. Les conditions de stationnarité de ce modèle impliquent que Q_0 définie positive et que la constante $(\bar{V} - A'\bar{V}A - B'\bar{V}B - G'\bar{N}G)$ soit semi définie positive.
Exploitant la même idée, Vargas (2006) suggère une extension asymétrique du modèle Block-DCC de Billio, Caporin, and Gobbo (2003). Une version asymétrique du DCC basée sur la loi multivariée de Laplace asymétrique est réalisée par Cajigas and Urga (2006).

Semi/non paramétrique DCC Une approche semi-paramétrique (modèle SPCC) est présentée par Franses, Hafner, and van Dijk (2005). C'est la matrice de covariances des résidus qui est calculée de façon non-paramétrique :

$$Q_t(x) = \frac{\sum_{t=1}^{T} \varepsilon_t \varepsilon'_t K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)} \tag{1.28}$$

où x_t est une variable de conditionnement observable (par exp. : $x_t = t$), $K_h(\cdot) = (1/h)K(\cdot/h)$ un noyau quelconque et h la fenêtre. Les corrélations sont donc calculées à partir d'un estimateur du type Nadaraya-Watson transformé.
Feng (2007) développe une estimation non paramétrique des corrélations à partir d'une méthode k-NN (k-nearest-neighbours) avec fenêtre variable (modèle local-DCC). L'idée des méthodes k-NN consiste à classer chaque observation selon une règle de décision afin de trouver les k plus proches observations.

Changements de régimes Plusieurs contributions se focalisent sur l'introduction de régimes dans le processus des corrélations. Diverses extensions du modèles à corrélations conditionnelles sont

donc apparues, basées aussi bien sur des spécifications déterministes (modèles à seuils, à transition douce) que stochastiques (modèle Markov-switching). Ces extensions sont en partie motivées par l'analyse de séries financières présentant des ruptures structurelles comme dans le cas d'une crise. Elles permettent ainsi d'isoler la dynamique correspondant à un état de crise de celui d'un état de non crise, en spécifiant un mécanisme de passage d'un état à un autre.
Reprenant l'approche de Tong and Lim (1980), Kwan, Li, and Ng (2010) proposent une spécification à seuil du modèle original de Tse and Tsui (2002) dans laquelle la variable de seuil est supposée connue (modèle VC-MTGARCH). Par ailleurs, la spécification de seuil ne concerne pas seulement les corrélations, mais aussi les volatilités univariées :

$$r_{i,t} = \Phi^j_{0,i} + \Phi^j_{1,i} r_{i,t-1} + e_{i,t} \tag{1.29}$$

$$h^j_{i,t} = \omega^j_i + \alpha^j_i r^2_{i,t-1} + \beta^j_i h_{i,t-1} \tag{1.30}$$

$$R_t = (1 - a^j - b^j)\overline{R} + a^j R_{t-1} + b^j \Psi_{t-1} \tag{1.31}$$

où l'indice j fait référence au $j$ème régime et i la $i$ème série. Les conditions de stationnarité impliquent que $\alpha^j_i + \beta^j_i < 1$ et $a^j + b^j \leq 1$. Dans ce modèle, la variable de seuil est endogène ou exogène retardée. Sur ce point, So and Yip (2010) proposent une version du modèle dans laquelle la variable de seuil s'écrit comme une moyenne pondérée ou une aggrégation de variables retardées (Aggregate Threshold DCC). .
Silvennoinen and Teräsvirta (2005) proposent le modèle Smooth Transition Conditional Correlation (STCC) dans lequel les corrélations évoluent entre deux matrices R_1 et R_2 constante dans le temps :

$$R_t = (1 - G_t)R_1 + G_t R_2 \tag{1.32}$$

Le passage de l'une à l'autre est réalisé pour une fonction de transition logistique conditionnelle à une variable de transition définie par :

$$G_t = (1 + e^{-\gamma(m_t - c)})^{-1} \tag{1.33}$$

où, γ le paramètre de lissage, c le paramètre de localisation et m_t la variable de transition. Cette approche de STAR a été par la suite enrichie par Silvennoinen and Teräsvirta (2009) avec l'introduction d'une seconde fonction de transition conditionnelle pour aboutir au Double STCC (DSTCC).
Si l'approche (S)TAR spécifie explicitement et de façon purement déterministe la fonction permettant de catégoriser les régimes, il est également possible de relâcher cette hypothèse en supposant que le mécanisme de passage d'un régime à un autre est de nature probabiliste. Diverses extensions basées sur un processus stochastique de changement de régime ont été formulées. Elles peuvent s'écrire comme :

$$R_t = \sum_{i=1}^{N} \mathbb{1}_{\{s_t = i\}} R_i \tag{1.34}$$

où R_i est le processus des corrélations correspondant au régime $i = 1, ..., N$. Une première possibilité pour modéliser un changement de régime de nature probabiliste consiste à supposer que les corrélations peuvent être réparties aléatoirement et indépendamment dans des composantes caractérisées par un régime spécifique. La variable indicatrice $\mathbb{1}_{\{s_t = i\}}$ est une variable latente qui permet de d'attribuer une probabilité non conditionnelle de répartition des corrélations dans les différentes classes spécifiques. Putintseva (2010) propose un modèle de mélanges finis en retenant l'équation (1.23) comme spécification pour les corrélations (modèle Mixture DCC 1). Ce type de modélisation est également retenu pour le modèle de Galeano and Ausín (2009), avec comme spécification celle de Tse and Tsui (2002) (modèle Mixture DCC 2).

Pelletier (2006) présente un modèle à corrélations dynamiques basé sur une extension du modèle de mélange de distribution : l'approche Markov-switching (Regime Switching for Dynamic Correlation, RSDC). L'approche Markov-switching ajoute à la variable indicatrice une dimension temporelle, en supposant que cette dernière suive une chaîne de Markov homogène de degré un. La structure latente permet alors d'attribuer l'appartenance à un régime conditionnellement à l'information passée. A l'instar du modèle STCC, les corrélations du modèle RSDC évoluent entre des matrices de corrélations constantes dans le temps de taille $K \times K$ ou $R_n \neq R_{n'}$ pour $n \neq n'$. Cette spécification assez lourde en paramètres est complétée dans l'article de Pelletier (2006) par une version parcimonieuse. La contribution du chapitre 2 de cette thèse présente un cas particulier du RSDC basé sur une structure hiérarchique Markovienne cachée (HRSDC).
L'approche Markov-switching est également appliquée par Billio and Caporin (2005) pour introduire un changement de régime dans l'équation (1.19) :

$$Q_t = (1 - a_{s_t} - b_{s_t})\overline{Q}_{s_t} + a_{s_t}\varepsilon_{t-1}\varepsilon'_{t-1} + b_{s_t}Q_{t-1} \tag{1.35}$$

où a_{s_t}, b_{s_t} et $\overline{Q}_{s_t}$ dépendent des valeurs de la variable aléatoire s_t. Dans la chapitre (3), nous présentons également une extension de modèle DCC basée sur une structure factorielle cachée.

Grandes matrices de corrélations conditionnelles Malgré les efforts pour limiter le nombre de paramètres les spécifications des modèles DCC restent au mieux délicates à estimer voire inapplicables pour la modélisation de grandes matrices de corrélations. Ainsi, un axe de recherche se focalise sur des spécifications permettant de prendre en compte plusieurs centaines de séries financières. Palandri (2009) propose une solution dont l'idée est de décomposer un problème numérique complexe en une somme de problèmes faciles à résoudre. Son approche, intitulée Sequential Conditional Correlation (SCC) est construite sur une expression séquentielle de la matrice de corrélation, en séparant les corrélations et les corrélations partielles :

$$R_t = [\prod_{i=1}^{K-1}\prod_{j=i+1}^{K} K_{i,j,t}][\prod_{i=1}^{K-1}\prod_{j=i+1}^{K} K_{i,j,t}]' \tag{1.36}$$

où matrice $K_{i,j,t}$ est une matrice triangulaire inférieure dont les éléments sont :

$$K_{i,j,t}[row, col] = \begin{cases} \rho_{i,j,t} \text{ si } row = j \text{ et } col = i \\ \left(1 - \rho^2_{i,j,t}\right)^{1/2} \text{ si } row = j \text{ et } col = j \\ I_{[row,col]} \text{ sinon} \end{cases}$$

et I est la matrice identité. Ce modèle n'impose pas de forme particulière pour la formulation des corrélations et n'importe quelle spécification peut être utilisée[9].
L'idée de décomposer un problème compliqué en une somme de problèmes plus simples est également exploitée par Engle (2008) avec la méthode *MacGyver*. L'idée est d'estimer chaque corrélation séparément et ensuite de retenir la médiane de ces estimateurs.
Une autre stratégie est déployée par Engle and Kelly (2009) avec le modèle Dynamic Equicorrelation (DECO). Cette formulation suppose que toutes les paires de rendements possèdent la même corrélation à l'instant t, mais que cette corrélation change à chaque période. Dans ce modèle, la matrice de corrélation s'écrit :

$$R_t = (1 - \rho_t)\mathrm{I}_K + \rho_t \mathrm{J}_{K \times K} \tag{1.37}$$

[9] Palandri (2009) expose d'ailleurs dans son analyse une spécification originale pour le processus des corrélations basée sur la transformation de Fisher.

avec $\mathrm{I_K}$ une matrice identité de taille K et $\mathrm{J_{K\times K}}$ une matrice carrée de un de taille K. Les équicorrélations sont représentées par la variable ρ et se calculent comme :

$$\rho_t = \frac{1}{K(K-1)} \sum_{i \neq j} \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}} \tag{1.38}$$

où $q_{ij,t}$ est un élément de la matrice de covariance Q_t issue d'un modèle DCC. Outre son hypothèse simplificatrice d'équicorrélation, l'expression des corrélations définit par l'équation (1.37) présente un avantage numérique dans la mesure où l'inverse et le déterminant de R_t s'écrirent sous une forme analytique[10].

Structures par arbre de décision Plusieurs approches basées sur un arbre de décision ont été suggérées. Barone-Adesi and Audrino (2006) élaborent une structure par arbre développée précédemment dans le cadre univarié (modèle Tree-GARCH, voir Audrino and Buhlmann (2001) et Audrino (2006)).

$$R_t = \left(1 - \sum_{k=1}^{N} \lambda_k \mathbb{1}_{\{(\overline{\rho}_{t-1}, r_{t-1}) \in \mathscr{R}_i\}}\right) \overline{Q}_{t-p}^{t-1} + \left(\sum_{k=1}^{N} \lambda_k \mathbb{1}_{\{(\overline{\rho}_{t-1}, r_{t-1}) \in \mathscr{R}_i\}}\right) \mathrm{I}_d \tag{1.39}$$

où $\lambda_k \in]0,1] \forall k$, $\overline{Q}_{t-p}^{t-1}$ est une matrice de covariance non conditionnelle sur un horizon de p périodes. La partition est effectuée dans la plan $(\overline{\rho}_{t-1}, r_{t-1})$ où $\overline{\rho}_{t-1}$ est une moyenne pondérée des éléments de la matrice de corrélation variant dans le temps. Ce modèle nommé Rolling Window Tree Averaging Conditional Correlations (RW-TACC) nécessite un faible nombre de partitions pour être exploitable en pratique. Cette approche est par la suite perfectionnée dans un cadre non paramétrique par Audrino (2006). Par ailleurs, une autre spécification, toujours inspirée de la méthode développée pour le Tree-GARCH, est appliquée par Trojani and Audrino (2005). Dans ce modèle un Tree-GARCH est appliqué pour extraire les volatilités individuelles et une structure par arbre similaire est appliquée pour le processus des corrélations. Dans ce cas, la structure par arbre ne concerne pas seulement les corrélations mais aussi les dynamiques des volatilités univariées.

Dellaportas and Vrontos (2007) développent également une structure par arbre de décision binaire établissant un lien entre la volatilité individuelle et les corrélations.

Le chapitre 4 de cette thèse propose également un modèle par arbre. A la différence des spécifications précédentes, basées sur une approche déterministe des arbres de décision, nous proposons une approche stochastique basée sur une extension du modèle HMM.

Ainsi, suite aux articles pionniers de Tse and Tsui (2002) et de Engle and Sheppard (2001) introduisants une dynamique des corrélations dans les modèles à corrélations conditionnelles, de nombreux développements ont été proposés. Nous venons d'évoquer dans cette section la plupart d'entre elles. Il n'en reste pas moins que ce bref survol de la littérature reste loin d'être exhaustif. De nombreux compléments figurent dans le survey de Bauwens, Laurent, and Rombouts (2006), qui présente une typologie des GARCH multivariés, ainsi que l'article de synthèse de Silvennoinen and Teräsvirta (2009). La figure 1.1 présente une cartographie des différentes extensions du modèle à corrélations conditionnelles dynamiques en prenant comme point de départ les approches *historiques* de Tse and Tsui (2002) et de Engle and Sheppard (2001).

[10] Dans le cas de grands systèmes de corrélation, l'inversion de R_t à chaque instant t quand cette dernière n'a pas d'expression analytique simple pose des problèmes numériques qui se traduisent par des durées d'estimation longues. La possibilité d'avoir l'expression analytique de R_t^{-1} est donc un avantage non négligeable.



Changement de régime :
- *VC MTGARCH* : Multivariate Threshold Varying Conditional Correlations Model, Kwan, Li, and Ng (2010)
- *AT MTGARCH* : Aggregate Threshold DCC, So and Yip (2010)
- *STCC* : Smooth Transition Conditional Correlation, Silvennoinen and Teräsvirta (2005)
- *DSTCC* : Double Smooth Transition Conditional Correlation, Silvennoinen and Teräsvirta (2009)
- *RSDC* : Regime Switching for Dynamic Correlation, Pelletier (2006)
- *HRSDC* : Hierarchical RSDC, chapitre 2 de cette thèse
- *MS-DCC* : Markov-switching DCC, Billio and Caporin (2005)
- *FHM-DCC* : DCC with Factorial Hidden Markov decompostion, chapitre 3 de cette thèse
- *Mix-DCC 1* : Mixture DCC (approche Engle), Putintseva (2010)
- *Mix-DCC 2* : Mixture DCC (approche Tse-Tsui), Galeano and Ausín (2009)

Ecritures matricielles :
- *F-DCC* : Flexible DCC, Billio, Caporin, and Gobbo (2006)
- *G-DCC* : Generalized DCC, Hafner and Franses (2009)
- *BD-DCC* : Block Diagonal DCC, Billio and Caporin (2009)

Strutures par arbre de décision :
- *RWTA-DCC* : Rolling Window Tree Averaging Conditional Correlations, Barone-Adesi and Audrino (2006)
- *Tree DCC 1* : Tree structured DCC, Trojani and Audrino (2005)
- *Tree DCC 2* : Tree structured DCC, Dellaportas and Vrontos (2007)
- *HMDT DCC* : Hidden Markov Decision Tree based DCC, chapitre 4 de cette thèse

Semi/non-paramétrique :
- *NP DCC* : Non Parametric DCC, Long and Ullah (2005)
- *L DCC* : Local DCC, Feng (2007)
- *SPCC* : Semi Parametric DCC, Franses, Hafner, and van Dijk (2005)

Simplifications :
- *Clus. DCC 1* : Clustered DCC, Zhou and Chan (2008)
- *Clus DCC 2* : Clustered DCC, So and Yip (2009)
- *PS DCC* : Proximity Structured DCC, Caporin and Paruolo (2009)

Asymétrie :
- *ADCC* : Asymmetric DCC, Cappiello, Engle, and Sheppard (2006)
- *AB DCC* : Asymmetric Block DCC, Vargas (2006)
- *AML DCC* : Asymmetric Multivariate Laplace Innovations, Cajigas and Urga (2006)

Grandes matrices de corrélations :
- *SCC* : Sequential Conditional Correlation, Palandri (2009)
- *DECO* : Dynamic Equicorrelation, Engle and Kelly (2009)
- *MacGyver DCC* : MacGyver estimation based DCC, Engle (2008)

FIG. 1.1: Cartographie des différents axes de développements des modèles à corrélations conditionnelles intégrant une spécification dynamique.

1.3 Cadre général des modèles graphiques probabilistes.

1.3.1 Origines et formalisme

Les modèles graphiques probabilistes (MGP), aussi appelés réseaux bayésiens, représentent un formalisme issu de la recherche en intelligence artificielle. Quoique trouvant ses origines dans les années soixante, cette approche a connu un développement important dans les années quatre-vingt dix afin de modéliser des problèmes complexes provenant de la reconnaissance d'écriture manuscrite, de la séparation de sources auditives, de la reconnaissance d'images ou encore de la classification de données.
Ce formalisme repose sur deux disciplines : la théorie des probabilités et la théorie des graphes[11]. Dans ce cadre, l'impact d'une variable, d'un système, d'un évènement sur un/une autre est représenté par un graphe. Cette représentation graphique de la causalité permet de modéliser une relation de dépendance entre ces variables ; cette relation est alors quantifiée avec des probabilités. Ainsi, plus formellement, un MGP peut être représenté par le couple $\{\mathcal{G}, \{p\}\}$ tel que :

- $\mathcal{G} = (X, E)$ est un graphe acyclique orienté dans lequel X représente les noeuds mais aussi des variables aléatoires et E les arcs, permettant de décrire les relations de dépendances entre les variables.
- $\{p\}$ représente un ensemble de lois de probabilités conditionnelles, où chaque élément correspond à la distribution de X_i, $i = 1, ..., n$, conditionnellement à l'ensemble de ses variables parentes.

Un MGP est donc un formalisme possédant une dimension qualitative, avec un graphe acyclique orienté représentant les relations causales du modèle, et une dimension quantitative, représentée par une probabilité conditionnelle.
Les MGP que nous considérons dans cette thèse sont construits sur une hypothèse Markovienne. En d'autres termes, la caractéristique de ce cas particulier de MGP suppose que les variables du modèle à l'instant t dépendent de l'instant $t-1$. Le modèle à chaîne de Markov caché (appelé selon son abréviation anglaise HMM, pour Hidden Markov Model) constitue ainsi le modèle de référence de cette classe particulière de MGP. La section suivante rappelle les principales propriétés de ce modèle.

1.3.2 Modèle de Markov caché

Origines Le modèle HMM trouve ses origines dans la construction des modèles de mélange de lois[12]. L'objectif du modèle de mélanges consiste à estimer une densité de probabilité en supposant que cette densité est un mélange fini de densités. En notant par $\{\mathcal{Y}_t\}_{t\in\mathbb{N}^*} = \{y_1, ..., y_T\} \subset \mathbb{R}^k$ un vecteur d'observations indépendantes de dimension k ayant pour fonction de densité $f(y_t)$ sur $\mathbb{R}^k$, la densité de $\{\mathcal{Y}_t\}$ a pour expression :

$$f(y_t) = \sum_{i=1}^{N} \pi_i f(y_t; \phi_i) \tag{1.40}$$

où $f(y_t; \phi_i)$ désigne une de densité de probabilité de paramètres ϕ_i. Le paramètre $\pi_i \in (0,1)$ correspond à la proportion de la composante i dans le mélange. Pour que la fonction $f(y_t)$ soit une densité de probabilité, la suite positive $(\pi_i)_{i\in\mathbb{N}^*}$ doit alors vérifier :

$$\sum_{i=1}^{N} \pi_i = 1 \tag{1.41}$$

[11] "Graphical models are a marriage between probability theory and graph theory." (Jordan (1999)).
[12] Voir par exemple l'ouvrage de McLachlan and Peel (2000) pour un exposé général et l'article de Böhning, Seidel, Alfo, Garel, Patilea, and Walther (2007) pour un survey des derniers développements.

L'estimation du modèle de mélange tel qu'il vient d'être présenté est particulièrement délicate pour la simple raison que l'on ne peut affecter les observations aux composantes du mélange. Une solution usuelle à ce problème consiste alors à reconsidérer le modèle de mélange dans un cadre de données incomplètes. Cette méthode associe à chaque vecteur y_t la variable indicatrice $\{\mathscr{S}_t\}_{t\in\mathbb{N}^*} = \{s_1, ..., s_T\} \subset \mathbb{R}$ telle que :

$$\mathbb{1}_{\{s_n=i\}} = 1 \tag{1.42}$$

Le modèle de mélanges en données incomplètes est donc représenté par le couple $\{\mathscr{Y}_t, \mathscr{S}_t\}_{t\in\mathbb{N}^*}$. Les données étant supposées réparties dans différentes composantes, le poids d'une composante peut s'interpréter comme la probabilité qu'une observation appartienne à la composante i : $\pi_i = \mathbb{P}[s_t = i]$. Le modèle est alors construit à partir d'une structure sous-jacente latente formée de la variable indicatrice $iid\ \mathbb{1}_{\{s_n=i\}}$. La probabilité conditionnelle et la probabilité jointe des observations s'écrivent respectivement :

$$\mathbb{P}[y_t|s_t] = \sum_{i=1}^{N} f(y_t;\phi_i)\mathbb{1}_{\{s_n=i\}} \text{ et } \mathbb{P}[y_t, s_t] = \sum_{i=1}^{N} \pi_i f(y_t;\phi_i)\mathbb{1}_{\{s_n=i\}} \tag{1.43}$$

Le modèle de mélanges de lois peut être vu comme un modèle à variable latente en supposant qu'aux données observées correspond une indicatrice d'appartenance à une composante. Cette approche peut être enrichie en supposant que l'indicatrice suit une chaîne de Markov cachée, ce qui permet d'introduire une dimension temporelle sur l'indicatrice des composantes.

Chaîne de Markov Rappelons brièvement les principales définitions et propriétés nécessaires concernant les chaînes de Markov[13]. Soit $\mathscr{S}$ un espace dénombrable appelé *espace d'états*, dans lequel chaque $i \in \mathscr{S}$ est appelé *état*. Une chaîne de Markov d'ordre k est alors définie par une suite de variables aléatoires $(s_t)_{t\in\mathbb{N}}$ à valeur dans $\mathscr{S}$ telle que, pour tout $i_t \in \mathscr{S}$, $t \geq 1$:

$$\mathbb{P}[s_{t+1} = i_{t+1}|s_t = i_t, ..., s_0 = i_0] = \mathbb{P}[s_{t+1} = i_{t+1}|s_t = i_t, ..., s_{t-k+1} = i_{t-k+1}] \tag{1.44}$$

Les modèles de mélange basés sur une chaîne de Markov utilisent généralement une chaîne de Markov d'ordre un[14], qui admet une interprétation intuitive simple : *l'état futur ne dépend que de l'état présent, et pas de l'état passé*. La dynamique est donc entièrement basée sur la probabilité de transition entre les différents états, représentée par :

$$a_{ij} = \mathbb{P}[s_{t+1} = j|s_t = i] \tag{1.45}$$

La figure 1.2 est un exemple de représentation graphique d'une chaîne de Markov à trois états. Une chaîne de Markov $(s_t)_{t\in\mathbb{N}}$ est alors dite homogène si les probabilités de transition d'un état à un autre sont indépendantes de t :

$$\mathbb{P}[s_{t+1} = j|s_t = i] = \mathbb{P}[s_1 = j|s_t = i] \tag{1.46}$$

Les probabilités de passage d'un état à un autre d'une chaîne de Markov homogène à n états permettent d'établir une matrice de transition :

$$\mathscr{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{1.47}$$

[13] Pour un exposé plus complet, nous renvoyons à (Billingsley, 1986, chap. 1.8) and (Shiryaev, 1996, chap. 1.12 and 8).
[14] $\Pr[S_{t+1} = s_{t+1}|S_t = s_t, ..., S_0 = s_0] = \Pr[S_{t+1} = s_{t+1}|S_t = s_t]$.

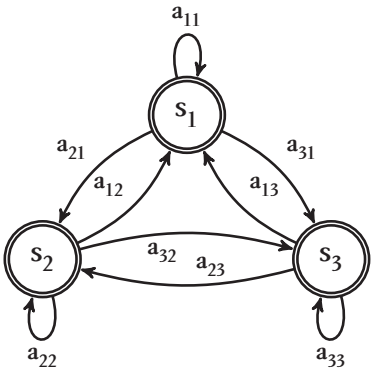


FIG. 1.2: Représentation des probabilités de transition d'une chaîne de Markov à trois états.

avec :

$$\sum_j a_{ij} = 1 \text{ et } 0 \leq a_{ij} \leq \tag{1.48}$$

Toute matrice de transition peut être représentée par un graphe orienté valué dont les sommets représentent les états de la chaîne. Ainsi, en supposant que la distribution initiale des états s'écrivent π_t, la distribution à la période $t+1$ s'obtient par : $\pi_{t+1} = \pi_t \mathscr{A}$. Le comportement de la matrice de transition pour des instants supérieurs à 1 s'étudie via le théorème de Chapman-Kolmogorov. La matrice de transition de l'instant 0 à l'instant m s'écrit :

$$\mathscr{A}^{(m)} = \begin{bmatrix} a_{11}^{(m)} & \cdots & a_{1n}^{(m)} \\ \vdots & \ddots & \vdots \\ a_{n1}^{(m)} & \cdots & a_{nn}^{(m)} \end{bmatrix} = \mathscr{A}^m \tag{1.49}$$

où $a_{ji}^{(m)}$ représente la probabilité d'aller de l'état i à l'instant 0 vers l'état j à l'instant m et pour tout $n, k \in \mathbb{N}$ et $h = 1, ..., k-1$:

$$a_{ji}^{(k)} = \sum_{l \in S} a_{li}^{(h)} a_{lj}^{(k-n)} \tag{1.50}$$

Il est alors possible d'identifier différents types d'états en remarquant que l'espace $\mathscr{S}$ peut être partitionné en différentes classes d'états. Pour i et j deux états de la chaîne, on dit que :
– i *mène à* j, noté $i \to j$, si $\mathbb{P}[s_{t+k} = j | s_t = i] > 0$.
– i et j *communiquent*, $i \leftrightarrow j$, si $i \to j$ et $j \to i$.
– un état est *récurrent* si $\mathbb{P}[\cup_{n=1}^{\infty}(s_n = i)|s_0 = i] = 1$.
– un état est *transitoire* si $\mathbb{P}[\cup_{n=1}^{\infty}(s_n = i)|s_0 = i] < 1$[15].
– une chaîne de Markov dans laquelle tous les états communiquent est dite *irréductible*.

Modèle HMM L'idée du modèle à chaîne de Markov caché (HMM) est que l'état du système n'est pas spécifié par l'observation mais grâce à un processus caché qui établit une probabilité de correspondance entre une observation et un état. Chaque observation y_t est donc reliée à une variable aléatoire s_t suivant une chaîne de Markov par une probabilité. L'hypothèse selon laquelle les variables aléatoires associées aux observations suivent une chaîne de Markov d'ordre un implique

[15] La quantité $\mathbb{P}[\cup_{n=1}^{\infty}(S_n = i)|S_0 = i]$ s'interprète comme la probabilité conditionnelle de retour à l'état i sachant que la chaîne est démarrée à l'état i.

une dimension temporelle puisque la probabilité d'être dans un état en $t+1$ dépend de l'état en t. La différence principale avec le modèle de mélange indépendant porte donc sur l'indicatrice, qui admet une dépendance temporelle. La variable aléatoire s_t suit généralement (mais pas forcément) une chaîne de Markov d'ordre un mais surtout respecte les conditions d'homogénéité, d'irréductibilité et d'ergodicité. En notant par N le nombre d'états, un modèle HMM correspond au couple $\{\mathscr{Y}_t, \mathscr{S}_t\}_{t\in\mathbb{N}}$. Il est donc caractérisé par trois éléments :

1. un vecteur des transitions initiales $\nu = \mathbb{P}[s_1 = i]\ \forall i \in S$ de taille $(N \times 1)$.
2. une matrice de transition $A = (a_{ij})$ de taille $(N \times N)$ telle que

$$a_{ij} = \mathbb{P}[s_{t+1} = j | s_t = i] \forall i, j \in S \tag{1.51}$$

3. des probabilités d'observations indépendantes conditionnellement aux états cachés :

$$\mathbb{P}[y_t | s_t = i] = f(y_t; \phi_i) \tag{1.52}$$

où ϕ_i correspond aux paramètres d'une famille de loi discrète paramétrique. Dans le cas discret, la loi conjointe suivie par le modèle de Markov caché défini par le couple $\{\mathscr{Y}_t, \mathscr{S}_t\}_{t\in\mathbb{N}}$ a pour expression :

$$\mathbb{P}[y_{1:T}, s_{1:T}] = \mathbb{P}[y_{1:T} | s_{1:T}] \times \mathbb{P}[s_{1:T}] = \prod_{t=1}^{T} \mathbb{P}[y_t | s_t] \times \prod_{t=1}^{T-1} \mathbb{P}[s_{t+1} | s_t] \times \mathbb{P}[s_1] \tag{1.53}$$

Pour représenter le modèle HMM graphiquement sous forme de réseaux bayésiens dynamiques (Dynamic Bayesian Network, DBN), la convention graphique consiste à représenter le modèle par tranche de temps t successives, comme c'est le cas sur la figure 1.3.

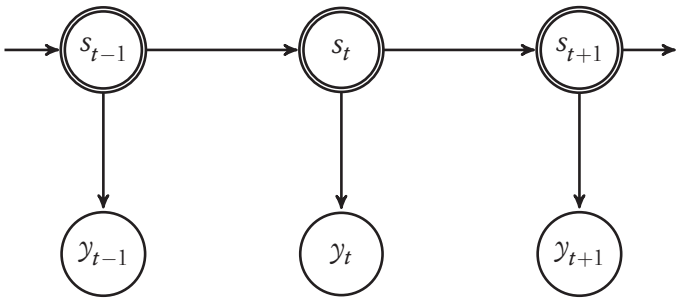


FIG. 1.3: Représentation d'un HMM sous forme de DBN.

1.3.3 Extensions

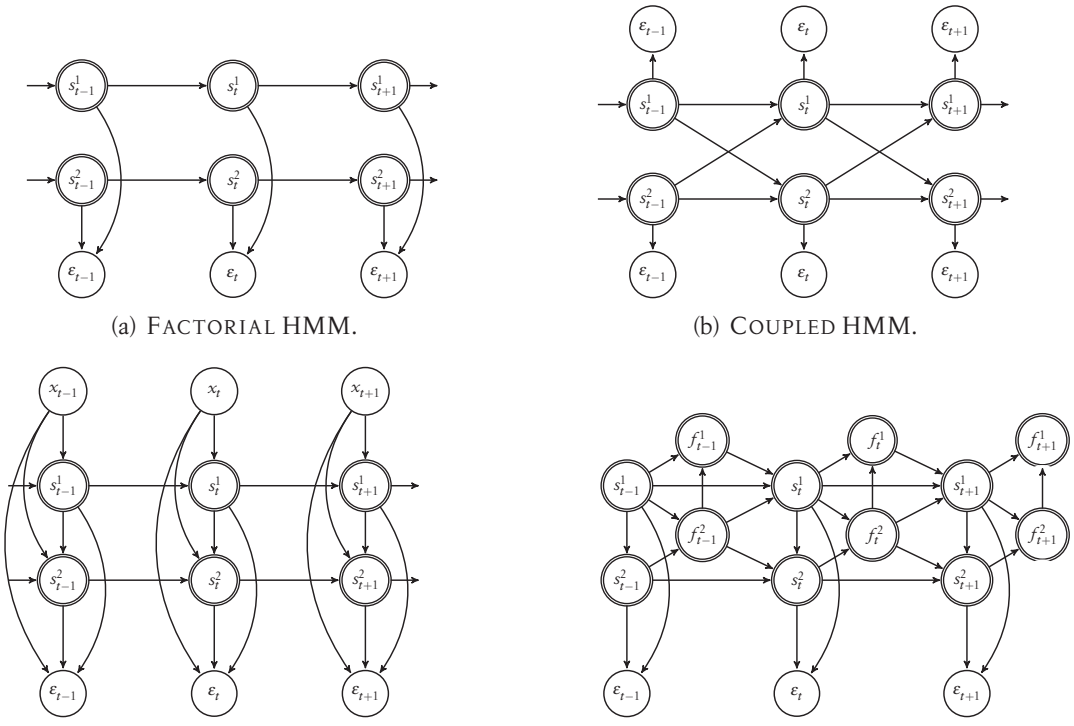
Le modèle HMM constitue l'un des plus simple MGP. De nombreuses extensions sont apparues dans la littérature, répondant à des besoins spécifiques de modélisation. Nous présentons ici celles que nous avons utilisé dans les contributions de cette thèse.

HMM factoriel Introduit par Ghahramani and Jordan (1997), le HMM factoriel (FHMM) suppose que l'état caché est décomposé en plusieurs états sous une forme factorisée. La chaîne de Markov principale du modèle est donc décomposée en plusieurs chaînes évoluant en parallèle. Une représentation graphique du FHMM sous forme de réseau bayésien dynamique (DBN) est visible sur la figure 1.4(a), où $\{s^1\}$ et $\{s^2\}$ sont deux chaînes de Markov évoluant en parallèle.

HMM couplé Le HMM couplé (Coupled HMM, CHMM) est exposé en détail dans l'étude de Brand (1997). Le modèle est bâti sur l'hypothèse qu'une chaîne de Markov peut avoir une influence sur une autre chaîne de Markov du modèle. Cette extension relâche donc l'hypothèse d'indépendance des chaînes du modèle FHMM, comme l'illustre la figure 1.4(b). Dans cette $\{s^1\}$ et $\{s^2\}$ sont deux chaînes de Markov évoluant en parallèle mais possédant des interdépendances.

Arbre de décision Markovian Jordan, Ghahramani, and Saul (1997) proposent un arbre de décision stochcastique (Hidden Markov Decision Tree, HMDT). Ce modèle est un fait une spécification hybride puisque qu'un HMDT est un HMM qui est factorisé et couplé. C'est donc le mariage entre le FHMM et le CHMM. Ce modèle est représenté sur la figure 1.4(c). Ce modèle autorise la présence d'un *input*, représenté par la variable x_t, pour conditionner les dépendances de l'arbre.

HMM hiérarchique Le HMM hiérarchique (Hierarchical HMM, HHMM) est une extension du modèle HMM de base introduite par Fine, Singer, and Tishby (1998). L'idée est de construire un processus stochastique à plusieurs niveaux en adoptant une structure par arbre afin d'obtenir un entrelacement de régimes. Une représentation sous forme de DBN est donnée sur la figure 1.4(d). La composante f_t^i, $i = 1, 2$, du graphique est une variable indicatrice prenant la valeur un ou zéro suivant la profondeur de l'arbre nécessaire pour produire une observation. Il est peut être nécéssaire de préciser que la représentation DBN n'est certainement pas la représentation la plus intuitive qui soit pour représenter ce modèle. Une représentation sous forme d'arbre est visible dans le chapitre 2.



(a) FACTORIAL HMM. (b) COUPLED HMM.

(c) HIDDEN MARKOV DECISION TREE. (d) HIERARCHICAL HMM.

FIG. 1.4: Quatre extensions de HMM utilisés dans cette thèse, où la variable s_t représente une chaîne de Markov, ε_t les observations, x_t une variable d'input et f_t une variable indicatrice.

Cette présentation des MGP et des extensions HMM présentée dans cette section est volontairement brève. Nous nous sommes limités à présenter les concepts généraux nécessaires à la lecture de cette thèse. Le lecteur intéressé trouvera de nombreux développements, notamment concernant l'apprentissage de ces modèles et l'inférence dans les ouvrages de Jordan (1999) et Bishop (2006). L'article de Ghahramani (1998) offre un exposé concis de ces modèles tandis que la thèse de Murphy (2002) explicite d'une manière exhaustive la théorie des MGP. Une présentation des modèles HMM est présentée dans l'incontournable article de Rabiner (1989) ; la monographie de Cappé, Moulines, and Ryden (2005) fournit un exposé théorique complet du modèle HMM.

1.4 Constributions de cette thèse

Les contributions de cette thèse se placent dans la littérature des modèles GARCH multivariés à corrélations conditionnelles dynamiques. L'apport des travaux qui y sont présentés concerne la façon de modéliser les changements de régimes avec des modèles à chaîne de Markov cachée. En effet, l'approche Markov-Switching[16] introduite par Lindgren (1978) en économie puis fortement popularisée par Hamilton (1989), a donné lieu à une abondante littérature[17].
Une caractéristique commune aux travaux de cette littérature est l'utilisation du modèle de base, à savoir le HMM. Dans cette thèse, nous avons considéré le modèle de base comme trop restrictif. Aussi, les trois modèles que nous présentons sont construits à partir d'extensions du HMM issues des travaux du domaine des MGP.

Contribution du chapitre 2 Ce chapitre présente un cas particulier du modèle RSDC Pelletier (2006) appelé Hierarchical-RSDC (HRSDC). Il est construit avec une structure markovienne cachée hiérarchique (HHMM) introduite par Fine, Singer, and Tishby (1998). Le gain de cette approche par rapport au Markov-Switching classique est d'augmenter la granularité des régimes. Le modèle HRSDC est comparé au modèle DSTCC, de classe STAR, présenté par Silvennoinen and Teräsvirta (2009) dont il est l'analogue dans la classe Markov-Switching. Nous présentons également deux applications : une première sur données simulées et une seconde sur des données réelles.

Contribution du chapitre 3 Cette seconde contribution présente un modèle à changement de régime basée de la spécification de Engle and Sheppard (2001). Cette approche avait déjà été formulée

[16] Markov-switching et HMM sont deux termes renvoyant au même concept. Cependant, ces deux dénominations ne sont pas encore vraiment standardisées et un modèle Markov-switching sera souvent qualifié de HMM (voir Cappé (2005, chap. 1.2).

[17] Citons, sans être exhaustif, le travail de Hamilton (1990) pour la mise au point d'un filtre d'estimation des probabilités de transition, ainsi que Krolzig (1997) pour une extension très complète du modèle original au cadre multivarié. Kim (1994) introduit un filtre de lissage des probabilités affinant le calcul des probabilités a posteriori.

Le premier modèle ARCH avec Markov-switching (SWARCH) est proposé par Hamilton and Susmel (1994) ; une première généralisation au GARCH (GRS) apparaît dans une analyse des taux d'intérêts courts de Gray (1996). L'innovation du modèle développé par Gray réside dans l'écriture de la variance conditionnelle. Pour le modèle SWARCH, chaque variance conditionnelle dépend non seulement du régime, mais aussi de tout l'historique[18]. Dans le modèle GRS, la variance conditionnelle dépend seulement du régime courant en intégrant l'espérance de la variance passée. Klaassen (2002) améliore ce dernier modèle en élargissant le spectre de l'information pour la variance retardée. Haas, Mittnik, and Paollela (2004b) proposent une approche un peu différente avec le modèle MSG(N) qui consiste à estimer N modèles GARCH parallèlement et à sauter d'une variance à une autre.Ils contournent ainsi le problème de path-dependancy. Cette solution généralise l'approche GARCH avec Markov-switching et plusieurs modèles, comme ceux de Haas, Mittnik, and Paollela (2004a) et de Franq, Roussignil, and Zakoian (2001) apparaissent comme des cas particuliers. Enfin, Liu (2006) finalise le modèle MSG(N) en étudiant ses conditions de stationnarités avec des hypothèses moins restrictives que celles de Haas, Mittnik, and Paollela (2004b).

Les modèles GARCH multivariés basés sur une approche Markov-switching ont été évoqué dans la section 1.2.2.

par Billio and Caporin (2005). Dans ce modèle, les corrélations évoluent entre deux régimes, dont le passage de l'un à l'autre s'effectue via une chaîne de Markov cachée. Dans cette configuration, tous les éléments de la matrice des corrélations conditionnelles sont alors soumis à la même dynamique de saut. Dans ce chapitre, cette hypothèse est considérée comme trop restrictive et nous permettons à tous les éléments de la matrice de corrélation d'avoir leur propre dynamique de saut. Cette possibilité de dynamique des corrélations par paire est réalisée en utilisant une structure Markovienne factorisée, proposée par Ghahramani and Jordan (1997).

Contribution du chapitre 4 Cette dernière contribution présente un arbre de décision mettant un établissant un lien entre les volatilités individuelles et les corrélations. Diverses modélisations par arbre de décision ont déjà été proposées dans le domaine des modèles à corrélations conditionnelles. Cependant, ces modèles adoptent tous une structure déterministe. Dans ce travail, nous utilisons une structure stochastique provenant de l'approche développée par Jordan, Ghahramani, and Saul (1997), le modèle HMDT. Dans un premier temps, nous utilisons une structure Markovienne factorisée afin de discriminer entre des régimes de faibles et de fortes volatilités ainsi qu'entre un régime de faibles ou fortes corrélations. Le lien permettant de relier des états spécifiques de volatilité à un état spécifique de corrélations est alors réalisé en couplant les chaînes du modèle factorisé.

CHAPITRE 2

Hierarchical Hidden Markov Structure for Dynamic Correlations : the Hierarchical RSDC Model.

2.1 Introduction[1]

Since the seminal papers of Engle (2002) and Engle and Sheppard (2001), the study of multivariate GARCH with dynamic correlations has given rise to many extensions and developments. This has been motivate by several empirical studies about stock market behaviours (like Longin and Solnik (1995, 1996, 2001)). They show that the hypothesis of constant correlations (CCC model of Bollerslev (1990)) is not realistic. However, if the assumption of dynamic correlations is now widely accepted in the literature, the question of the form of the dynamic is still an open problem ; see Bauwens, Laurent, and Rombouts (2006) or Silvennoinen and Teräsvirta (2009) for a recent survey.

The purpose of this paper is to present a new multivariate GARCH model with dynamic correlations. We propose a regime switching model that is part of the Markov-Switching class. It is a special case of the RSDC model of Pelletier (2006). This model is halfway between CCC of Bollerslev (1990) and DCC of Engle and Sheppard (2001). Correlations are constant within regime, but vary from one regime to another and the transition between the different regimes are performed by a Markov chain. Silvennoinen and Teräsvirta (2005) have proposed a model with smooth transition between regime for correlations (STCC model), which can be seen, under certain assumptions, as the competitive STAR approach of Pelletier's model. The STCC impose the correlations to varying between two matrix of constant correlations. The transition between this two extreme matrix is governed by a conditional logistic function. Recently, they build an extension of this model, the Double-STCC (see Silvennoinen and Teräsvirta (2007)) in which conditional correlations vary across four matrix of constant correlations through two logistic functions. Our new model, called Hierarchical-RSDC (HRSDC) can be seen as a Markov-Switching version of the DSTCC. In this new model, correlations vary between four correlation matrix constant in time but the transition from one to another is determined by hierarchical hidden Markov structure. The originality of this structure is to establish a hierarchy between the hidden states in order to increase the granularity of the regime. This hierarchical hidden structure was first developed by Fine, Singer, and Tishby (1998) for handwriting recognition. In our context, this specific structure will allow us to bring out a finer definition of regimes that do not appear with the classical Markov-Switching approach.

The paper is organized as follows. In section 2.2, we briefly give a review of the dynamic conditional correlations models and present a overview of regime switching model, including STAR and Markov-switching approach, applied to correlation modeling. The Hierarchical-RSDC model is introduced in section 2.3.1. Section 2.4 presents results of Monte-Carlo simulations and two empirical applications and compare the HRSDC with the DSTCC and DCC models. Section 2.5 presents concluding remarks and expose some directions for future research.

2.2 DCC-MGARCH and Regime Switching models

The general framework of multivariate GARCH models with dynamic correlations assumes that a the stochastic process $\mathbf{r}_t$ followed by the observations of size $(K \times T)$ is defined by :

$$\mathbf{r}_t | \mathscr{F}_{t-1} \sim \mathscr{L}(0, H_t) \tag{2.1}$$

where $\mathscr{L}$ is a parametric distribution function with mean equal to zero and conditional variance H_t follows :

$$H_t = D_t R_t D_t \tag{2.2}$$

[1] Joint work with Vêlayoudom MARIMOUTOU.

with $D_t = \text{diag}(h_{1,t}^{1/2}, ..., h_{K,t}^{1/2})$, a diagonal matrix composed of the standard deviation of the K univariate series, obtained for example with a GARCH filtering. This filtering permits to calculate the standardized residuals, which are expressed as $\varepsilon_t = D_t^{-1}\mathbf{r}_t$. The expectation of these standardized residuals gives the conditional correlations :

$$\mathbb{E}_{t-1}[\varepsilon_t \varepsilon_t'] = D_t^{-1} H_t D_t^{-1} = R_t \tag{2.3}$$

The first multivariate volatility model to take into account the conditional correlations was the CCC of Bollerslev (1990). He adopts the simplest dynamic ever by taking the conditional correlations constant through time, i.e. $R_t = R \ \forall t$. The correlations matrix is then equal to the unconditional correlation matrix. However, several empirical studies show that correlations among assets are not constant. These results imply that the CCC seems too restrictive and give rise to an extensive literature about the process follows by the conditional correlations.
Tse and Tsui (2002) suggested to express the correlations with an autoregressive form (DCC_T model) :

$$R_t = (1 - \theta_1 - \theta_2)R + \theta_2 R_{t-1} + \theta_2 \Psi_{t-1}, \ \theta_1 + \theta_2 \leq 1 \tag{2.4}$$

where R is the unconditional correlation matrix while the elements of the correlation matrix Ψ_{t-1} are defined as :

$$\psi_{ij,t-1} = \varepsilon_{i,t-1}\varepsilon_{j,t-1} / \sqrt{\varepsilon_{i,t-1}^2 \varepsilon_{j,t-1}^2} \text{ with } 1 \leq i \leq j \leq K \tag{2.5}$$

In the same way, Engle and Sheppard (2001) adopt an autoregressive formulation for R_t (DCC_{ES} model) with :

$$R_t = \text{diag}\{Q_t\}^{-1/2} Q_t \text{diag}\{Q_t\}^{-1/2} \tag{2.6}$$

where Q_t is the conditional covariance matrix such as :

$$Q_t = (1 - \alpha - \beta)\overline{Q} + \alpha \varepsilon_{i,t-k} \varepsilon_{j,t-k}' + \beta Q_{t-l} \tag{2.7}$$

with $\overline{V}$ the unconditional covariance matrix. These initial formulations define the basic framework of dynamic correlations models and represent the starting point of many extensions and we refer to Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009) for recent surveys of this field of research.
In this work, we are interested by two competing approaches is to modeling regime switching for the correlations. The first one is the STAR type model of Silvennoinen and Teräsvirta (2005). The second is from Pelletier (2006) and is builded in a Markov-Switching setup. Recall that this two main classes of models -STAR and Markov-Switching- are build on two ways to treat the problem. The STAR class assumes that the switch between the regimes is deterministic and determinate by a observable transition variable. The Markov-Switching class is quite different. It assumes that the process followed by the observation is governed by an underlying hidden stochastic process. To better understand our proposed model, both approaches are described in more details in the following subsection.

2.2.1 (S)TAR approach

Theorical aspects The Threshold Autoregressive (TAR) model of Tong and Lim (1980) was one of the first approach to model regime switching. It assumes that at each time t regime is determined by a threshold variable and a threshold value. A two regimes model will then be composed by two parallel sub-regimes where the transition from one to another depends of the position of the threshold variable against the threshold value. The SETAR model, for Self-Exciting TAR, is an direct extension in which the threshold variable is the lagged time series itself.

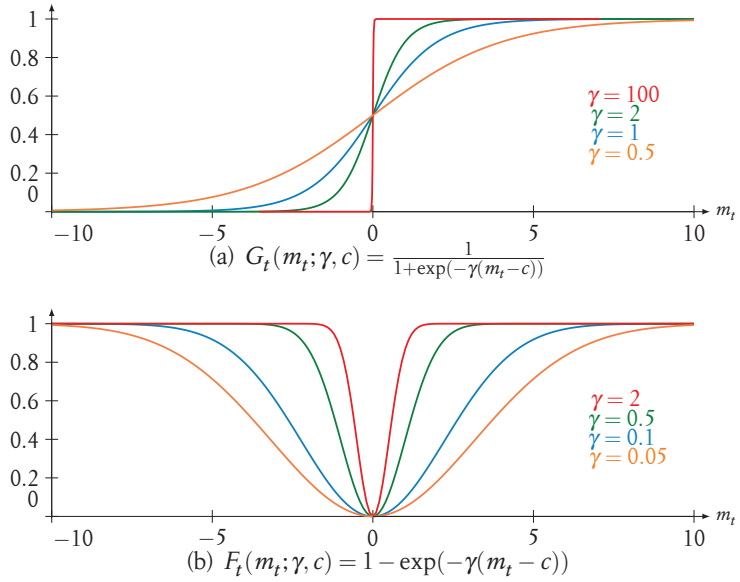


FIG. 2.1: Examples of logistic and exponential functions (resp. subfigures 2.1(a) and 2.1(b)) with $c = 0$ and different values of γ.

However, these early models are distinguished by a very sudden switch between regimes. Teräsvirta (1994)) proposes another extension in order to avoid this problem with the Smooth Transition AR (STAR). His solution consists to replace the indicator function by a continuous function $G(y_{t-d};\gamma,c)$ which vary smoothly and monotically from zero to one. This transition function depends on three variables. The first is the transition variable y_{t-d} ; the two others are parameters which determined the localization (c) and the smoothness (γ). Along all possible functions, the most commonly-used are the logistic function (LSTAR model) :

$$G(y_{t-d};\gamma,c) = (1+\exp(-\gamma\prod_{i=1}^{n}(y_{t-1}-c_i)))^{-1},\ c_1 < \ldots < c_n,\ \gamma > 0 \tag{2.8}$$

and the exponential function (ESTAR model) :

$$G(y_{t-d};\gamma,c) = 1-\exp(-\gamma\prod_{i=1}^{n}(y_{t-1}-c_i)^2),\ c_1 < \ldots < c_n,\ \gamma > 0 \tag{2.9}$$

Figure 2.1 shows this two functions for various smoothness values. The main advantage of the STAR model is the effect of the smoothness parameter. It provides a strong transition when $\gamma \to +\infty$ and the STAR can be seens as a TAR. Consequently, when $\gamma \to 0$, the transition is very smooth and when $\gamma = 0$, the STAR performed as a linear model (see van Dijk, Teräsvirta, and Franses (2000) for a complete survey, or Franses and van Dijk (2004)).
These methodologies, first developed for autoregressive processes, were quickly applied to GARCH models. Tong's approach has been taken over by Glosten, Jagannathan, and Runkle (1993) for the GJR-GARCH model, in which the conditional variance follows two different processes depending on the sign of error terms. In the same spirit, the model TGARCH developed by Rabemananjara and Zakoian (1993)) specifies various regimes for the dynamics of the conditional standard deviation depending on the sign of error terms. Finally, Li and Li (1996) develop the Double-Threshold ARCH (DTARCH), which imposes a threshold structure for conditional variance and also for the conditional mean. Following the example of the latter model, Koutmos (1998) proposes a dual-threshold model in order to test the asymmetrical response of mean and conditional variance with

past information (asAR-TGARCH model). The STAR approach also results in various solutions capable of taking into account the asymmetry. Hagerud (1996) makes two STAR-GARCH models. The first type is LSTAR where the transition variable is based on the sign of residuals. The second model is based on a smoothing ESTAR and the transition variable is the absolute value of residuals. These two models are complemented by the analysis of Gonzales-Riviera (1998) with take as transition variable the asymmetry parameter of the variance. Lundbergh and Teräsvirta (1998) build the STAR-STGARCH model in which the mean follows a STAR process and the conditional variance follows a STAR-GARCH. This approach extends two other nonlinear approaches that are GJR-GARCH and Quadratic GARCH of Sentana (1995). Nam, Pyun, and Arize (2002) apply a non-linear structure based on the STAR approach to the mean and conditional variance to capture the asymmetry (ANST-GARCH model). There is also the approach of Lanne and Saikkonen (2005), in which the transition variable is the conditional variance and whose goal is to model series with high persistence in the conditional variance. Finally, Medeiros and Veiga (2004) establish a general formulation with the Flexible Coefficient Smooth Transition GARCH model (FC-STGARCH), encompassing most of the models mentioned above.

STCC model The construction of a STAR structure for the process followed by the correlations of a multivariate GARCH model was produced by Silvennoinen and Teräsvirta (2005) and Berben and Jansen (2005), leading to the Smooth Transition Conditional Correlation GARCH model (STCC-GARCH). The idea is to establish a conditional correlation matrix R_t as a combination of two extremes correlation matrix R_1 and R_2 constants over time. The transition from one to another is provided by the logistic function as defined by the equation (2.8). Specifically, the STCC specifies dynamic of the correlations as :

$$R_t = (1 - G_t)R_1 + G_t R_2 \text{ with } G_t = (1 + e^{-\gamma(m_t - c)})^{-1} \tag{2.10}$$

where γ is the smoothing parameter, c is setting the localization and m_t is the transition variable. To ensure the semi-positivity of the conditional matrix R_t, the matrix R_1 and R_2 must themselves be positive definite, via a Cholesky decomposition for example. The transition variable m_t can be both deterministic or stochastic, but its choice will remain crucial. If there is no real consensus to choice this variable, the study of empirical analysis using the STCC model shows two types of transition variable. It can be :

- a endogenous or exogenous variable. In Silvennoinen and Teräsvirta (2005), the authors apply the model STCC to study the correlations of assets from the daily S&P500. The transition variable is the value of the lagged returns of the squared S&P500. Their strategy is builded on the idea that a strong difference value of the index at a given time represents a strong turbulence of the assets of the S&P500 to the next period.
- calendar time, i.e. $m_t = t/T$. This specification results in the terminology of Silvennoinen and Teräsvirta (2005) to the Time-Varying Smooth Transition Conditional Correlation (TVSTCC) model. That's the choice of Berben and Jansen (2005) in order to study correlations of equity returns.

Silvennoinen and Teräsvirta establish a strategy of inference for choosing the best transition variable. The idea is to find the most explicative variable controlling the dynamics of correlations. In this context, the transition variable varies depending on the purpose of the analysis. In the case where preliminary analysis did not reach to discriminate clearly this transition variable, the choice of calendar time seems to be the most consensual.

Silvennoinen and Teräsvirta (2007) extended the STCC model by introducing two transitions functions to evolve correlations between four correlation matrix constants over time. The conditional

correlations process then takes the following form :

$$R_t = (1-G_{1t})R_{(1)t} + G_{1t}R_{(2)t} \text{ avec } \begin{cases} R_{(i)t} = (1-G_{2t})R_{(i1)} + G_{2t}R_{(i2)} \\ G_{it} = (1+\exp(-\gamma(m_{it}-c_i)))^{-1} \end{cases}, i = 1,2 \tag{2.11}$$

As for the STCC model, the transition variable can be of various kinds (endogenous or exogenous). We now focus on a special case of the DSTCC, in which transition functions are $m_{1t} = m_{2t} = t/T$. In this version, the correlations dynamic can be rewritten in a more illustrative way as :

$$R_t = (1-G_{2t})\underbrace{((1-G_{1t})R_{11} + G_{1t}R_{12})}_{R_{1t}} + G_{2t}\underbrace{((1-G_{1t})R_{21} + G_{1t}R_{22})}_{R_{2t}} \tag{2.12}$$

In this formulation, the correlation matrix is a linear combination of two conditional matrix R_{1t} and R_{2t}. These two matrix are themselves a linear combination of two extreme matrix constants over time and semi-positive definite. Estimation is done by maximizing the likelihood with the usual iterative optimization methods (i.e. Gradient), in one or two steps in accordance with the sample size (see Engle and Sheppard (2001)). However, as often in the regime switching models, the existence of numerous local maxima require to test many initial conditions.

2.2.2 Markov-Switching approach

Theorical aspects The Markov Switching approach has its origins in the construction of mixtures models[2]. The objective of this models is to estimate a density assuming that the density is a finite mixture of densities, i.e. a convex combination of densities. In this model, a latent indicator variable is use to attribute an observation to a component. The Markov-switching approach enhances the mixture model by assuming that the indicator follows a hidden Markov chain, which allows us to introduce a temporal dimension on the indicator components. Thus, each observed variable ε_t has a specific relation with a hidden state s_t which follow a first order Markov chain. The switch from one state to another is specified by transition matrix. This linking relation is generally a first order Markov chain, that means the state s_t is independant of state lagged state except s_{t-1} :

$$\mathbb{P}[s_{t+1} = i | s_t = j] = p_{ij} \tag{2.13}$$

These probabilities are grouped together to make the transition matrix of size $N \times N$, which can be expressed as :

$$\mathrm{P} = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \tag{2.14}$$

with $\sum_{j=1}^{N} a_{ij} = 1$ for each $i = 1,...,N$. In its more general representation, any Markov-switching model can be written as :

$$\varepsilon_t | \mathscr{F}_{t-1} \sim f(\varepsilon_t; \theta) \tag{2.15}$$

with $\mathscr{F}_{t-1}$ denotes the information set at time $t-1$, θ the parameter set. The density function $f(\varepsilon_t; \theta)$ given past observations over N states of the Markov chain can then be expressed as :

$$f(\varepsilon_t; \theta) = \sum_{n=1}^{N} f(\varepsilon_t | s_t = n; \theta_n)\mathbb{P}[s_t = n; \theta_n] \tag{2.16}$$

[2]See, for example, McLachlan and Peel (2000) for a briefing and Böhning, Seidel, Alfo, Garel, Patilea, and Walther (2007) for a survey of recent developments.

with θ_n the parameters of each regime n for $n = 1, ..., N$.
Partly introduced by Lindgren (1978) in economics and highly popularized by Hamilton (1989), the Markov-Switching approach gave rise to a huge literature. We can quote, without being exhaustive, the work of Hamilton (1990) for the development of a filter to estimate the transition probabilities, as well as Krolzig (1997) for a comprehensive extension of the original model to a multivariate framework. Kim (1994) also introduces a smoothing filter refining the calculation of posteriors probabilities.
The first ARCH model with Markov-switching (SWARCH) is proposed by Hamilton and Susmel (1994) while the first GARCH extension (GRS) appears in an analysis of the interest rates by Gray (1996). The innovation for the model developed by Gray lies in writing the conditional variance. Indeed, for SWARCH, each conditional variance depends not only on the regime, but on all history[3]. Whereas in the GRS, the conditional variance depends only on the current regime by the expectation of the lagged variance. Klaassen (2002) improves the latter model by broadening the spectrum of information for the lagged variance. Haas, Mittnik, and Paollela (2004b) suggest a slightly different approach with the MSG(N) model, which consists of estimating N GARCH models in parallel and switching directly from a variance to another. They then avoid the problem of path-dependency. It generalizes several Markov-switching GARCH approaches like the models proposed by Haas, Mittnik, and Paollela (2004a) and by Franq, Roussignil, and Zakoian (2001), which appear as special cases. Finally, Liu (2006) deepens the MSG(N) model by studying the conditions of his steady states with less restrictive assumptions than Haas, Mittnik, and Paollela (2004b).

RSDC model The literature on Markov-Switching GARCH models in a multivariate framework is actually much less dense than in the univariate context. Lee and Yoder (2007)) apply the Gray's method to build a bivariate Markov-switching BEKK model. The multivariate GARCH model of Pelletier (2006) is the first, and for now the only one, to adopt a Markov-Switching structure for the correlations process. By imposing constants correlations in each regime and establishing the switch of each other through a Markov chain of order one, this model falls between the CCC and the DCC. Formerly, the RSDC model assumes that the conditional correlation matrix has the form :

$$R_t = \sum_{n=1}^{N} \mathbb{1}_{\{s_n=i\}} R_n \tag{2.17}$$

where $\{s_t\}_{t\in\mathbb{N}}$ is a sequence follows a homogeneous first order Markov chain with N state. R_n is a conditional correlation matrix of size $K \times K$ where $R_n \neq R_{n'}$ for $n \neq n'$. The distribution of the process defined by $\{s_t\}_{t\in\mathbb{N}}$ is determined by a transition matrix written A. So that the conditional correlation matrix is indeed a correlation matrix[4], it is necessary to impose a PSD constraint on the matrix corresponding to the regimes. This can be a constraint Choleski decomposition, i.e. $R_n = C_n C_n'$ with C_n a lower triangular matrix and imposing that diagonal elements of C_n to be written as :

$$c_{j,j} = \sqrt{1 - \sum_{i=1}^{j-1} c_{j,i}^2} \tag{2.18}$$

Pelletier also proposes a more parsimonious specification of the RSDC whith the following expression :

$$R_t = \bar{Q}\lambda_{s_t} + \mathbf{I}_K(1 - \lambda_{s_t}) \tag{2.19}$$

[3] It is the well-known path-dependency problem.
[4] Recall that by definition, for R_t be is a correlation matrix, it must be symmetrical, PSD, have its diagonal elements equal to 1 and the non-diagonal elements must belong to the interval $[-1;1]$.

With two states, the latter expression as being equal to the weighted average of two extreme states. The case $\lambda_{s_t} = 0$ is where the returns are not correlated while $\lambda_{s_t} = 1$ corresponds to the case where they are very correlated. This second formulation requires an identification constraint as follows :

$$\lambda_1 = 1 \text{ with } \lambda_1 > \lambda_2 > ... > \lambda_N \text{ or also } \max_{i \neq j} |R_{ij}| = 1 \text{ with } 1 > \lambda_1 > \lambda_2 > ... > \lambda_N \tag{2.20}$$

Estimation of the RSDC model is done via maximum likelihood, with a nuance according to the specification in question. For the first specification, corresponding to the equation (3.13), Pelletier favors the use of the EM algorithm. The second specification, define by equation (3.14), can be estimated through Gradient methods by exploiting the idea of correlation targeting of Engle and Mezrich (1996) and using Hamilton's filter. In that case, the correlations can be simply calculated as :

$$\mathbb{E}[R_t] = \bar{Q} \sum_{n=1}^{N} \lambda_n \pi_n + I_K \sum_{n=1}^{N} (1 - \lambda_n) \pi_n \tag{2.21}$$

2.3 Hierarchical Hidden Markov Structure for Dynamic Correlations

As we have seen in section 3.2.1, the DSTCC model of Silvennoinen and Teräsvirta (2007) is an expanded version of the STCC. Its feature is to express the conditional correlation matrix as a combination of two extreme matrix that are themselves dynamics. As the authors explain, at the beginning of the sample, the correlations defined by the equation (2.12) will vary between the two states $R11$ and $R12$, and then towards the end of the sample between R_{21} and R_{22}. Couples of matrix (R_{11}, R_{12}) and (R_{21}, R_{22}) define, according to the second transition variable, two major states bounded by two extreme dynamic matrix. We can say that the dynamic of the DSTCC works on two *primaries* regimes, themselves build under four *secondaries* regimes. The Markov-Switching approach of Pelletier (2006) doesn't consider the case of sub-regimes, unless, in our case, to take four states. However, a classical Markov-Switching model will not reach to capture the existence of sub-regimes. The idea of the model that we present is that, unlike the basic Markov-Switching, it makes possible to establish a hierarchy between states that define sub-regimes. The aim is to show nuances induced by the existence of sub-regimes in the dynamic of the correlations. As a first step, we will present the class of hierarchical hidden Markov models, the structure which is build our model. In a second step, we will expose our model.

2.3.1 Structure

The Hierarchical Hidden Markov Model

The Hierarchical Hidden Markov Model (HHMM) has been proposed by Fine, Singer, and Tishby (1998) in order to generalize the HMM model. The idea is to build a stochastic process with several levels by adopting a tree structure to obtain an interlacing of regimes. The hierarchy of the tree is built with *internal states*, which are abstract states (i.e. they do not produce observations). The internal states can lead so-called *emitting states*, which produce observations. Internal and emitting states can also lead to a third type of state which are called *exiting states*. The exiting states allow to quit a level of the tree. Each internal state produce a sub-HMM, which can also be itself a HHMM. In this framework, the classical HMM is a special case of the HHMM with only one level. The main advantage of the HHMM compared to the HMM is to improve the *granularity* of the regimes. The HHMM approach permits to break up a time series into several types of regimes. For example, a HHMM with two levels has two types of regimes : the firsts are called *primaries regimes* and the seconds *secondaries regimes*. The combination of the secondaries permits

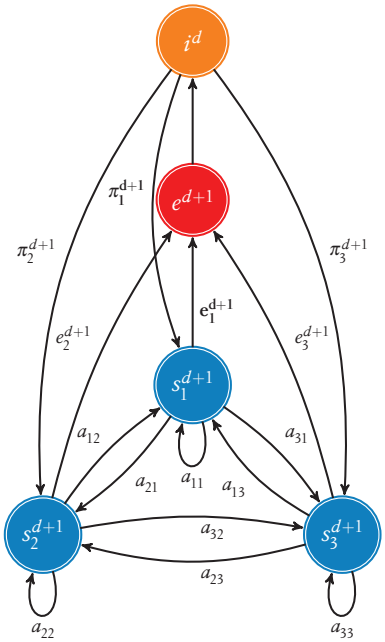


FIG. 2.2: Basic structure of a HHMM with three emitting states.

to deduce the primaries ones. This increased granularity allows us to bring out nuances in regimes which can not occur with the simplified structure of HMM.

Formally, a HHMM can be represented as the process $\{\mathscr{Y}_t, \mathscr{Q}_t\}_{t\in\mathbb{N}}$ with :

- $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$ is the process followed by the observations, which are supposed to be conditionally independent to the hidden states.
- $\{\mathscr{Q}_t\}_{t\in\mathbb{N}}$ is a homogeneous first order Markov chain. Each state of a HHMM q_i^d belongs to the set $\mathscr{Q} = \{\mathscr{S}, \mathscr{I}, \mathscr{E}\}$ where $\mathscr{S}$ is the set of emitting states, $\mathscr{I}$ the set of internal states and $\mathscr{E}$ the set of exiting states. The superscript d corresponds to the index Hierarchy (vertical location) in the tree, with $d \in \{1, ..., \mathscr{D}\}$ and subscript i is its horizontal location.

The tree structure is obtained by imposing an internal state at the root (level $d = 1$)[5]. This initial state then have several descendants which can be internal or emitting states. These sub-internal states can themselves have descendants which can be internal or emitting and so on. The transition from level d to level $d+1$ is provided by the probabilities of vertical transitions through an internal state. The return from level $d+1$ to level d is done with the exiting states and corresponds to a probability of exiting. Internal and emitting states of the same sub-HMM of level d communicate with a transition matrix as in a classical HMM model. An internal state leading to another internal/emitting states of a lowest level is called *parent* state. A parent state leads to *child* states. Finally, three probabilities govern the dynamic of the hidden structure :

- $\mathscr{A}_k^d = (a_k^d(i,j))$: is the matrix of horizontal transition of state i to state j where i and j are two states of the same sub-HMM of level d, i.e. $a_k^d(i,j) = \mathbb{P}[q_{t+1}^d = j | q_t^d = i]$, $q_{t+1}^d, q_t^d \in \{\mathscr{S}, \mathscr{I}\}$.
- $e_i^d = \mathbb{P}[q_{t+1}^d | q_t^{d+1}]$, with $q_t^{d+1}, q_{t+1}^d \in \{\mathscr{S}, \mathscr{I}\}$: vertical probability to leads from a child at level $d+1$ state to parent state at level d.
- $\pi_i^d = \mathbb{P}[q_{t+1}^{d+1} | q_t^d]$ with $q_t^{d+1}, q_{t+1}^d \in \{\mathscr{S}, \mathscr{I}\}$: vertical probability to lead from a parent state to its child state.

[5] We will see later that this condition is not always necessary. The HRSDC model has not an internal state at the summit.

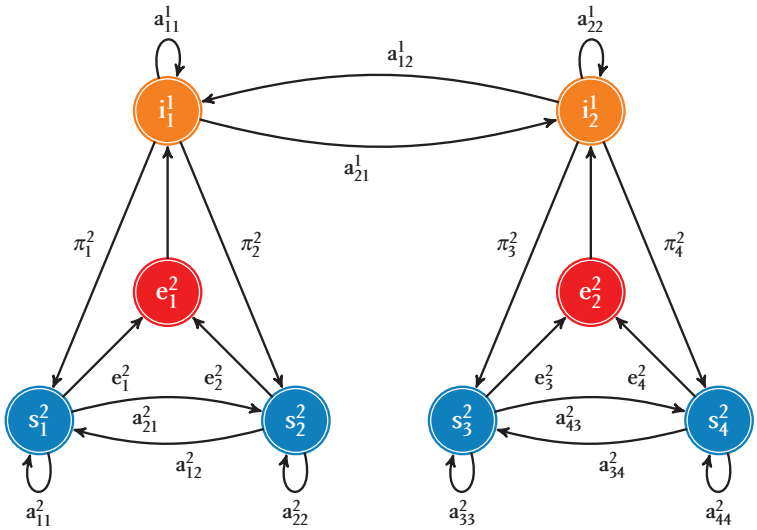


FIG. 2.3: Hierarchical Hidden structure of the HRSDC.

To ensure the transition matrix to be stochastic[6], it is necessary to impose that, for each sub-model depending of parent k :

$$\sum_{j \in ch(k)} a_k^d(i,j) + e_i^d = 1 \text{ and } \sum_{i \in ch(k)} \pi_i^d = 1 \tag{2.22}$$

where $i, j \in ch(k)$ are two states with parents k. Figure 2.2 is an illustration of the basic structure of a HHMM and figure A.1 is a complete example (see appendix A.1). As we can seen, the number of parameters of the model depends on the structure of the tree. Xie (2005) uses a useful notation to represent parameters space of a HHMM. Let Q the size of the space state of each sub-model, Xie assume that global configuration of a HHMM starting from the root up to the d^{th} level can be written as :

$$q^{(d)} = \overline{(q^1 q^2 ... q^d)} = \sum_{i=1}^{d} q^i Q^{d-i} \tag{2.23}$$

Thus, assuming that there is only one state at the root, all the parameters comprising a HHMM matches to :

$$\theta = (\cup_{d=2}^{D} \cup_{i=0}^{Q^{d-1}-1} \{\mathscr{A}_k^d, \pi_i^d, e_i^d\}) \cup (\cup_{i=0}^{Q^{D-1}} \{\phi_i\}) \tag{2.24}$$

where ϕ_i corresponds to the parameters of a parametric probability distribution.

The Hierarchical RSDC model

The objective of the model we are building is to vary the correlations between two extreme major regimes, while allowing the existence of secondaries regimes. Like Silvennoinen and Teräsvirta (2007), the correlation process is bounded by four states of constant correlations over time. The structure highlights two primaries regimes, depending on abstract states i_1^1 and i_2^2. Each of these abstract states is connected with emitting states. Thus, the regime corresponding to i_1^1 is determinated by the emitting states s_1^2 and s_2^2 ; that of i_2^1 by s_3^2 et s_4^2. Figure 2.3 shows the hierarchical hidden structure of the Hierarchical RSDC model (HRSDC).

[6] $\sum_j a_{ij} = 1$ et $0 \leq a_{ij} \leq 1$.

The hierarchical structure allows states to increase the granularity of the regimes. It establishes different types of regimes, which in our case are primaries and secondaries. The primaries regimes correspond to the regimes obtained with a classical Markov Switching model. To a higher level of granularity, these primaries regimes are built with sub-regimes, known as secondaries regimes. The structure allows secondaries's to capture nuances of dynamics that are thinner than the primaries's. The idea of granularity is illustrated on the figure 2.4. Our model has two levels and four secondaries regimes, corresponding to two sub-HMM models.

The pair of emitting states defined by (s_1^2, s_2^2) forms a Markov-Switching model and the same is true for (s_3^2, s_4^2). The link between these sub-models is provided by the abstract states i_1^1 and i_2^1. The model is then building on two sub-models with two emitting states each, which transition matrix are respectively[7] :

$$A_1^2 = \begin{bmatrix} a_{11}^2 & a_{12}^2 \\ a_{21}^2 & a_{22}^2 \end{bmatrix} \text{ and } A_2^2 = \begin{bmatrix} a_{33}^2 & a_{34}^2 \\ a_{43}^2 & a_{44}^2 \end{bmatrix}$$

and verified constraints :

$$\begin{cases} a_{11}^2 + a_{21}^2 + e_1^2 = 1 \\ a_{12}^2 + a_{22}^2 + e_2^2 = 1 \end{cases} \text{ and } \begin{cases} a_{33}^2 + a_{43}^2 + e_3^2 = 1 \\ a_{34}^2 + a_{44}^2 + e_4^2 = 1 \end{cases}$$

where e_i^2, $i = 1, ..., 4$ is the probability of exiting from a state of level two and go to a parent state at level one. The two sub-HMM communicate via exiting states through abstract states i_1^1 and i_2^1. The dynamic in the transition from one to another of these abstract states is defined by the transition matrix :

$$A^1 = \begin{bmatrix} a_{11}^1 & a_{21}^1 \\ a_{12}^1 & a_{22}^1 \end{bmatrix}$$

which verifies :

$$a_{11}^1 + a_{12}^1 = 1 \text{ and } a_{21}^1 + a_{22}^1 = 1$$

Parameters π_i^2, $i = 1, ..., 4$ represents the probability to move from a parent state of first level to one of its children at the second level. These probabilities must verify :

$$\pi_1^2 + \pi_2^2 = 1 \text{ and } \pi_3^2 + \pi_4^2 = 1$$

The specification for the four correlation matrix constants in time is that outlined by Pelletier (see equations 3.13 and 3.14)[8]. In fact, the only difference with the RSDC is the hierarchical hidden structure which introduce us to see the RSDC as a special case of the HRSDC with only one level. As in the RSDC model, the specification defined by equation 3.13 can be estimate by EM algorithm whereas formulation defined by equation 3.14 allows to use iterative methods like Gradient.

2.3.2 Estimation

Estimation of the HRSDC model can be done using the multi-step estimation of Engle and Sheppard (2001) and Engle (2002). This computationally attractive method splits up the log-likelihood $\ell(\varphi, \theta)$ as the sum of two parts : the volatility component $\ell_v(\varphi)$ and the correlations term $\ell_c(\varphi, \theta)$,

[7] Recall that the probability that the state q which was in i at time $t-1$ to be in j at time t is written $\mathbb{P}[q_t = j | q_{t-1} = i] = p_{ji}$.

[8] The four correlation matrix then represent $4K(K+1)/2$ parameters.

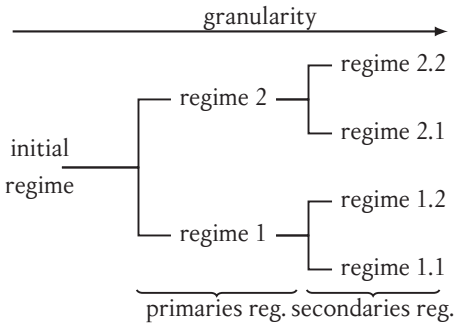


FIG. 2.4: The increase of the granularity in the HRSDC model.

where φ is the parameters space of the K univariate GARCH model and θ corresponding of the parameters of the correlations. More formally, this is written as :

$$\ell(\varphi,\theta) = \ell_v(\varphi) + \ell_c(\varphi,\theta)$$

with :

- $\ell_v(\varphi|\mathbf{r}_t) = -\frac{1}{2}\sum_{t=1}^{T}(K\log(2\pi) + 2\log(|D_t|) + \mathbf{r}_t' D_t^{-2}\mathbf{r}_t)$
- $\ell_c(\theta|\mathbf{r}_t,\varphi) = -\frac{1}{2}\sum_{t=1}^{T}(\log(|R_t|) + \varepsilon_t' R_t^{-1}\varepsilon_t)$

Estimation of the volatility part ℓ_v is done by maximizing the sum of the individual GARCH likelihoods. Estimation of the correlation part is trickier because of the abstract states. Various solutions have been proposed.
In their seminal article, Fine, Singer, and Tishby (1998) used a *generalized Baum-Welch* algorithm based on an modified version of the *Inside-Outside* algorithm. This method requires to calculate all sequences that could be generated by each sub-model of each level of the hierarchy. With a computational complexity of $\mathcal{O}(NT^3)$, where N represents the number of hidden states and T the number of observations, this approach is suitable for small sequences but remains impractical for long sequences[9]. An alternative approach, first developed in Murphy and Paskin (2001) -and taking up again but with some modifications by Hung, Phung, and Venkatesh (2004) - provides an estimate by $\mathcal{O}(T)$. The idea is to consider the HHMM as a special case of Dynamic Bayesian Network (DBN). Note that a mechanic of a DBN is to repeat a chain of Bayesian network for each observation. The chains of the network are connected in time by a causal relationship which permit to track the evolution of the process. After the HHMM converted into a DBN, the estimation uses an classical Baum-Welch algorithm. However, if this is attractive from a computational point of view, it is still relatively complicated to implement. A third approach, suggested by Wierstra (2004), is to circumvent the problem posed by vertical transitions, which implies the existence of several paths to go from one state to another, by transforming the HHMM into a equivalent HMM. This *flattening* method, has the advantage of allowing to use the usual estimation methods of HMM with a complexity of $\mathcal{O}(T)$. Wierstra's tip is to redistribute vertical and horizontal probabilities such that no longer exists self-referential probabilities for internal states. This transformation allows to have a flat version of the HHMM in which there is only one path from a state to another. Xie (2005) succeeds to estimate the HHMM with HMM's standard tools while respecting the vertical dynamic of the model. In contrast to previous approaches, Xie's is builded on a particular expression of the transition matrix. Instead of considering the whole dynamic of

[9]The inability of the Fine, Singer, and Tishby (1998)'s algorithm to deal with long sequence is not presented as a problem by the authors they apply their model for unsupervised handwriting recognition. In such applications, the size of the sequences are about fifty characters.

the model, the transition matrix is broken down into several sub-transition matrix for each hierarchical level. Each level of the tree is then linked with a transition matrix. This rewriting of the transition matrix allows to estimate the correlations with a version of the Baum-Welch algorithm very similar to that used in the context of standard HMM. The complexity of Xie's approach is of order $\mathcal{O}(DTN^{2N})$.
To estimate our HRSDC model, we follows a modified Xie's approach. The main advantage of its method is the speed of implementation. Xie builds the transition matrix by successive layers in order to have an expression of the likelihood of complete data that can be factorizable. In this framework, vertical transitions, which allows to activate the child of a parent state are given by :

$$\pi_q = \prod_{d=1}^{D} \pi_{q^d}^d, \ q = 0, ..., Q^D - 1 \tag{2.25}$$

Transition matrix is rewritten by layers, each layer represents a level of the hierarchy of the tree :

$$\tilde{a}^d(q', q) = \prod_{i=d}^{D} e_{q'^i}^i \pi_{q^i}^i \cdot a(q'^d, q^d) \tag{2.26}$$

This probabilities are called *hypertransition probabilities*. Aggregation of these probabilities leads to the *hypertransition matrix* of our model :

$$\tilde{A}(q', q) = \sum_{i=1}^{D} \tilde{a}^i(q', q) \tag{2.27}$$

Finally, the hypertransition matrix has size $(S \times S)$ where $S = card(\mathcal{S})$. This formulation is attractive in the sense that it leads to work with a transition matrix similar to the classical HMM or Markov-Switching model. The hypertransition matrix of our HRSDC model is then written as :

$$\tilde{A} = \begin{bmatrix} a_{11}^2 + e_1^2\pi_1^2 a_{11}^1 & a_{21}^2 + e_2^2\pi_1^2 a_{11}^1 & e_3^2\pi_1^2 a_{12}^1 & e_4^2\pi_1^2 a_{12}^1 \\ a_{21}^2 + e_1^2\pi_2^2 a_{11}^1 & a_{22}^2 + e_2^2\pi_2^2 a_{11}^1 & e_3^2\pi_2^2 a_{12}^1 & e_4^2\pi_2^2 a_{12}^1 \\ e_1^2\pi_3^2 a_{21}^1 & e_2^2\pi_3^2 a_{21}^1 & a_{33}^2 + e_3^2\pi_3^2 a_{22}^1 & a_{43}^2 + e_4^2\pi_3^2 a_{22}^1 \\ e_1^2\pi_4^2 a_{21}^1 & e_2^2\pi_4^2 a_{21}^1 & a_{34}^2 + e_3^2\pi_4^2 a_{22}^1 & a_{44}^2 + e_4^2\pi_4^2 a_{22}^1 \end{bmatrix} \tag{2.28}$$

The hidden hierarchical structure requires twenty parameters. However, in practice, only sixteen parameters will be to estimate due to constraints of stochastic matrix. In the next sub-section, we propose two ways to estimate the correlations. The first is based on the EM algorithm. The second is done with the Hamilton's filter.

Estimation by EM algorithm

To run EM algorithm[10], we need to write the quantity $\mathcal{Q}(\theta|\theta_k)$. With our hypertransition matrix, it is simply written as :

$$\begin{aligned} \mathcal{Q}(\theta|\theta_p) &= \mathbb{E}_{\theta_p}[\log f(\varepsilon_{1:T}, s_{1:T}; \theta)|\varepsilon_{1:T}] \\ &= \sum_{t=1}^{T}\sum_{i=1}^{S} \log(f(\varepsilon_t; \phi_i))\mathbb{P}_{\theta_p}[s_t = i|\varepsilon_{1:T}] \\ &\quad + \sum_{t=1}^{T-1}\sum_{d=1}^{D}\sum_{i=1}^{S}\sum_{j=1}^{S} \log(\tilde{a}_{ij}^d)\mathbb{P}_{\theta_p}[s_{t+1} = j, s_t = i|\varepsilon_{1:T}] \\ &\quad + \sum_{i=1}^{S} \log(\nu_i)\mathbb{P}_{\theta_p}[s_1 = i|\varepsilon_{1:T}] \end{aligned} \tag{2.29}$$

[10] See appendix A.2 for a very brief review.

To simplify the later formula 2.29, we define :

$$\begin{aligned}\gamma_t^p(q) &\overset{def}{=} \mathbb{P}_{\theta_p}[q_t = q|\varepsilon_{1:T}] \\ \xi_t^p(q',q,d) &\overset{def}{=} \mathbb{P}_{\theta_p}[q_t = q', q_{t+1} = q, e_t = d|\varepsilon_{1:T}]\end{aligned} \tag{2.30}$$

Reestimation formulas are obtained by maximizing the expected value of the complete-data log-likelihood $\mathscr{Q}(\theta|\theta_p)$. In order to speed up this maximization, we can rewriting $\gamma_t^p(q)$ and $\xi_t^p(q',q,d)$ with the auxiliary variables defined by the Forward-Backward algorithm.

Forward-Backward algorithm The Forward-Backward algorithm is a useful method to reduce the calculus complexity[11]. It is to define two auxiliary quantities from which will be expressed reestimation formulas of the M step.

- Forward algorithm : the auxiliary quantity Forward, written $\alpha_t(q)$, is defined as the probability that the sequence $(\varepsilon_1,...,\varepsilon_t)$ to be in state q in the model θ :

$$\alpha_t(q) \overset{def}{=} \mathbb{P}_\theta[\varepsilon_{1:t}, q_t = q] \tag{2.31}$$

  Let $f^p(\varepsilon_t;\phi_q)$ the value of the likelihood at time t and iteration p. To simplify the notations, we will omit the superscript p and use the short-hand notation of $f_q(y_t)$. The two steps calculus for Forward quantity is as follow :

$$\begin{aligned}&① \text{ Initialization : } \alpha_1(q) = \pi_q f_q(\varepsilon_1) \\ &② \text{ Iteration : } \alpha_{t+1}(q) = f_q(\varepsilon_{t+1})\textstyle\sum_{q'}\sum_d \alpha_t(q_t)\tilde{a}^d(q',q)\end{aligned} \tag{2.32}$$

  for $t = 1,...,T-1$ and $q = 0,...,Q^D-1$.
- Backward Algorithm : the Backward variable, written $\beta_t(q)$, is the inverse of the Forward. It is defined as the probability to generate the sequence $(\varepsilon_{t+1},...,\varepsilon_T)$ by leaving from state q in the model θ :

$$\beta_t(q) \overset{def}{=} \mathbb{P}_\theta[\varepsilon_{t+1:T}|q_t = q] \tag{2.33}$$

  We have then two steps to compute the Backward variable :

$$\begin{aligned}&① \text{ Initialization : } \beta_T(q) = 1 \\ &② \text{ Iteration : } \beta_t(q) = \textstyle\sum_{q_t}\sum_d \beta_{t+1}(q')f_{q'}(\varepsilon_{t+1})\tilde{a}^d(q,q')\end{aligned} \tag{2.34}$$

  for $t = 1,...,T-1$ and $q = 0,...,Q^D-1$.

Auxiliary variables $\alpha_t(q)$ and $\beta_t(q)$ allows to rewriting the probabilities $\gamma_t^p(q)$ and $\xi_t^p(q',q,d)$ as follow[12] :

$$\begin{aligned}\gamma_t^p(q) &= \alpha_t(k)\beta_t(k) \\ \xi_t^p(q',q,d) &= \alpha_t(q')\tilde{a}^d(q',q)f_{q'}(\varepsilon_{t+1})\beta_{t+1}(q)\end{aligned} \tag{2.35}$$

As these two variables are probabilities[13], they shall normalize as :

$$\begin{aligned}&\bullet\ \textstyle\sum_k \gamma_t^p(q) = 1 \\ &\bullet\ \textstyle\sum_k\sum_{k'}\sum_d \xi_t^p(q',q,d) = 1 \\ &\bullet\ \gamma_t^p(k') = \textstyle\sum_k\sum_d \xi_t^p(k',k,d)\end{aligned} \tag{2.36}$$

Then, we can easily rewriting the reestimation formulas with $\gamma_t^p(q)$ and $\xi_t^p(q',q,d)$.

[11] it lets you go of a complexity from $\mathscr{O}(e^T)$ to $\mathscr{O}(T)$ (see (Cappé, Moulines, and Ryden, 2005, chap. 3.2) and also (Baum and T., 1966, p. 170).

[12] For a detailed calculation, see (Cappé, Moulines, and Ryden, 2005, chap. 3).

[13] In practice, the calculation of auxiliary variables Forward and Backward revealed problems of *underflow*. It is therefore advisable to apply a scaling factor in order to facilitate the computation (see appendix A.3).

Reestimation formulas Reestimation formulas can be obtained by maximizing $\mathscr{Q}(\theta|\theta_p)$ subject to the stochastic constraints on the parameters (see sub-section 2.3). Then, we have only to differentiate with respect to the constraints and adding the appropriate Lagrange factors. We take the Xie's notations to indexing the states. The writing $q = \overline{(rir')}$ and $q' = \overline{(rir'')}$ are two states configuration which are identical up to the d^{th} level and have the same $(d-1)^{th}$ parent r such that $r = q^{1:d-1} = q'^{1:d-1}$. Moreover, $r' = q_t^{d+1:D}$ and $r'' = q_{t+1}^{d+1:D}$ are two state configurations of a level below d. For each level $d \in \{1,...,\mathscr{D}\}$, the reestimation variables at iteration $p+1$ can be displayed as :

$$\begin{aligned}
&\bullet\ \hat{\pi}_r^d(j) = \frac{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_i \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)}{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_j\sum_i \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)} \\
&\bullet\ \hat{e}_r^d(i) = \frac{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_{q'}\sum_{d'\leq d} \xi_t^p(\overline{(rir')},q',d')}{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''} \gamma_t^p((rir'))} \\
&\bullet\ \hat{a}_r^d(i,j) = \frac{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''} \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)}{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_j \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)} \times (1-\hat{e}_r^d(i)) \\
&\bullet\ \hat{R}_q = \frac{\sum_{t=1}^{T} \varepsilon_t \cdot \gamma_t^p(q) \cdot \varepsilon_t'}{\sum_{t=1}^{T} \gamma_t^p(q)}
\end{aligned} \tag{2.37}$$

We simply use as stopping rule the conventional difference $\|\hat{\theta}^{p+1} - \hat{\theta}^p\| \leq 10^{-6}$.

Estimation Gradient methods

Using Gradient methods to estimate a HMM is made possible with the so-called Hamilton's filter (see Hamilton (1989, 1990, 1994)). This iterative filter allows to make inference on the state of the unobserved Markov chain. With Hamilton's notations, let $\hat{\xi}_{t|t}$ a vector of size $(N \times 1)$ which elements $\xi_{jt} = \Pr[s_t = j|\mathscr{F}_{t-1},\theta]$, $j = 1,...,N$ are the conditional probabilities to be in regime j given the information set $\mathscr{F}_{t-1}$ at time $t-1$ and $\tilde{A}$ our hypertransition matrix of size $(N \times N)$ (see equation 2.27). Then, the vector η_t of size $(N \times 1)$, which elements are the densities under the N regimes, i.e. $f(\varepsilon_t|s_t = j,\mathscr{F}_{t-1};\theta)$, $j = 1,...,N$. Then, Hamilton shows that the observed density ε_t given the information set is written as :

$$f(\varepsilon_t|\mathscr{F}_{t-1};\theta) = \imath'(\hat{\xi}_{t|t-1} \odot \eta_t) \tag{2.38}$$

where $\imath'$ is a $(N \times 1)$ vector with all elements equal to 1. The conditional distribution of s_t can be deduced :

$$\begin{aligned}
\frac{f(\varepsilon_t, s_t = j|\mathscr{F}_{t-1};\theta)}{f(\varepsilon_t|\mathscr{F}_{t-1};\theta)} &= \mathbb{P}_\theta[s_t = j|\varepsilon_t,\mathscr{F}_{t-1}] \\
&= \mathbb{P}_\theta[s_t = j|\mathscr{F}_t]
\end{aligned} \tag{2.39}$$

Combining with equation 2.38, relation 2.39 has the following expression :

$$\mathbb{P}_\theta[s_t = j|\mathscr{F}_t] = \frac{f(\varepsilon_t, s_t = j|\mathscr{F}_{t-1};\theta)}{\imath'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{2.40}$$

As the joint density $f(\varepsilon_t, s_t = j|\mathscr{F}_{t-1};\theta)$ is written as :

$$f(\varepsilon_t, s_t = j|\mathscr{F}_{t-1};\theta) = \mathbb{P}_\theta[s_t = j|\mathscr{F}_{t-1}] \times f(\varepsilon_t|s_t = j,\mathscr{F}_{t-1};\theta) \tag{2.41}$$

which is the density of the j^{th} regime. By collecting for the N regimes, the filtered probabilities become :

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{\imath'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{2.42}$$

with forecasts calculated by :

$$\hat{\xi}_{t|t+1} = \tilde{A} \times \hat{\xi}_{t|t} \tag{2.43}$$

Inference for each date t is then found by iterating equations 3.31 and 2.43. For choosing starting value, we use one of the options proposed by (Hamilton, 1994, p. 693). He suggests to take the vector of unconditional probabilities π, which is the solution of the following system :

$$\begin{cases} \tilde{A}\pi = \pi \\ \imath'\pi = 1 \end{cases} \tag{2.44}$$

The starting value are then given by :

$$\pi = (M'M)^{-1} M e_{N+1} \text{with } M = \begin{bmatrix} \mathbf{I}_N - \tilde{A} \\ \imath' \end{bmatrix} \tag{2.45}$$

where e_{N+1} is the $(N+1)^{th}$ column of $\mathbf{I}_{N+1}$.

Smoothed probabilities

Hamilton's filter allows to make inference about the state of the Markov chain at time t conditional on the information set up to time t. Kim (1994) has developed an filter in order to make inference with the whole information set. Instead of computing $\xi_{t|t}$, it permit to compute $\xi_{t|T}$ with $t < T$. The Kim's filter is computed with our hypertransition as in the classical case :

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \{\tilde{A}'[\hat{\xi}_{t+1|T} \oslash \hat{\xi}_{t+1|t}]\} \tag{2.46}$$

where $\oslash$ denotes element-by-element division.

2.4 Applications

This section contains Monte-Carlo experiments and two applications based on real dada. The first real database is that used by Colacito and Engle (2006) and contains daily data from S&P500 and 10-year bond futures (see appendix A.5). The second real database application is done with that of Pelletier (2006). It contains four week-days close exchange rates against US dollar which are Pound, Deutschmark, Yen and Swiss-Franc (see appendix A.4).

2.4.1 Simulated data

In this sub-section, we compare the correlations estimates of our HRSDC model and the DCC_{ES} of Engle and Sheppard (2001) in a setting where the true correlation structure is known. For simplicity, it is done in a bivariate framework. We simulate six DGP. The first one has the following variance equations :

$$\begin{aligned} h_{1,t} &= 0.01 + 0.05 r_{1,t-1} + 0.85 h_{1,t-1} \\ h_{1,t} &= 0.12 + 0.1 r_{1,t-1} + 0.85 h_{1,t-1} \end{aligned}$$

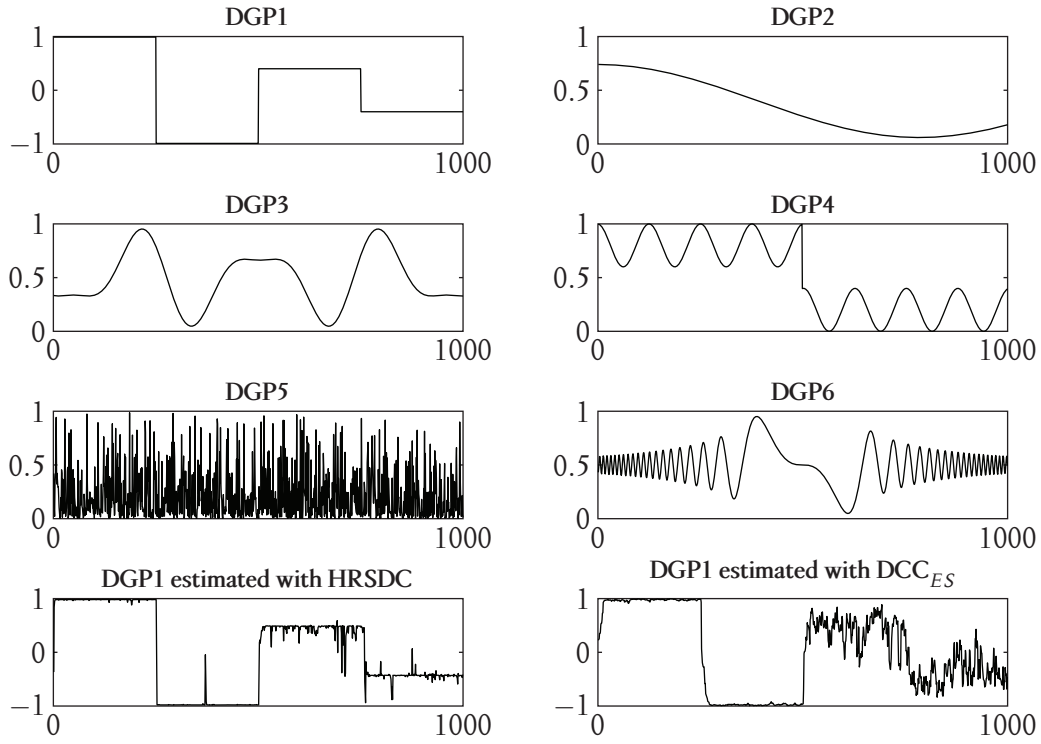


FIG. 2.5: Simulated Correlation Processes. The last two graphs are an example of the estimated correlations of the two models for the DGP1.

and the others (DGP2 :5) :

$$h_{1,t} = 0.01 + 0.04 r_{1,t-1} + 0.95 h_{1,t-1}$$
$$h_{1,t} = 0.01 + 0.2 r_{1,t-1} + 0.5 h_{1,t-1}$$

For the DGP1, the correlations process follow a TAR-CCC model with constants correlations. The DGP2 :5 are builded with dynamic processes. The different DGP are labeled as :

- DGP1 : $R_{12,t} = \begin{cases} 0.99 \text{ if } t \in [1;250] \\ -0.99 \text{ if } t \in [251;500] \\ 0.4 \text{ if } t \in [501;750] \\ -0.4 \text{ if } t \in [751;1000] \end{cases}$
- DGP2 : $R_{12,t} = 0.4 + 0.34\cos(t/250)$
- DGP3 : $R_{12,t} = 0.5 + \cos(d/s) - (1/3)\cos(3d) + (1/7)\cos(5d)$, $d = (t-50)/145$, $s = 35$
- DGP4 : $R_{12,t} = \begin{cases} .8 + .2\cos(t/20) \text{ if } t \in [1;500] \\ 0.2 + .2\cos(t/20) \text{ if } t \in [501;1000] \end{cases}$
- DGP5 : $R_{12,t} = 0.99 - \frac{1.98}{1+\exp(0.5\max(\varepsilon^2_{1,t-1},\varepsilon^2_{2,t-1}))}$
- DGP6 : $R_{12,t} = 0.5 + \sin(s^3)/(1+\sqrt{|s^3|})$, $s = 5 - t/100$

The simulated correlations are voluntarily pathological in order to test the accuracy of the correlations estimates corresponding of a very volatile/distress periods of financial markets. The DGP2 :4 ;6 are builded with sinusoidal functions to create correlations with different regimes or sub-regimes. The DGP5 is used by Long and Ullah (2005) and corresponds of the stylized fact pointed out by Longin and Longin and Solnik (2001) that correlations among assets tend to increase during volatile periods.

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
MAE_1						
DCC	.1331	.5125	.3774	.4237	.5298	.2764
HRSDC	.0562	.5521	.3836	.4642	.5386	.2831
MAE_2						
DCC	.0553	.2901	.1700	.2361	.2824	.0778
HRSDC	.0235	.3408	.1660	.2659	.3008	.0835
DQ 5% *(strategy EW)*						
DCC	.7594	.5884	.6499	.9299	.5784	.9586
HRSDC	.5966	.6985	.7814	.9699	.4263	.9800
DQ 1% *(strategy EW)*						
DCC	.3359	.9944	.9970	$1e^{-4}$	.0217	.9925
HRSDC	.4382	.9733	.9995	$2e^{-4}$	.0215	.9929
DQ 5% *(strategy LS)*						
DCC	.7969	.1616	.5136	.9532	.7284	.7068
HRSDC	.8731	.0912	.3509	.9216	.6574	.7321
DQ 1% *(strategy LS)*						
DCC	.4294	$4.3e^{-5}$	.9986	.9996	.01	.0523
HRSDC	.3896	.2685	.9959	.9994	.01	.0136

TAB. 2.1: Performance measures results.

The performance measures we use are very similar of theses used by Engle (2002). We first calculate two versions of a very classical loss function, which are computed as follow :

$$MAE_1 = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{R}_t - R_t\right|$$
$$MAE_2 = \frac{1}{T}\sum_{t=1}^{T}(\hat{R}_t - R_t)^2$$

For the second type of measure, we follow the methodology of Engle (2002) by considering the loss function of the Value-at-Risk (VaR). Recall that for a portfolio with a share w invested in the in the first asset and $(1-w)$ in the second, the VaR assuming normality can be computed as :

$$VaR_t^{\alpha} = \Phi_t^{-1}(\alpha)\sqrt{(w^2\hat{H}_{11,t} + (1-w)^2\hat{H}_{22,t} + 2w(1-w)\hat{R}_{12,t}\sqrt{\hat{H}_{11,t}\hat{H}_{22,t}})}$$

The loss function of the VaR is then defined by :

$$hit_t = \mathbb{1}_{\{wr_{1,t}+(1-w)r_{2,t}<-VaR_t\}} - \alpha$$

We then use the *in-sample Dynamic Quantile* (DQ) test introduced by Engle and Manganelli (2004). This consists to test if all the coefficients of the regression of the violation process hit_t with its lagged values and others exogenous variables are equals to zero. To compute this F test, we use as explanatory variables five lags hit_t and the current value of the VaR. We test two numbers of rejections (1% and 5%) and two portfolios : an equal-weighted ($w = 0.5$, strategy EW) and a long-short (1 and -1, strategy LS). Table 2.1 presents the results of the performance measures.
The results show that the HRSDC performs better than the DCC when correlations are piecewise constants. This result is consistent with what could be expected since the HRSDC varied correlations between several constants correlations matrix over time. Once the simulated correlations are no longer piecewise constants, the DCC model has better MAE, which, once again, seems normal. However, differences in the MAE between the DCC and the HRSDC are rather weak. The

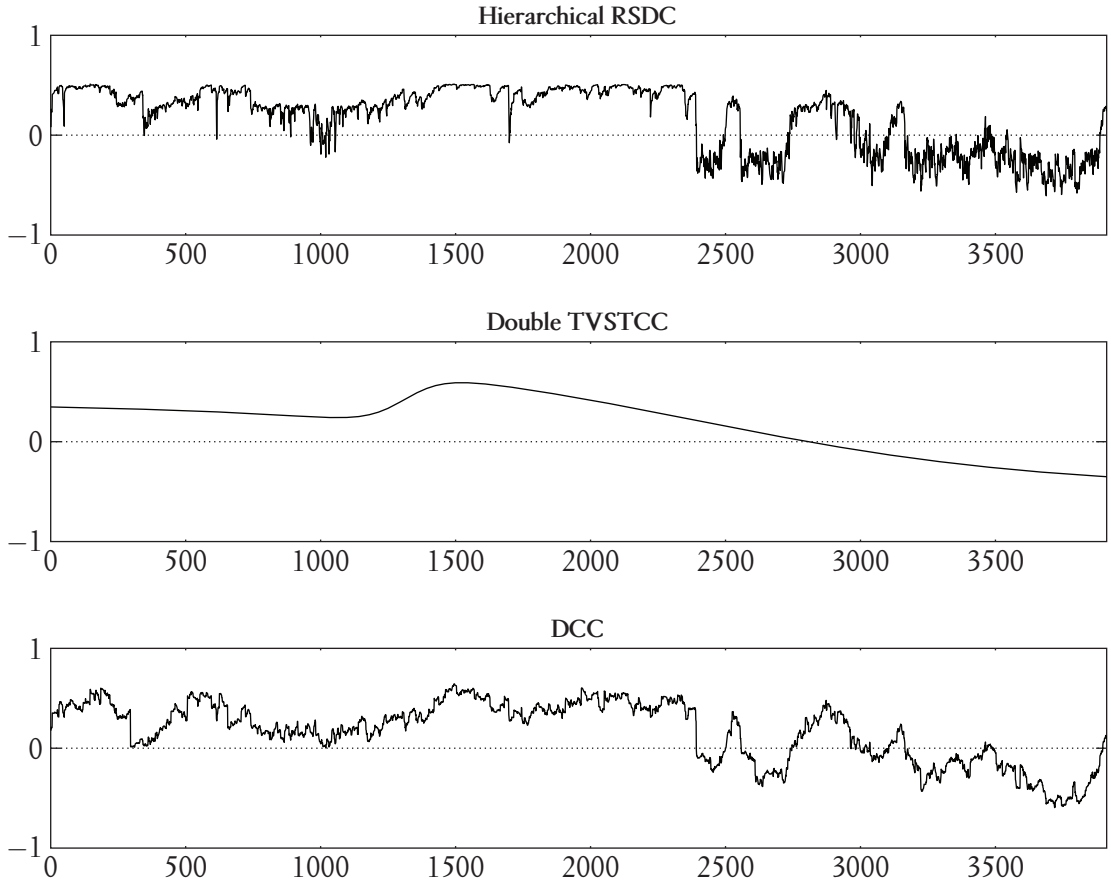


FIG. 2.6: Estimated correlations of the HRSDC, DCC_{ES} and DSTCC with the sample of Colacito and Engle (2006).

dynamic process resulting from a combination of regimes of the HRSDC seems to be not too bad face to the autoregressive dynamic of the DCC.
The results of the DQ test are more mixed and appear highly dependents on the strategy chosen. The DCC is the best for EW strategy, while HRSDC dominates on the LS strategy. A simple addition of the best values shows that the HRSDC seems preferable to DCC. Nevertheless, it would be unfair to conclude to the dominance of HRSDC under the DCC. It is just better to point out that the HRSDC remains credible face to the DCC.
Finally, the HRSDC remains a formidable efficiency when correlations are piecewise constants (DGP1), but becomes less effective when it is no longer the case (DGP2 :6). It seems that the autoregressive dynamic of the DCC stay very efficient when correlations oscillate around a trend value. However, the interest of the HRSDC is in its explanatory power. Beyond a simple measure of correlations, its structure allows to bring out the components of the overall dynamic by increasing the granularity of regimes.

2.4.2 Correlations between S&P500 index futures and 10-year bond futures

The first application with a real database is based on the bivariate sample of Colacito and Engle (2006). It contains daily returns of S&P500 index futures and 10-year bond futures from January 1990 to August 2003 (see appendix A.5). This data are also used by Silvennoinen and Teräsvirta (2007) to apply their DSTCC model. The individuals volatilities are obtained by running a

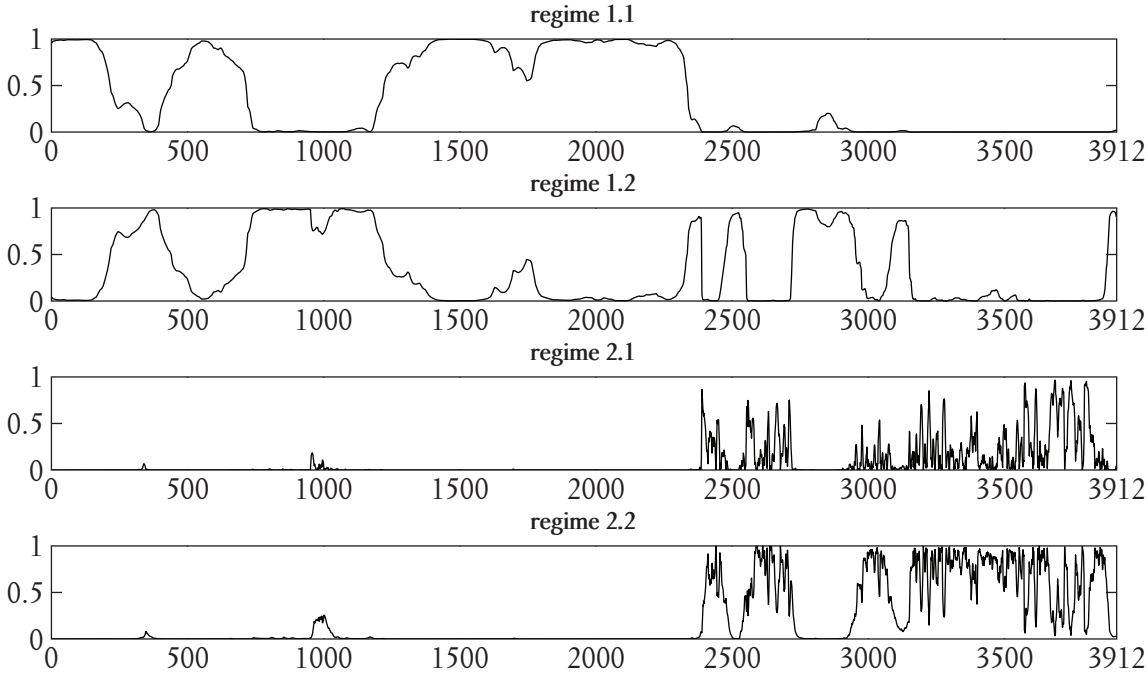


FIG. 2.7: Smoothed probabilities of the HRSDC. Regimes 1.1 and 1.2 correspond to the emitting states $\{s_1^2, s_2^2\}$ and regimes 2.1 and 2.2 to $\{s_3^2, s_4^2\}$.

GARCH(1,1) model. We later obtain the correlations by running three different models in order to make comparisons. These models are HRSDC, DCC_{ES} and DSTCC. For this sample, the correlations of the DCC_{ES} and DSTCC are estimated using gradient methods while the correlations of the HRSDC are computed with EM algorithm. The figure 2.6 shows the estimated correlations for the three models.

In our application, transition variables for the DSTCC are defined as $m_{1t} = m_{2t} = t/T$ (calendar time). In their study, Silvennoinen and Teräsvirta (2007) used another specification : one of the transition variable is calendar time while the other is VIX index[14]. This particular approach is motivated by investigate the fact that correlation among S&P500 and bond futures increase with distress of the market. As our model can not incorporate exogenous variable, calendar time allows to the DSTCC to be similar to the HRSDC. They are together conditional on time.

Results of the estimated parameters of the HRSDC are presented in appendix A.6 and the smoothed probabilities corresponding to the secondary regimes are in the figure 2.7. The model has correctly identified two sub-regimes with positive correlations and two sub-regimes with negative correlations. Adding the smoothed probabilities of each pairs of sub-regimes give the smoothed probabilities that we would have had with Markov-Switching model with two regimes. This is shown on the Figure 2.8. The results for the DSTCC (with the calendar time specification) show that it does not accurately captures the *punctual* changes in regime. Correlations estimate have the form of a curve. Rather than a local measure, this model allows for a trend. For example, in the DCC and HRSDC models, correlations go to a negative correlations regime around the 2400^{th} observation whereas they are still positives in the DSTCC. These is due to the fact that the DSTCC

[14]The Vix index (for Chicago Board Options Exchange volatility index) is build with of selected basket of the implied volatility of S&P500 index options. It gives a measure of expectation of the market's volatility over the next 30 days. The VIX index can be freely downloaded at : `http://www.cboe.com/micro/vix/introduction.aspx`

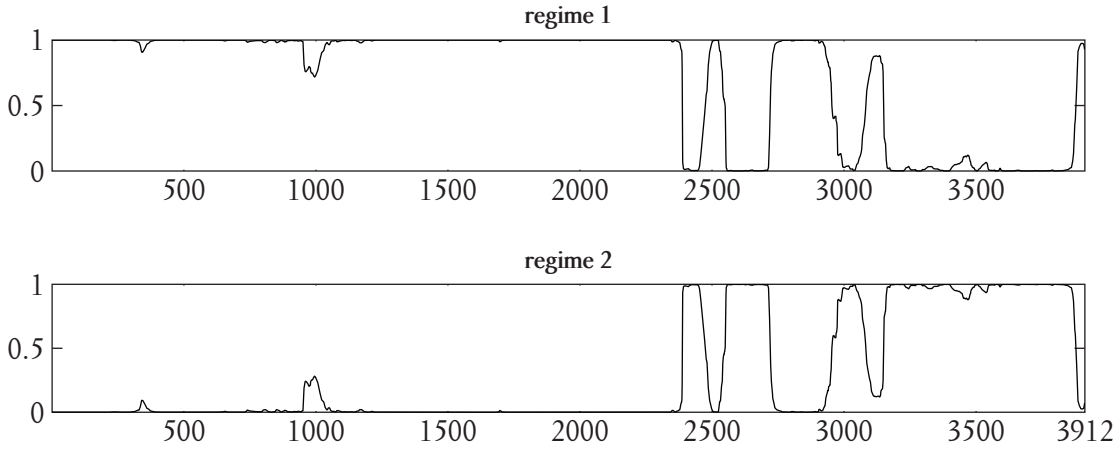


FIG. 2.8: Smooth probabilities of the regimes defined by the internal states i_1^1 and i_2^1 (first application).

has only two transition functions. A solution to this problem could be to introduce more transition functions as in Amado and Teräsvirta (2008). If the DCC and HRSDC correlations estimate seem similar, the advantage of HRSDC in the case study is to explain nuances in the dynamics through the decomposition in sub-regimes. The increasing of the granularity ends up to a finer definition of the dynamic.

2.4.3 Correlations between exchange rate data

In this second application, we apply our HRSDC model to the sample used by Pelletier (2006). This series are exchange rate data and are plotted on the figure A.2 (see also appendix A.4). As before, the dynamic of the standard deviations is obtained with a GARCH(1,1) model. For this case study, only DSTCC and HRSDC models are considered. We use the two-step maximum likelihood estimation using gradient methods for both models. Figure 2.9 shows plots of the estimated correlations for the two models. Parameters estimated for the HRSDC are in the table A.2 (see appendix A.6). Smooth probabilities for the HRSDC can be seen in the figure 2.10. As we can seen on figure 2.9, as in the previous application, the DSTCC (with its calendar time specification, i.e. $m_{1t} = m_{2t} = t/T$) allows mainly to establish a trend on the evolution of correlations in the sample. The reason is often that, with only two transition functions, the DSTCC can not capture more than three regimes. The direct consequence of this formulation is that it fails to capture the punctual extreme regimes, as it shows by example on the Swiss Fr/Deutschmark correlations. These are indeed marked by two peaks of correlations close to zero as they oscillate the rest of the time around a value about 0.8.

On this sample, the HRSDC clearly identifies four sub-regimes. Recall that the first primary regime associated with the internal state i_1^1 is a combination of the secondaries regimes depending of the two emitting states s_1^2 and s_2^2. Increasing the granularity provided by the hidden hierarchical structure allows to highlight the existence of a sub-regime linked to s_2^2. This sub-regime occurs only very rarely along the sample, around the 450^{th} observation. But it permit to capture an extreme and punctual behavior of the correlations process. This element of the global dynamic of correlations would be go unnoticed with a classical Markov-Switching model (because too short and not significant enough in relation to the size of the sample).

This application with a sample of four series brings to light a problem in the correlations specification. As we have said before, estimated parameters have been obtained using Gradient methods to

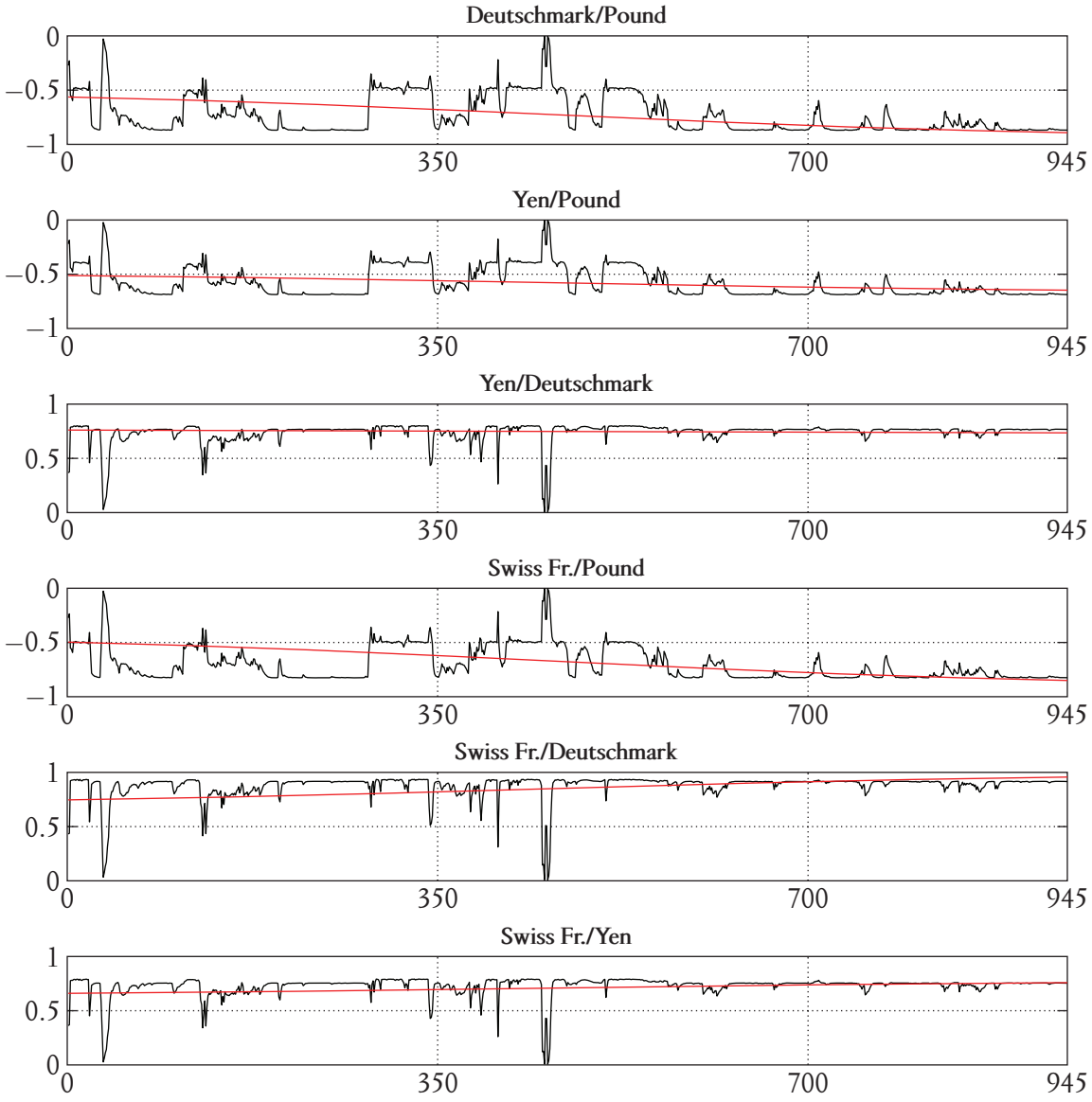


FIG. 2.9: Estimated correlations for the HRSDC (in black) and DSTCC (in red).

estimate the HRSDC. This choice is not fortuitous. In fact, our experiments shows that the specification defined by equation (3.13) has some difficulties to return the maximum of the objective function by EM algorithm. The reason comes from the constraint Choleski representation used for the correlation matrix. Without this transformation, the derivatives of the objective function gets a formulation of the correlation matrix with no guarantees to have all one in the diagonal and PSD. But, the substitute of the Choleski representation for the diagonal elements :

$$c_{j,j} = \sqrt{r_{i,i} - \sum_{i=1}^{j-1} c_{j,i}^2} \tag{2.47}$$

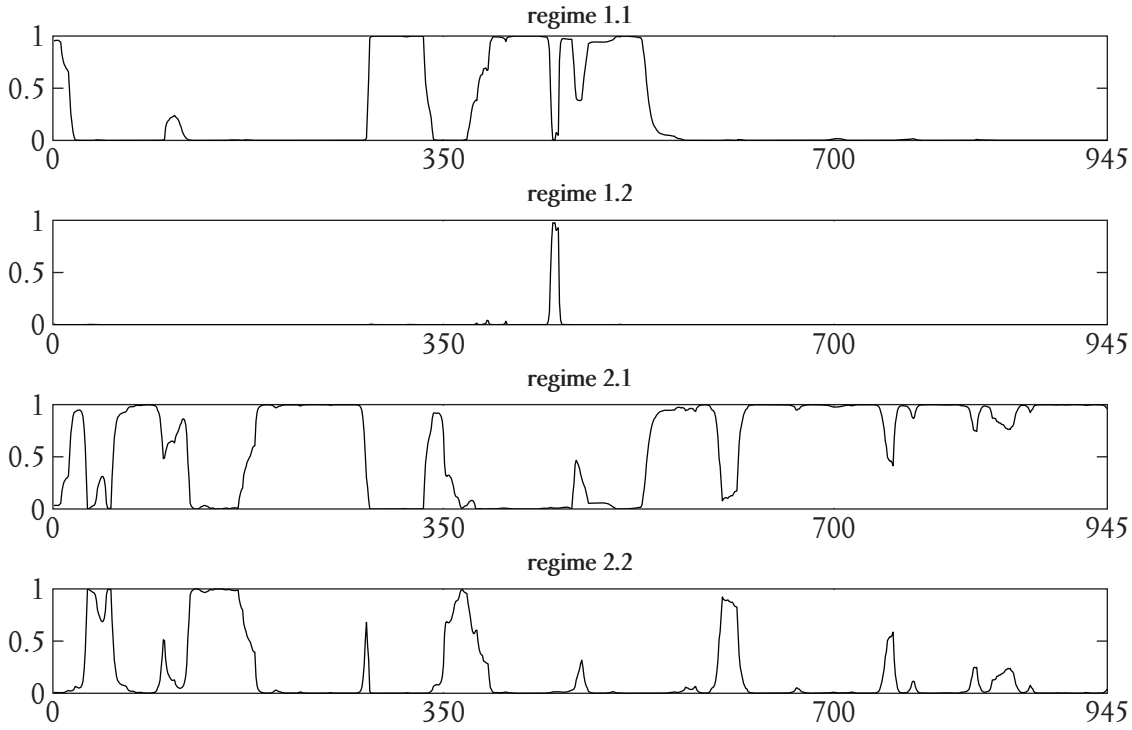


FIG. 2.10: Smoothed probabilities for the HRSDC (exchange rate data).

by its constraint version :

$$c_{j,j} = \sqrt{1 - \sum_{i=1}^{j-1} c_{j,i}^2} \tag{2.48}$$

strongly disrupts the convergence of the EM algorithm. That's why it is preferable to use the specification defined by equation (3.14) which can be estimated by gradient methods using Hamilton's filter. In this example, despite of a huge number of iterations, the EM algorithm has not converge whereas Gradient methods rapidly reach the optimum of the objective function. Nevertheless, even the second specification (equation (3.14)) need numerous parameters and the estimation of large correlation matrix could turn out to be cumbersome.

2.5 Conclusion

In this paper, we have presented a new multivariate GARCH with dynamic correlations. This further extension, called Hierarchical RSDC (HRSDC), can be view as a special case the RSDC model of Pelletier (2006). The HRSDC is a Markov-Switching class model with a correlation process bounded by four correlation matrix constants over time. The innovation of our model is that it is builded on a hierarchical hidden structure introduced by Fine, Singer, and Tishby (1998).
The main advantage of this hidden tree-like structure is to increase the granularity of the regime. It permits to define different types of regime ; in our case, primary and secondary regime. Apply to correlations modeling, the HRSDC allows to capture thinner nuances than with the classical Markov-Switching approach. Monte carlo experiments and applications on real data show that this approach could improve understanding of the dynamic of the correlations. The application of

the HRSDC to estimate the correlations between S&P500 index futures and 10-year bond futures or between exchange rate data has exhibited existence of sub-regime which can't appear with the another regime switching model.

While the results in this paper show that the HRSDC has a good explanatory power, several limitations have yet to be overcome. The first of them is about model selection. In this study, our model is builded from a symmetric hidden tree, with two primaries regimes which are formed with one sub-HMM each. This is a very simple structure. It is possible to build an asymmetric tree with much levels of depth. But finding the best hierarchy, i.e. split/merge/swap levels, is yet an open problem. Xie (2005) use the Reversible-Jump MCMC (RJMCMC) of Green (1995) for model selection. This approach seems interesting but it must be expanding. A second problem is about the specification of the correlation matrix. The specification of Pelletier doesn't suit to model large correlation matrix. As the hierarchical hidden structure model is a plug-in method, it could be interesting to find a specification for large correlation matrix. These tasks await future research.

CHAPITRE 3

A Markov-switching Dynamic Conditional Correlation model with factorial hidden Markov representation.

3.1 Introduction

In the last years, financial econometrics has shown a wide range of new development in the modelling of variances and correlations of financial series. In this new field, the work of Bollerslev (1990), which introduced a multivariate garch with constant conditional correlations (the wellknown CCC), has given to conditional correlations models a huge centre of attention. Since empirical studies (see Longin and Solnik (1995, 1996, 2001) among others) have pointed out the fact of fluctuations in correlation among series according to the degree of volatility of financial market, the focus is actually on dynamic correlations models. Engle and Sheppard (2001) (and also around the same time Tse and Tsui (2002)) proposed a generalization of the CCC by making correlations to be time-varying (the DCC model). This seminal approach has been the starting point of many extensions which allow for richer correlation dynamics while making estimation feasable[1].
Since Hamilton (1989), Markov-switching has been extensively used in econometrics in the field of time series. First applications of Markov-switching to ARCH models came from Cai (1994) and Hamilton and Susmel (1994). Gray (1996) solved the problem of path-dependancy and proposed a MS-GARCH by taking the conditional variance depends only on the current regime by the expectation of the lagged variance. This specification was improved by Klaassen (2002) with a broader information set. Finally, Haas, Mittnik, and Paollela (2004b) generalized the previous approaches by making the GARCH equations corresponding to hidden state evolving in parallel. Extensions to the multivariate case have been proposed by Pelletier (2006), Billio and Caporin (2005) and also Haas and Mittnik (2008)[2].
In this paper, we propose an extension of the Markov-switching version of the DCC based on the approach of Haas and Mittnik (2008). Indeed, one limitation of the previous models is that all elements of the correlation matrix must switch together. In the case of the model of Pelletier (2006) with two states, correlations evolve between two extrem constant correlations. In Billio and Caporin (2005) and also in Haas and Mittnik (2008), correlations are defined by two non-linear processes. When a switch occurs, all elements of the correlations matrix switch together. We argue this condition may be too restrictive. The main innovation of our proposal approach is to allow each correlations to have its own switching dynamic. Thus, elements of the correlations can jump to different times while in the classical models, all elements switch at the same time.
This constraint of a common switch for all the elements of the correlation matrix is relaxed by introducing Markov switch following the factorial representation developed by Ghahramani and Jordan (1997). The factorial representation should not be confused with the factor models usually used in econometrics. The factorial hidden Markov model (FHM) can be view as a generalization of the classical hidden Markov model where hidden states are factored into multiple hidden state variables. Each element of the correlation matrix follows its own Markov chain so that the correlations between the first and the second time series can have a different switching dynamic of correlation between the first and the third time series.
The paper is organized as follows. The model is defined section 3.2. We firts briefly recall the basics of the DCC of Engle and Sheppard (2001) and then outline our model. Section 3.3 the model is applied to data examples. Final remarks concludes and proposes areas for further research in section 3.4.

3.2 The Model

In this section we first define the basic framework of dynamic correlation models and then explicite the FHM approach. Related extension dynamic correlations is then described.

[1] see Bauwens, Laurent, and Rombouts (2006) or Silvennoinen and Teräsvirta (2009) for a complete survey.
[2] These models will be reviewed in the next section.

3.2.1 Conditional Correlations Models

Starting with Bollerslev (1990), the basic framework for conditional correlation models defines the conditional variance-covariance matrix of returns r_t of size $K \times 1$ expressed as follows :

$$H_t = D_t R D_t \tag{3.1}$$

where R is a $K \times K$ constant correlation matrix and D_t a $K \times K$ diagonal matrix containing univariate time varying standard deviations such that :

$$D_t = \text{diag}\{h_{i,t}^{1/2}\},\ i = 1, ..., K \tag{3.2}$$

This elements can be extracted by applying a GARCH(p,q) for each series $i = 1, ..., K$:

$$h_{i,t} = \omega_i + \sum_{j=1}^{q} \alpha_{i,j} r_{i,t-j}^2 + \sum_{j=1}^{p} \beta_{i,j} h_{i,t-j} \tag{3.3}$$

with usuals stationarity and non-negativity restrictions. Another way of understanding this setup is to rewriting equation (4.2.2). Let ε_t the standardized residuals expressed as :

$$\varepsilon_t = D_t^{-1} \mathbf{r}_t \tag{3.4}$$

The conditional correlations are simply the expectation of the standardized residuals :

$$\mathbb{E}_{t-1}[\varepsilon_t \varepsilon_t'] = D_t^{-1} H_t D_t^{-1} = R \tag{3.5}$$

Therefore, despite its computational attractivity, the assumption of constant correlations seems to be too parsimonious and seriously unrealistic in empirical applications (see Longin and Solnik (1995, 1996, 2001) among others). Engle and Sheppard (2001) extented this previous approach by making the correlations to be time-varying :

$$H_t = D_t R_t D_t \tag{3.6}$$

This model (called DCC) first defines a dynamic for the covariance of the residuals from which conditional correlations are computed. The conditional covariance dynamic of the standardized residuals V_t follow a BEKK-like process :

$$Q_t = (1 - a - b)\overline{Q} + a\varepsilon_{t-1}\varepsilon_{t-1}' + bQ_{t-1} \tag{3.7}$$

where $\overline{Q}$ is the unconditional covariance matrix of the standardized residuals and a and b are two positive scalars such that $a + b < 1$. Under this restrictions the matrix V_t is positive definite and the conditional correlation matrix is then obtained by :

$$R_t = \text{diag}\{Q_t\}^{-1/2} Q_t \text{diag}\{Q_t\}^{-1/2} \tag{3.8}$$

One interesting feature in this approach is it involves only two parameters to define cavariances dynamics and a such specification has the advantage to be very parcimonious in the field of multivariate GARCH models. This original parametrization introduced a new field in multivariate variance modelling and dynamic correlations models has been a new centre of attention. We no longer outline these extensions and we refers to Silvennoinen and Teräsvirta (2009) and Bauwens, Laurent, and Rombouts (2006) for an updated review of these approaches.

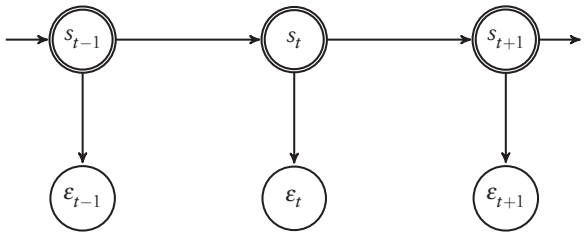


FIG. 3.1: The Markov-switching model views as a dynamic Bayesian networks.

3.2.2 The main approach

Our approach extends the Markov-Switching model to modelling dynamic correlations. We briefly recall the main idea of the Markov-Switching model. This approach assumes that the observed variable ε_t has a specific relation with a hidden state s_t which follow a first order Markov chain. Each observation is linked with a hidden state by a probability and the switch from one state to another is specified by transition matrix. This linked relation is generally a first order Markov chain, that means the state s_t is independant of state lagged state except s_{t-1} :

$$\mathbb{P}[s_{t+1} = i | s_t = j] = p_{ij} \tag{3.9}$$

These probabilities are grouped together to make a $N \times N$ transition matrix :

$$\mathrm{P} = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \tag{3.10}$$

where $\sum_{j=1}^{N} a_{ij} = 1$ for each $i = 1, ..., N$. In its more general representation, any Markov-switching model can be written as :

$$\varepsilon_t | \mathscr{F}_{t-1} \sim f(\varepsilon_t; \theta) \tag{3.11}$$

with $\mathscr{F}_{t-1}$ denotes the information set at time $t-1$, θ the parameter set. The density function $f(\varepsilon_t; \theta)$ given past observations over N states of the Markov chain can then be expressed as :

$$f(\varepsilon_t; \theta) = \sum_{n=1}^{N} f(\varepsilon_t | s_t = n; \theta_n) \mathbb{P}[s_t = n; \theta_n] \tag{3.12}$$

with θ_n the parameters of each regime n for $n = 1, ..., N$. The main idea is to allows the parameters to switch according to value of the hidden Markov chain at each point of time. This can be expressed graphically by a dynamic Bayesian networks (DBN) (see figure 3.1). The idea of introducing Markov-switching in dynamic correlations models has been motivated by severals empirical applications. In particular, Longin and Solnik (2001) has shown that correlations tend to increase when volatility is hight and Ang and Chen (2002) and Campbell, Koedijk, and Kofman (2002) have founded a significant increased correlation in international equity returns in bear markets. In this way, Pelletier (2006) suggested a specification where correlations switch between various constant correlations matrix :

$$R_t = \sum_{n=1}^{N} \mathbb{1}_{\{s_n = i\}} R_n \tag{3.13}$$

where R_n is the constant correlation matrix of regime n and $\{s_t\}_{t \in \mathbb{N}}$ a first order Markov chain. Because correlations envolve between constant correlations matrix, this model can be view as a

hybrid approach between CCC and DCC. But despite its simple specification, this representation involve a high number of parameters. To avoid this problem, Pelletier (2006) proposed a constraint formulation :

$$R_t = \Gamma \lambda_{s_t} + \mathbf{I}_K \left(1 - \lambda_{s_t}\right) \tag{3.14}$$

with $\lambda_{s_t} \in (0,1)$ a regime-dependent weighting variable, Γ a constant correlation matrix and $\mathbf{I}_K$ a $K \times K$ identity matrix. Nevertheless, this parcimonious formulation does not permit the correlations to change sign. In order to bypass these drawbacks, Billio and Caporin (2005) introduced regime switch in equation (4.6) such as :

$$Q_t = \left(1 - a_{s_t} - b_{s_t}\right)\overline{Q}_{s_t} + a_{s_t}\varepsilon_{t-1}\varepsilon'_{t-1} + b_{s_t}Q_{t-1} \tag{3.15}$$

where a_{s_t} and b_{s_t} are two regime-dependant scalars and $\overline{Q}_{s_t}$ an unconditional regime-dependant covariance matrix of the standardized residuals. The restrictions to have a positve definite correlation matrix is to have $a_{s_t} + b_{s_t} < 1$ and $\overline{Q}_{s_t}$ definite positive. Therefore, this approach has the drawback to require approximations to be estimated. This leads to difficulties in the interpretation of the covariance process of each regime.
Haas and Mittnik (2008) develop a multivariate Markov-switching GARCH model based of the assumption of Haas, Mittnik, and Paollela (2004b). This approach is applied on a full BEKK formulation but can easly extented to the DCC as the dynamic of the residuals in the follows a BEKK process. Haas and Mittnik (2008) suggest that the conditional covariance process has K independant covariance processes. It can be written as follow :

$$Q_{t,s_t} = C_{s_t} + A_{s_t}\varepsilon_{t-1}\varepsilon'_{t-1}A'_{s_t} + B_{s_t}Q_{s_t,t-1}B'_{s_t} \tag{3.16}$$

where A_{s_t} and B_{s_t} are symetric matrix and C_{s_t} is a positive definite matrix. This specification does not require approximations as the conditional covariance matrix is written as a function of the current state s_t and the squared standardized residuals. This specification has several advantages. As pointed out by Haas and Mittnik (2008), it seems to be the *most natural* approach for multi-regime GARCH model since the K covariance processes evolve independantly which make easier the identification of low- and high-volatility periods. Estimation using maximum likelihood also renders this model interesting from a practical point of view.

3.2.3 Definition of the Model

Despite its high capabilities of modelling regime switches, we argue in this paper that the classical Markov-Switching can suffer of a lack of flexibility. Indeed, the switch between regime then imposes strong restrictions on the parameters set. In the case of equation (3.16), the switch from one regime to another assumes that all the conditional covariance switch togother from one regime to another. *In fine*, all the covariance are assumed to have the same switching date. That means there is no possibility to differentiate the switches of each covariance of the conditional covariance matrix. Although, this condition may be too rectrictive and it seems interesting to relax it.
From an economic point of view, our model has an easy interpretation. Taking the exemple to study the linkage in the correlations of exchange rates. Using standart DCC, the dynamic of correlations is constant over all the period. This may be strongly unrealistic during turbulent period and a DCC based on the MMS-GARCH stucture of Haas and Mittnik (2008) seems to be a natural specification to modelise a possible change in the correlations. Nevertheless, all the correlations switch at the same time. While it provides an interesting explanatory power and allows to clear identification of crisis/non-crisis period, this assumption may appear as a too strong simplification. To overcome this shortcoming, our proposed model assumes individual switching dynamic conditions for each element of the correlation matrix.

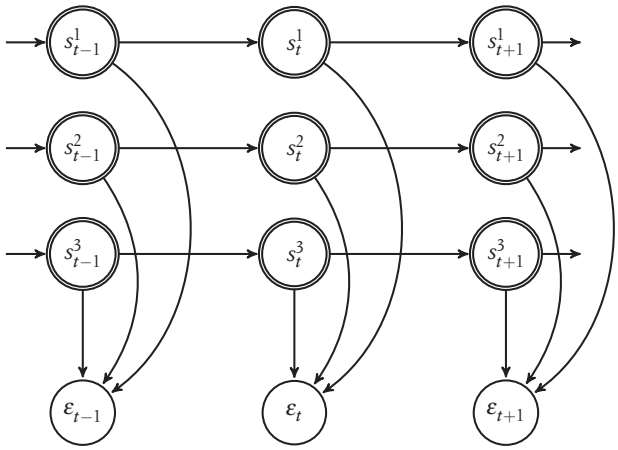


FIG. 3.2: fHMM view as DBN.

Our model is based on the factorial hidden Markov model (FHMM) where each covariance of the conditional covariance matrix has its own dynamic. This extension of the hidden Markov model, developed by Ghahramani and Jordan (1997), generalized the previous Markov-Switching approach by representing the hidden state in a factored form. Each hidden state is factored into multiple hidden state variables evolving in parallel. Each chain has similar dynamics to a basic hidden Markov model. We consider in this paper the chains are independant and make the usuals assumptions of aperiodicity, irreducibility and ergodicity. The difference is the probability to produce an observation at each time depends upon the current state in all of the chains. Representation as dynamic Bayesian networks is shown in figure 3.2. Comparing to equation (3.12), the density function $f(\varepsilon_t;\theta)$ given past observations over the set of M parallel Markov chain $s^1,...,s^M$ with N states each can then be expressed as :

$$\mathbb{P}[s^1_{t+1}=i,...,s^1_{t+1}=i|s^1_t=j,...,s^1_t=j]=\prod_{m=1}^{M}\mathbb{P}[s^m_{t+1}=i|s^m_t=j] \tag{3.17}$$

Since the state space is making as the cross product of these state variables, this representation that can model several processes with their own independent dynamics. It makes possible to build a Markov-switching set up for equation (3.16) where covariances are linked to its own Markov chain so that it allows allow them to switch independently from others. Formally, assuming to simplify notation that :

$$s_t=(s^1_t,...,s^M_t) \tag{3.18}$$

We express the expression of the conditional covariance process of the standardized residuals can be written as :

$$Q_{t,s_t}=C_{s_t}+A_{s_t}\varepsilon_{t-1}\varepsilon'_{t-1}A'_{s_t}+B_{s_t}Q_{t-1,s_t}B'_{s_t} \tag{3.19}$$

Each Markov chain s^i_t is linked to a parameter and has its own transition matrix, respectively $\mathrm{P^i}$ with $i=1,...,M$. For convenience, we will assume that all Markov chains have N states. The resulting state space is then based on N^M combinations.
To avoid an inflation of parameters while maintaining estimation feasible, we choice to use diagonal matrix for A_{s_t} and B_{s_t} such that :

$$A_{s_t}=\mathrm{diag}(a_{s^1_t},a_{s^2_t},...,a_{s^M_t}) \tag{3.20}$$

and similarly :

$$B_{s_t} = \text{diag}(b_{s_t^1}, b_{s_t^2}, ..., b_{s_t^M}) \tag{3.21}$$

In that way, the ij^{th} element of V_{t,s_t} has the following expression :

$$Q_{ij,t,s_t} = c_{ij,s_t} + a_{i,s_t^m} a_{i,s_t^m} \varepsilon_{t-1} \varepsilon'_{t-1} + b_{i,s_t^m} b_{i,s_t^m} Q_{ij,t-1,s_t} \tag{3.22}$$

for $m = 1, ..., M$ and $i, j = 1, ..., K$.
Stationnarity condition yet assumes that Q_{t,s_t} is positive definite and :

$$a_{i,s_t^m} a_{i,s_t^m} + b_{i,s_t^m} b_{i,s_t^m} < 1 \tag{3.23}$$

with $m = 1, ..., M$ and $i = 1, ...K$. It also requires the matrix C_{s_t} to be symetric and positive definite and the initial covariance matrix to be positive definite. Given this specification our model does not suffer of the problem of path dependency. It is build on the same idea of Haas and Mittnik (2008) and can be view as a special case of their approach. Indeed, Haas and Mittnik (2008) assume a covariance matrix Q_t defined by K separate BEKK models, where each element switch together from one regime to another. Whereas our model assumes that each element of the matrix Q_t evolve independently from each other in their switching dynamic.

3.2.4 Estimation via equivalent HMM representation

It is wellknown that the Holy Grail of multivariate GARCH model is to make compatible two contradictory objectives : having a model with a good explanatory power, which requires few parameters while making estimation feasable. This problem has motivated the choice of a scalar BEKK representation for our covariance process. Nevertheless, our specification remains dependent of an order NK because of the parameters needed by the regime-dependent unconditional covariance matrix. Hopefully, our model can be estimated using step-wise procedure of Engle and Sheppard (2001). Assuming gaussian innovations, the first stage involve the estimation of K univariate GARCH and corresponds to the volatility component of the log-likelihood :

$$\ell_v(\varphi|\mathbf{r}_t) = -\frac{1}{2}\sum_{t=1}^{T}(K\log(2\pi) + 2\log(|D_t|) + \mathbf{r}'_t D_t^{-2}\mathbf{r}_t) \tag{3.24}$$

These variances estimates of r_t are used to construct the matrix D_t :

$$D_t = \text{diag}\{h_{i,t}^{1/2}\} \tag{3.25}$$

and allow to compute the standardized residuals :

$$\varepsilon_t = D_t^{-1}\mathbf{r}_t \tag{3.26}$$

In a second step, the correlation component is estimated by maximising the likelihood conditioning on the volatility parameters estimated in the first step :

$$\ell_c(\theta|\mathbf{r}_t, \varphi) = -\frac{1}{2}\sum_{t=1}^{T}(\log(|R_t|) + \varepsilon'_t R_t^{-1}\varepsilon_t) \tag{3.27}$$

Since the Markov chains s^i for $i = 1, ..., M$ are not observed, inference on the state of these Markov chains require to make a filtration step. Assuming each chains can have N distinct values, our model has M transition matrix of size $N \times N$. As pointed out by Ghahramani and Jordan (1997), any factorial hidden Markov representation can be express by a regular hidden Markov model

with a single transition matrix. Indeed, as state is factored into multiple states of parallel chains, the regular Markov-switching representation for the transition is given by the kronecker product of the transition matrix of the parallel chains, which give a transition matrix of size $N^M \times N^M$ (see appendix B.1 for details). Let's call Υ the resulting transition matrix of the regular Markov-switching formulation, then in the case of equation (3.19) we have :

$$\Upsilon = \bigotimes_{i=1}^{M} \mathrm{P}^{\mathrm{i}} \tag{3.28}$$

where $\otimes$ is the kronecker product. Estimation can now be performed used standart tools. Since the elements of the transition matrix Υ are a combination of elements of matrix P^{i}, $i = 1, ..., M$, we will denote them by υ_{ij}, with $i, j = 1, ..., N^M$. Thus, to peformed Hamilton's filter (see Hamilton (1989, 1990, 1994)), we first need to define the probability to be in regime j given the information set :

$$\xi_{jt} = \mathbb{P}[\upsilon_t = j | \mathscr{F}_{t-1}, \theta] \tag{3.29}$$

and the density under the regime j :

$$\eta_{jt} = f(\varepsilon_t | \upsilon_t = j, \mathscr{F}_{t-1}; \theta) \tag{3.30}$$

Inference on the state of the Markov chain is then defined by :

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{\imath'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{3.31}$$

where $\odot$ denotes element by-element multiplication. $\hat{\xi}_{t|t}$ contains the probability to be in each regime at time t given the observations set up to t, which can be expressed as :

$$\hat{\xi}_{t|t+1} = \Upsilon \times \hat{\xi}_{t|t} \tag{3.32}$$

The regular Markov-switching representation of a factorial hidden Markov model makes inference on states easier using Hamilton's filter. However, estimation of multivariate GARCH models stay difficult. Moreover, as point out by Engle (2009), definition and estimation of the intercept C_{s_t} is the most difficult part of our model. In standart DCC, Engle and Sheppard (2001) use *variance targeting*. This means that the intercept is estimated via an auxiliary estimator based on a moment condition. In the case on a Markov-switching representation of DCC, this solution is no longer possible. Hence, the specification of Haas and Mittnik (2008) needs to estimate one intercept matrix per regime. In our case, the the situation is more complicated since we have N^M intercepts to estimate[3].

As point out by Palandri (2009), maximisation of the likelihood can be view of a challenge for optimizer, and this is especially true in regime switching MGARCH. In this paper, we used Gradient methods based on numericals derivatives. This motivated to take a very parcimonious specification to avoid problems of local maxima.

3.2.5 Smoothing

Smoothed probabilities can be compute from the regular Markov-switching representation. These filtered probabilities refer to inferences on the state conditional on all the information set of the sample. Computation is done using the filter of Kim (1994) given by :

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \{\Upsilon'[\hat{\xi}_{t+1|T} \oslash \hat{\xi}_{t+1|t}]\} \tag{3.33}$$

[3] We explain in detail how to specify the intercepts in order to be parcimonious in paramaters.

where $\oslash$ denotes element-by-element division. Nevertheless, filtered probabilities computed with equation (3.33) refer to the regular Markov-switching representation. This means it gives inference of mixed states. In appendix B.2, we propose an easy method to recover smoothed probabilities from those of the regular representation.

3.2.6 Multi-step-ahead correlation matrix forecast

As in the DCC, forecasting the covariance matrix of our model can be decomposed in two separate parts : the forecast of the diagonal matrix D_t on one hand and the forecast of the correlation matrix on the other. The matrix contains the time varying standard variation can be calculated recursively. Using a GARCH(1,1) model, the volatility forecast at time $T-1$ can be computed such that :

$$h_{t+r} = \sum_{i=1}^{r-2} \omega(\alpha+\beta)^i + (\alpha+\beta)^{r-1} h_{t+1} \tag{3.34}$$

where $h_{t+1} = \omega + \alpha r_t^2 + \beta h_t$ is known at the period t. The second part is far from obvious. The first problem arises with the structure with the factorial structure of the chains. An easy way to compute the r-step ahead forecast of the correlation matrix is to adopt the regular MS representation. It gives :

$$\hat{R}_{T,T+r} = \sum_{\tau=1}^{r} \hat{R}_{T,T+\tau} = \sum_{\tau=1}^{r} \sum_{i=1}^{M^N} \mathbb{P}[\upsilon_\tau = i | \mathscr{F}_{T-1}] \hat{R}^i_{T,T+\tau} \tag{3.35}$$

with $\hat{R}_{T,T+r}$ the correlation matrix forecast from T for the next r steps and :

$$\mathbb{P}[\upsilon_{t+\tau} = i | \mathscr{F}_{T-1}] = \Upsilon^\tau \times \mathbb{P}[\upsilon_t = i | \mathscr{F}_{T-1}] \tag{3.36}$$

A second problem occurs for the formulae of covariances forecast. Indeed, the forecasting of the correlation matrix is a combination of the forecasting covariance of the residuals and the square root of residuals. According to Engle and Sheppard (2001), one way to deal this problem is to make an approximation. Assuming that $\overline{V}(s_t^{1,\dots,M}) \approx \overline{R}(s_t^{1,\dots,M})$ and $\mathbb{E}_t[Q_{t+1}(s^{1,\dots,M})] \approx \mathbb{E}_t[R_{t+1}(s^{1,\dots,M})]$, the r-step-ahead correlation matrix forecast is given by :

$$\mathbb{E}_t[R_{t+r}] = \sum_{i=1}^{r-2} \left(1 - a(s^{1,\dots,M}) - b(s^{1,\dots,M})\right) \overline{R}(s^{1,\dots,M}) \left(a(s^1) + b(s^{1,\dots,M})\right)^i + \left(a(s^{1,\dots,M}) + b(s^{1,\dots,M})\right)^{r-1} R_{t+1}(s^{1,\dots,M}) \tag{3.37}$$

with $\overline{R}(s^{1,\dots,M})$ is the average correlations of regime depending of $s^{1,\dots,M}$. Hence, in each regime, the forecast of the conditional correlation matrix will converge to the long run unconditional correlation matrix of the regime.

3.3 Applications

We apply our model an application on a dataset of daily data. This dataset contains three exchange rates (Canadian dollar, Yen and Pound) against the US dollar from march 1999 to july 2009. It contains 2697 daily observations (see figure 3.3). We expose an comparison of three models : standard DCC with diagonal-BEKK dynamic, Markov-switching DCC following the Haas and Mittnik (2008) specification with also a diagonal-BEKK dynamic and our proposed model. All time series were downloaded from DataStream and results are generated using Matlab on Linux.
In this application, returns are calculated by taking 100 times the first difference of the logarithm of each series minus the sample mean. Descriptives statistics of log-returns are given in table 3.0(a).

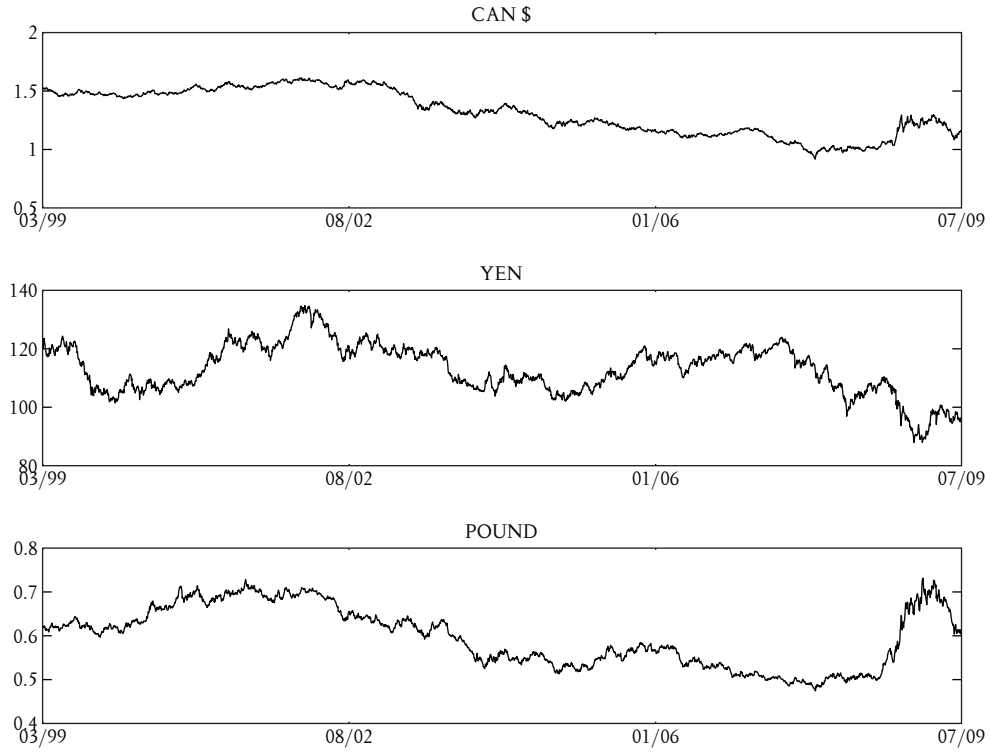


FIG. 3.3: Canadian dollar, Yen and Pound from march 1999 to july 2009.

This series are then filtered with a GARCH(1,1) model to get the standartized residuals. Results are shown in table 3.0(b).

The first model we apply is the DCC with diagonal BEKK dynamic and is written as :

$$Q_t = \overline{P} - \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} \overline{P} \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} - \begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix} \overline{P} \begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix} + \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} \varepsilon_{t-1}\varepsilon'_{t-1} \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix}' + \begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix} Q_{t-1} \begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix}' \tag{3.38}$$

where $\overline{P}$ refers to the unconditional covariance matrix of the standardized residuals. Using variance targeting, $\overline{P}$ can be estimated as :

$$\overline{P} = T^{-1} \sum_{t=1}^{T} \varepsilon_{t-1}\varepsilon'_{t-1} \tag{3.39}$$

This model is very parcimonious and needs only six parameters. Estimated correlations are plotted on figure 3.4.

The second model we compare is the DCC based on the diagonal-BEKK dynamic of Haas and

(a) DESCRIPTIVE STATISTICS FOR THE CANADIAN DOLLAR, YEN AND POUND.

	Can $	Yen	Pound
Minimum	-5.7236	-4.9343	-4.4737
Maximum	3.2253	3.2156	3.9187
Mean	-0.0105	-0.0081	-0.0009
Variance	-0.0105	-0.0081	-0.0009
Skewness	-0.3784	-0.4181	0.0656
Kurtosis	10.6898	6.3327	8.4107
Jarque-Bera	$6.7095e+03$ $(1.0000e-03)$	$1.3267e+03$ $(1.0000e-03)$	$3.2918e+03$ $(1.0000e-03)$
Box-Ljung Q(25)	56.15523 $(0.3465e-03)$	29.2504 (0.2536)	80.8896 $(0.8272e-07)$
Box-Ljung Q(50)	103.2819 $(0.0142e-03)$	65.6647 (0.0678)	131.0487 $(0.0344e-07)$

(b) GARCH(1,1) ESTIMATES.

	Can $		Yen		Pound	
Parameters	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
ω	0.0009	1.821e-07	0.0050	2.84e-06	0.0027	8.53e-07
α	0.0420	3.09e-05	0.0291	2.84e-06	0.0442	4.26e-05
β	0.9559	3.55e-05	0.9596	7.15e-05	0.9477	6.12e-05

(c) DIAG-DCC ESTIMATES.

Parameters	Estimate	Std. error
a_{11}	0.0984	3.2e-4
a_{22}	0.1462	0.0014
a_{33}	0.11030	3.0e-4
b_{11}	0.9938	9.0e-6
b_{22}	0.9869	4.6e-5
b_{33}	0.9937	7.4e-6

(d) MS-DCC.

Parameters	Estimate	standart errors
a_{11}	0.0998	0.0267
a_{22}	0.0712	0.0022
a_{33}	0.0730	0.001
b_{11}	0.9947	4.31e-06
b_{22}	0.9308	0.007
b_{33}	0.9960	2.15e-06
const	$\begin{bmatrix} 5.81e-09 & -7.70e-05 & 2.23e-04 \\ & 0.0085 & 5.12e-04 \\ & & 3.93e-04 \end{bmatrix}$	$\begin{bmatrix} 3.02e-09 & 4.08e-07 & 4.56e-08 \\ & 1.09e-04 & 2.75e-07 \\ & & 2.70e-08 \end{bmatrix}$
a_{11}	0.0120	2.09e-05
a_{22}	0.0300	9.24e-05
a_{33}	0.0676	2.93e-04
b_{11}	0.9896	3.40e-05
b_{22}	0.9902	1.04e-05
b_{33}	0.9914	4.84e-06
const	$\begin{bmatrix} 2.53e-04 & 1.07e-04 & 2.14e-04 \\ & 4.32e-04 & 3.29e-04 \\ & & 0.001 \end{bmatrix}$	$\begin{bmatrix} 1.11e-09 & 1.14e-09 & 3.96e-09 \\ & 3.03e-08 & 7.34e-09 \\ & & 1.16e-06 \end{bmatrix}$
p_{11}	0.9990	3.54e-06
p_{22}	0.9985	9.53e-07

TAB. 3.1: Descriptive statistics, GARCH(1,1) estimates, DCC and MS-DCC estimated parameters.

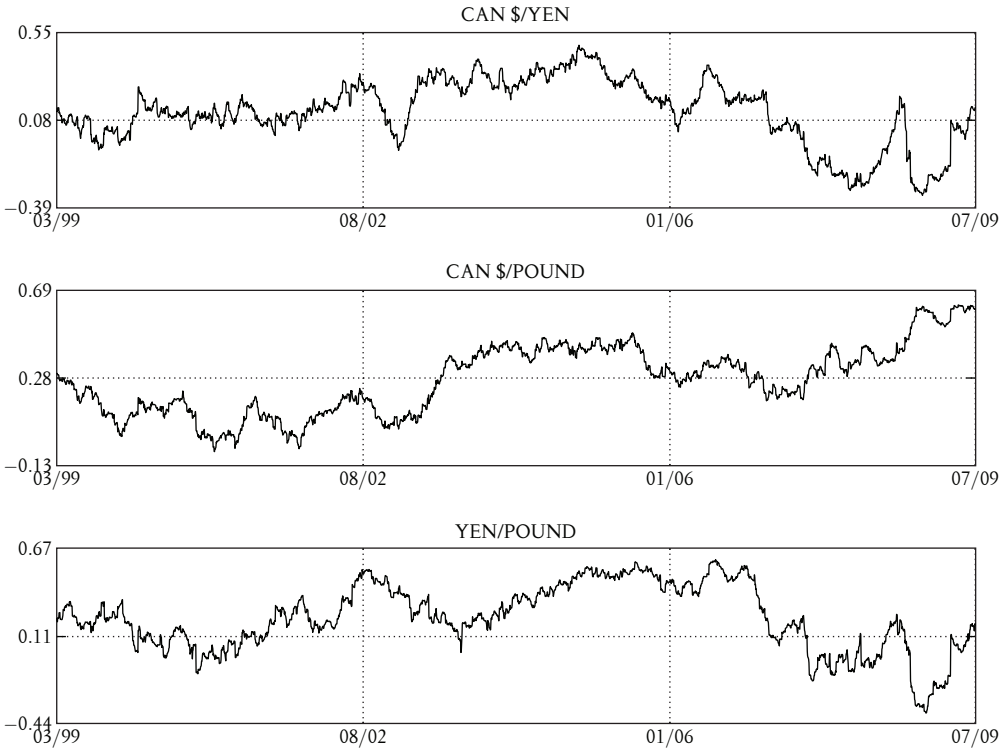


FIG. 3.4: Estimated correlations of diagonal DCC.

Mittnik (2008) with two regimes. It can be expressed as :

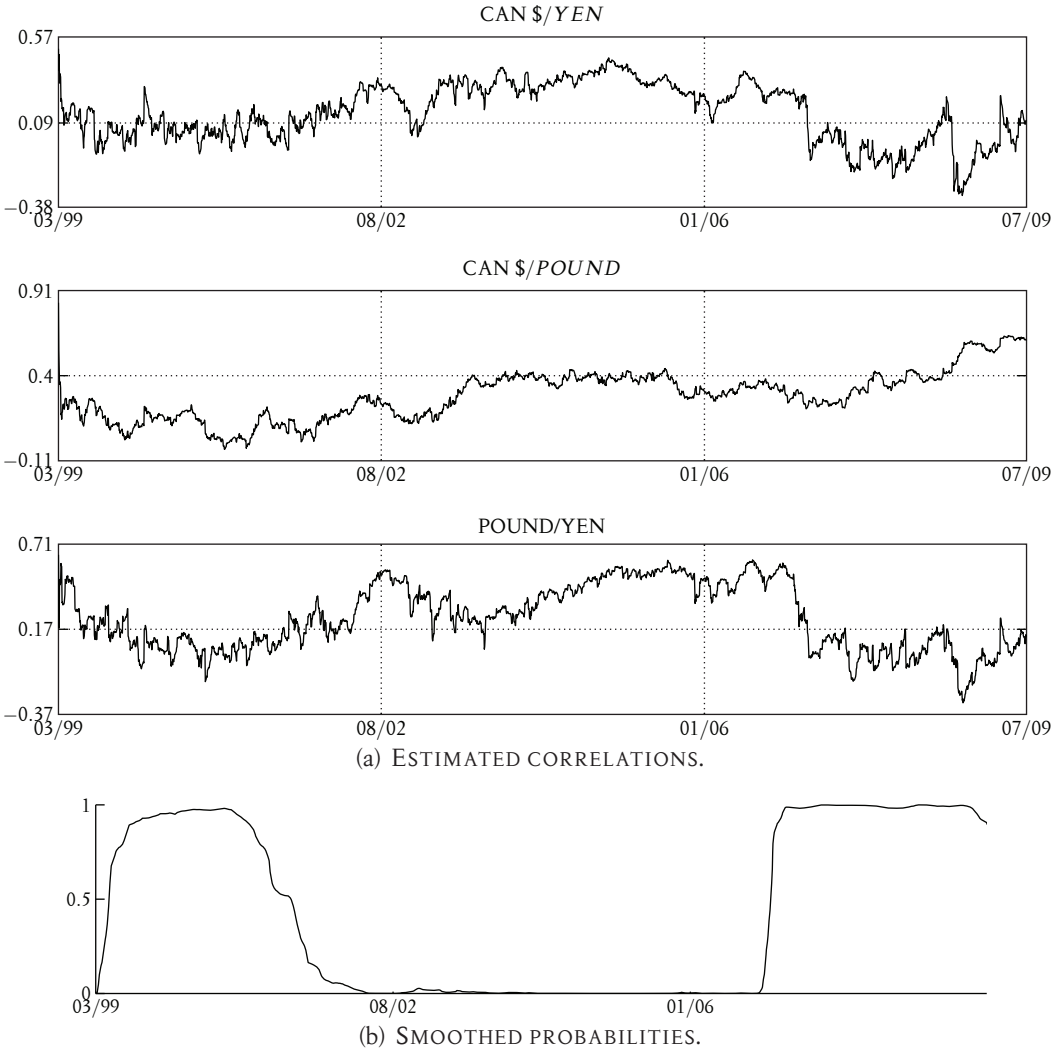
$$Q_{t,s_t} = \begin{bmatrix} q_{11,s_t} & q_{12,s_t} & q_{12,s_t} \\ & q_{22,s_t} & q_{23,s_t} \\ & & q_{33,s_t} \end{bmatrix} + \begin{bmatrix} a_{11,s_t} & 0 & 0 \\ 0 & a_{22,s_t} & 0 \\ 0 & 0 & a_{33,s_t} \end{bmatrix} \varepsilon_{t-1}\varepsilon_{t-1}' \begin{bmatrix} a_{11,s_t} & 0 & 0 \\ 0 & a_{22,s_t} & 0 \\ 0 & 0 & a_{33,s_t} \end{bmatrix}' + \begin{bmatrix} b_{11,s_t} & 0 & 0 \\ 0 & b_{22,s_t} & 0 \\ 0 & 0 & b_{33,s_t} \end{bmatrix} Q_{t-1,s_t} \begin{bmatrix} b_{11,s_t} & 0 & 0 \\ 0 & b_{22,s_t} & 0 \\ 0 & 0 & b_{33,s_t} \end{bmatrix}' \quad (3.40)$$

where s_t is an unobserved first order Markov chain process with a 2-by-2 transition matrix :

$$\mathrm{P} = \begin{bmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{bmatrix} \quad (3.41)$$

with $p_{ij} = \mathbb{P}[s_{t+1} = j | s_t = i],\ i, j = 1, 2$. In that case, variance targeting is not feasable and Q_{t,s_t} needs 7 parameters for the constant, 6 for the coefficients of the dynamic and 2 more parameters for the transition matrix. For two regimes, this makes a total of 28 parameters. Estimation was done using Gradient methods, filtered probabilities are computed with Hamilton's filter and smoothed probabilities are calculated with Kim's filter. Results appear on figure 3.5. The model clearly identifies two regimes. The switch from one to another is matched with the coming of the so-called *subprime crisis*. Comparing with the previous DCC model, this model has an intersting explanatory power.

We now apply the specification presented in this paper. With three series, our model has the



(a) ESTIMATED CORRELATIONS.

(b) SMOOTHED PROBABILITIES.

FIG. 3.5: Estimated correlations and smoothed probabilities of the multivariate Markov-switching DCC.

following representation :

$$Q_{t,s_t^{1,2,3}} = \overline{Q}_{s_t^{1,2,3}} + \begin{bmatrix} a_{11,s_t^1} & 0 & 0 \\ 0 & a_{22,s_t^2} & 0 \\ 0 & 0 & a_{33,s_t^3} \end{bmatrix} \varepsilon_{t-1}\varepsilon'_{t-1} \begin{bmatrix} a_{11,s_t^1} & 0 & 0 \\ 0 & a_{22,s_t^2} & 0 \\ 0 & 0 & a_{33,s_t^3} \end{bmatrix}' + \begin{bmatrix} b_{11,s_t^1} & 0 & 0 \\ 0 & b_{22,s_t^2} & 0 \\ 0 & 0 & b_{33,s_t^3} \end{bmatrix} Q_{t-1,s_t^{1,2,3}} \begin{bmatrix} b_{11,s_t^1} & 0 & 0 \\ 0 & b_{22,s_t^2} & 0 \\ 0 & 0 & b_{33,s_t^3} \end{bmatrix}' \tag{3.42}$$

where s_t^1, s_t^2 and s_t^3 are three unobserved first order Markov chains processes with transition matrix :

$$\mathrm{P^i} = \begin{bmatrix} p_{11}^i & 1-p_{22}^i \\ 1-p_{11}^i & p_{22}^i \end{bmatrix} \tag{3.43}$$

for $i = 1, 2, 3$. With three chains of two states, the factorial decomposition of the process, the model allows for eight combinations of regime. It implies that we need eight contants. There

are two ways to specify the constants. A naive approach is to estimate eight full matrix for the intercepts. But this solution is difficult from a numerical point of view. Another way to estimate these matrix is define the constants of the two extrem cases (all covariances are in regime 1, and all are in regime 2), and then construct the six other intercepts. Let

$$\overline{Q}_{s_t^1=1,s_t^2=1,s_t^3=1} = \begin{bmatrix} q_{11,s_t^1=1} & q_{12,s_t^1=1,s_t^2=1} & q_{13,s_t^1=1,s_t^3=1} \\ & q_{22,s_t^2=1} & q_{23,s_t^2=1,s_t^3=1} \\ & & q_{33,s_t^3=1} \end{bmatrix} \tag{3.44}$$

the constant corresponding at the case where all the covariances are in regime 1 and :

$$\overline{Q}_{s_t^1=2,s_t^2=2,s_t^3=2} = \begin{bmatrix} q_{11,s_t^1=2} & q_{12,s_t^1=2,s_t^2=2} & q_{13,s_t^1=2,s_t^3=2} \\ & q_{22,s_t^2=2} & q_{23,s_t^2=2,s_t^3=2} \\ & & q_{33,s_t^3=2} \end{bmatrix} \tag{3.45}$$

the intercept where all covariances are in regime 2. Given time series are ordered as CAD, YEN and POUND, the constant corresponding to the case where the covariance of CAD/YEN is in regime 1, CAD/POUND in regime 2 and YEN/POUND in regime 1 can be written using the parameters the previous matrix :

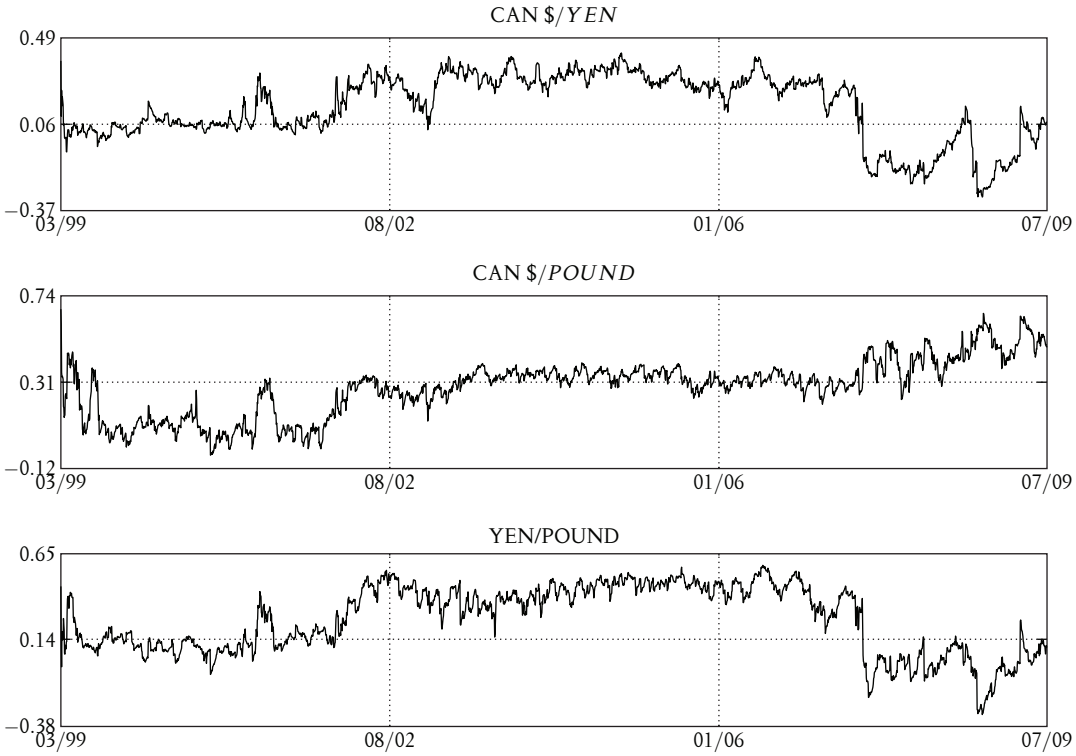
$$\overline{Q}_{s_t^1=1,s_t^2=2,s_t^3=1} = \begin{bmatrix} q_{11,s_t^1=1} & \boxed{q_{12,s_t^1=1,s_t^2=2}} & q_{13,s_t^1=1,s_t^3=1} \\ & q_{22,s_t^2=2} & \boxed{q_{23,s_t^2=2,s_t^3=1}} \\ & & q_{33,s_t^3=1} \end{bmatrix} \tag{3.46}$$

Non-boxed elements are coming from the constants of the two extrem cases. Boxed elements corresponds to the additional elements needed to define the intercept in that case. The constant corresponding to the case where the covariance CAD/YEN and CAD/POUND are in regime 1 and YEN/POUND in regime 2 can be written using the parameters the previous matrix :

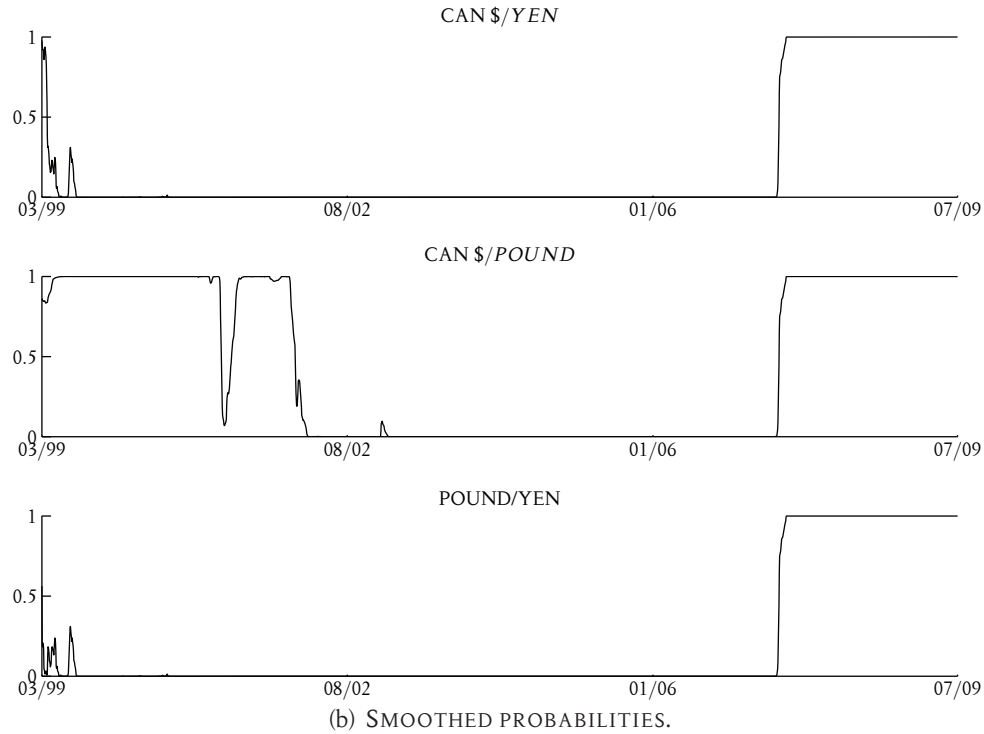
$$\overline{Q}_{s_t^1=1,s_t^2=1,s_t^3=2} = \begin{bmatrix} q_{11,s_t^1=1} & q_{12,s_t^1=2,s_t^2=1} & \boxed{q_{13,s_t^1=2,s_t^3=2}} \\ & q_{22,s_t^2=1} & \boxed{q_{23,s_t^2=2,s_t^3=2}} \\ & & q_{33,s_t^3=2} \end{bmatrix} \tag{3.47}$$

where boxed terms refers to the two additional elements to be estimated. By doing so for the other matrix, we need $14+6\times 2$ parameters for the constants (the complete formulation of the intercepts are given in appendix B.3). Given the three transition matrix need 3×2 parameters and the dynamics elements 12×2, the total number of parameters of our model is 42. Estimation was done using the regular HMM representation. This strategy has the advantage of allowing the use of standars tools (Hamilton filter and Kim's smoothing), but has the disadvantage of being a complex optimization problem. To avoid problem with local maxima, we have performed twenty estimation with various starting points. On our computer with Q6600 processeur, one estimate is approximately twelve hours. As point out by Engle (2009), the most complicated part of the estimation is the intercept matrix. Because variance targeting is not possible, we have to estimate eight intercept matrix. Even with the strategy exposed previously, the numerical constraints needed to keep these eight matrix positive definite make the maximisation of the likelihood difficult. Results are plotted on figure 3.6 and estimated parameters are in table 3.2.

The comparison of the smoothed probabilties plotted on figure 3.5(b) and 3.6(b) highlights the differences betwwen the classical Markov-switching approach and the factorial decomposition.



(a) ESTIMATED CORRELATIONS.



(b) SMOOTHED PROBABILITIES.

FIG. 3.6: Estimated correlations and smoothed probabilities of the multivariate Markov-switching DCC with FHMM specification.

With one Markov chain for all the conditional covariances, the DCC based on the specification of Haas and Mittnik (2008) identifies two regimes that appear once. The break is done at the

Parameters	Estimate	standart errors
a_{11}^1	0.1936	0.0124
a_{22}^1	0.1112	0.00556
a_{33}^1	0.3293	0.0080
a_{11}^2	0.0334	0.0092
a_{22}^2	0.2683	0.0163
a_{33}^2	0.1610	6.64e-04
b_{11}^1	0.9760	1.87e-04
b_{22}^1	0.9835	0.0163
b_{33}^1	0.9389	2.41e-05
b_{11}^2	0.9874	0.0049
b_{22}^2	0.9581	2.26e-04
b_{33}^2	0.9819	5.70e-04
$\overline{Q}_{s_t^1=1,s_t^2=1,s_t^3=1}$	$\begin{bmatrix} 0.0395 & -0.0025 & 0.0648 \\ & 0.0320 & 0.0087 \\ & & 0.1544 \end{bmatrix}$	$\begin{bmatrix} 2.84e{-}07 & 4.67e{-}05 & 5.25e{-}05 \\ & 9.10e{-}06 & 9.55e{-}05 \\ & & 2.08e{-}04 \end{bmatrix}$
$\overline{Q}_{s_t^1=1,s_t^2=1,s_t^3=2}$	$\begin{bmatrix} --- & --- & 0.0648 \\ --- & --- & 0.0087 \\ --- & --- & --- \end{bmatrix}$	$\begin{bmatrix} --- & --- & 1.59e{-}06 \\ --- & --- & 5.63e{-}05 \\ & & --- \end{bmatrix}$
$\overline{Q}_{s_t^1=1,s_t^2=2,s_t^3=1}$	$\begin{bmatrix} --- & 0.0549 & --- \\ --- & --- & 0.1248 \\ --- & --- & --- \end{bmatrix}$	$\begin{bmatrix} --- & 9.00e{-}05 & --- \\ --- & --- & 0.0051 \\ --- & --- & --- \end{bmatrix}$
$\overline{Q}_{s_t^1=1,s_t^2=2,s_t^3=2}$	$\begin{bmatrix} --- & -0.0466 & 0.0259 \\ --- & --- & --- \\ --- & --- & --- \end{bmatrix}$	$\begin{bmatrix} --- & 1.57e{-}05 & 5.81e{-}04 \\ --- & --- & --- \\ --- & --- & --- \end{bmatrix}$
$\overline{Q}_{s_t^1=2,s_t^2=1,s_t^3=1}$	$\begin{bmatrix} --- & 0.0527 & -0.0312 \\ --- & --- & --- \\ --- & --- & --- \end{bmatrix}$	$\begin{bmatrix} --- & 0.0076 & 7.68e{-}04 \\ --- & --- & --- \\ --- & --- & --- \end{bmatrix}$
$\overline{Q}_{s_t^1=2,s_t^2=1,s_t^3=21}$	$\begin{bmatrix} --- & -0.0278 & --- \\ --- & --- & 0.0144 \\ --- & --- & --- \end{bmatrix}$	$\begin{bmatrix} --- & 4.08e{-}07 & --- \\ --- & --- & 2.75e{-}07 \\ --- & --- & --- \end{bmatrix}$
$\overline{Q}_{s_t^2=1,s_t^2=2,s_t^3=1}$	$\begin{bmatrix} --- & --- & 0.0636 \\ --- & --- & 0.1200 \\ --- & --- & --- \end{bmatrix}$	$\begin{bmatrix} --- & --- & 0.0968 \\ --- & --- & 0.0046 \\ --- & --- & --- \end{bmatrix}$
$\overline{Q}_{s_t^1=2,s_t^2=2,s_t^3=2}$	$\begin{bmatrix} 0.1551 & 0.0885 & -0.0085 \\ & 0.1205 & 0.0033 \\ & & 0.0859 \end{bmatrix}$	$\begin{bmatrix} 6.20e{-}04 & 0.0012 & 0.0011 \\ & 4.69e{-}04 & 0.0052 \\ & & 0.0018 \end{bmatrix}$
p_{11}^1	0.9830	5.70e-04
p_{22}^1	0.4016	0.0375
p_{11}^2	0.9867	9.02e-04
p_{22}^2	0.9860	5.41e-04
p_{11}^3	0.8179	0.0096
p_{22}^3	0.4729	0.0425

TAB. 3.2: FHMM-DCC estimated parameters.

beginning of the subprime crisis. On the other hand, our specification, with one Markov chain for each conditional covariance, clearly shows it might be excessive to assume that all have the same switching dynamic. If all the conditional covariances switch at the beginning of the subprime crisis, it is worth noting they have truly distinct dynamics. This comparison clearly emerges that relaxing the constraint of common switching dynamic provides additional explanatory power.

3.4 Conclusion

In this paper, we have presented a generalization of the DCC model. Our model is based on the Markov-switching method introduced by Haas and Mittnik (2008). We have suggested a specification based on the factorial hidden Markov model (FHMM) of Ghahramani and Jordan (1997). Thus, in our model, each covariance of the conditional covariance matrix has its own dynamic. Consequently, our switching dynamic relax the classical assumption of Markov-switching models where all the conditional covariance switch togother from one regime to another. While the results of the application in this paper contribute to a better understanding of the dynamic of the correlations of exchange rate data, we must point out that several limitations must yet to be overcome. The most important of these is certainly the estimation of the model. Based on our specification, the increasing of time series leads to a complex numerical problem. These tasks await future research.

CHAPITRE

4

Modelling volatility and correlations with a hidden Markov decision tree.

4.1 Introduction

The past few years, multivariate volatility models have been a field of very active research. Related to this topic, the study of multivariate GARCH with conditional correlations has attracted a stong interest since the seminal paper of Bollerslev (1990). Introducing dynamic in the conditional correlations has led this pioneering approach to creating a new class of model, the wellknown Dynamic Conditional Correlation (DCC) models. Proposed by Engle and Sheppard (2001) and by a slightly different maner by Tse and Tsui (2002), the DCC becomes an attractive parametrization.

This new class now includes many extensions. Engle (2002) generalized the model by introducing full matrix instead of scalars to drive the dynamics of the process (Generalised DCC). Therefore, this unrestricted specification requires a large number of parameters and make estimation unfeasable. To solve the curse of dimensionality while relaxing constraints on the dynamics, various extensions have been proposed. Most of them are based on combinations of specific matrix, partionned vectors or block matrix (see Billio and Caporin (2009), Hafner and Franses (2009) and Billio, Caporin, and Gobbo (2006)) but also clustering techniques (see Zhou and Chan (2008)). A DCC incorporate exogenous variables is proposed by Vargas (2008) to identify factors that could lead correlations. A semi-parametric approach is presented by Franses, Hafner, and van Dijk (2005), using a transformed Nadaraya-Watson estimator. Feng (2007) also presents a local estimator for the correlations based on k-nearest-neighbours (k-NN) methods.

Others extensions focus on the asymmetric effect in the conditional correlation. Asymmetry refers to the upward motion of volatility more after a negative than after a positive shock. An asymmetric extension has been proposed by Cappiello, Engle, and Sheppard (2006) to study worldwide linkages in the dynamics of the correlations of selected bonds and equity markets. Vargas (2006) extends the Block DCC of Billio, Caporin, and Gobbo (2003) to take account of the asymmetry between blocks of asset returns. Cajigas and Urga (2006) also propose an asymmetric generalised DCC where standardized residuals follow an asymmetric multivariate Laplace distribution.

Another direction of extensions refers to introduce regime switch in the conditional correlations to take into account empirical findings in case of turbulent periods. In this context, Silvennoinen and Teräsvirta (2005) assume the conditional correlation matrix varies moothly between two constant correlations matrix. The link between these two extrem matrix is done via a conditonal logistic function according to a exogenous or endogenous transition variable. This model has been improved to allows another transition around the first one (see Silvennoinen and Teräsvirta (2009)). Differently from the previous STAR approach, Pelletier (2006) suggested to introduce Markov switch to the correlation. In this way, correlations are constants within regime but vary from one regime to another. Both previous approach assume correlations evolve between constant correlations matrix. Billio and Caporin (2005) explained a different strategy with an extension of the scalar DCC of Engle and Sheppard (2001) where both the parameters and the unconditional correlation are driven by a hidden Markov chain.

Alongside these developments, the technical issue of the curse of dimensionnality, which underlies most of the developments in the field of multivariate GARCH, has been the subject of recents advances. This problem of the explosion in the number of parameters arises fastly : with k time series, the numbers of correlation to estimate is $k(k-1)/2$. Several strategies have been proposed. Palandri (2009) decomposes the correlations and partial correlations. This decomposition transforms a high dimensional optimization problem into a serie of simple estimates. Engle, Shephard, and Sheppard (2008) also exploited the idea of decompose a complex estimation problem into a set of simpler models to construct a composite likelihood which is obtained from the summation of likelihood of subsets of assets. Engle and Kelly (2009) suggests a radically different strategy based on an updated approximation of the correlation. At each time, they assumes all pairwise correlations are equal. Engle (2008) also exposes an alternate solution based on approximation. This aptly

named *MacGyver* method is based on the bivariate estimation of each pair and then calculate the median of theses estimators.
This brief summary of the literature is far from exhaustive and we refer to Bauwens, Laurent, and Rombouts (2006) for a general survey on multivariate GARCH. The paper of Silvennoinen and Teräsvirta (2009) gives an useful review of recents advances and Engle (2009) outlines a thorough presentation of conditional correlations models.
The model we present is an extention of multivariate GARCH with conditional correlations based on a tree-structure. Models based on tree structure have already been proposed. Barone-Adesi and Audrino (2006) apply a tree structure developed by Audrino and Buhlmann (2001) and Audrino and Trojani (2006) to a rolling window averaged conditional correlation estimator. This method allows thresholds in the conditional correlations. The estimator is a convex combination of realized correlations and estimates of averaged correlations. As the averaged conditional correlation is constructed with univariate GARCH models, this approach has the advantge to required few parameters. This study has been extended using non-parametric functions based on functional gradient descent to estimate univariate volatilities by Audrino (2006). Also based on the idea of Audrino and Buhlmann (2001), the model suggested by Trojani and Audrino (2005) adopts the two-step procedure exposed in Engle and Sheppard (2001). In first time, conditional variance are extracted with tree-structured GARCH. Conditional correlations are then computed from standardized residuals and have also their own tree-structured dynamic. An tree-structured extension of the DCC is suggested by Dellaportas and Vrontos (2007) to study volatility and co-volatility asymmetries. Making a link between conditional variances and correlations, this model allows to exhibit multivariate thresholds. Using a sequence of binary decisions rules, each terminal node is matched with a multivariate GARCH.
One commun feature of theses approaches is they are all based on the idea of binary tree. Time series are recurcively partionned using binary decisions. The resulting tree is purely deterministic. At the opposite, in this paper, we propose en extension of the DCC based on a stochastic decision tree. Our model links the univariate volatilities with the correlations via a hidden stochastic decision tree. Hidden Markov Decision Tree (HMDT) model is an extension of Hidden Markov Model (HMM) introduced by Jordan, Ghahramani, and Saul (1997). Based on a Markov temporal structure, the architecture is at the opposite of the classical deterministic approach on a binary decision tree. Our proposed models can be estimated via maximum likelihood.
The paper is organized as follows. The model is defined section. We firts briefly recall the basics of the DCC of Engle and Sheppard (2001) and then outline our model. Section the model is applied to data examples. In the first application, we apply our model to exchange rate data with in-sample and out-of-sample analysis. Final remarks concludes and proposes areas for further research in section.

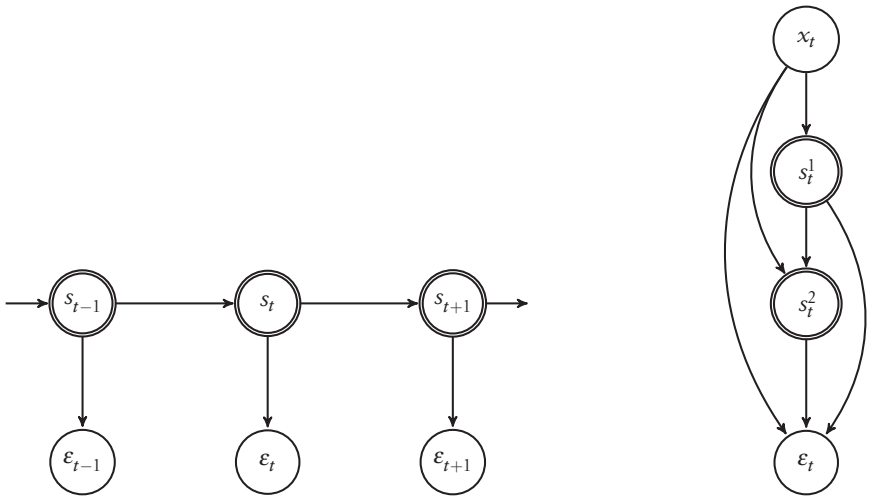
4.2 The Model

4.2.1 What is a Hidden Markov Decision Tree ?

The approach that we adopt for this paper is developed around the idea of hierarchical architecture with a Markov temporal structure. This model can be explained in one of two ways.

A Hierarchical Mixtures-of-Experts with Markov dependencies

This approach can be first interpretated as a hybrid model that blends the Hierarchical Mixtures-of-Experts (HME) of Jordan and Jacobs (1994) and Hidden Markov models (HMM). The HME is based on the *divide-and-conquer* algorithm. This methodology is to fit the data by dividing



(a) HIDDEN MARKOV MODEL.

(b) HIERARCHICAL MIXTURE OF EXPERTS.

FIG. 4.1: The hidden Markov model and the hierarchical mixture of experts as graphical models.

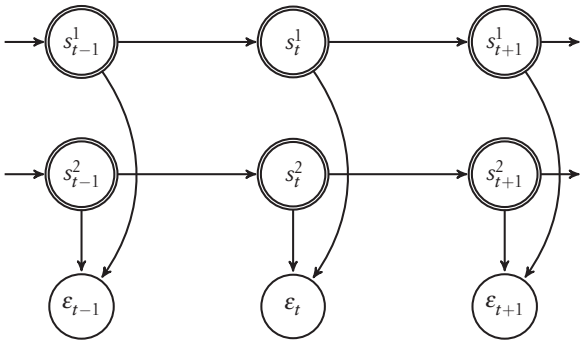
the space into nested subspaces and estimate a sub-model for each of these regions (see figure 4.1(a) for the representation of HME as a directed acyclic graph). The data are fitted by a model which is the overall combination of each sub-models. This philosophy has given rise to numerous approaches as, without being exhaustive the CART (see Breiman, Friedman, Olshen, and Stone (1984)), MARS (see Friedman (1991)) or GUIDE (see Loh (2002)). HME is itself an generalization of the Mixture-of-Experts (ME) models of Jacobs, Jordan, Nowlan, and Hinton (1991). It assumes that the data can be fitted by a piecewise function. The approach is based on two elements. First, the models has *n experts* to fit the data in *n* regions. An expert can be a constant, AR or GARCH process. The second element is a *gating network*, that associate to each region with a expert given an input variable. The final output is is a convex combination of the expert outputs given the input. The ME architecture has been applied to time series by Carvalho and Tanner (2006). This architecture has only one level of experts associated with one gating network. The HME generalize this approach by allowing different level of experts matched by various gating functions. Huerta, Jiang, and Tanner (2003) use a two layers HME to study the US industrial production index with *trend-stationary process experts* and *difference-stationary process experts*. As explained in McLachlan and Peel (2000), the main difference between CART/MARS/GUIDE approaches and HME divide the covariate space with probabilistic scheme whereas the first use hard bondaries. It also may be pointing out of the link between ME architecture and (S)TAR models. Indeed, in a STAR-GARCH model the transition function can be interpretated as the gating network and the two GARCH component as the experts. The same comparison is true for the HME and the multiple-regime STAR of van Dijk and Franses (1999).

Hidden Markov models has emerged as a standart tools to modeling time series with regime switching. This approach has been extensively used, whether in graphical models (see Rabiner (1989)) or in econometrics with its counterpart Markov-switching (see among others Hamilton (1994). HMM is based on the assumption that each observation is linked to a finite number of hidden states via a probability distribution (see figure 4.1(b)). The conditional probability distribution of the hidden variable generally assumes to be a first order Markov chain. In time series analysis, this has first been used by Hamilton (1989) to make a regime switching AR process and generalized by Krolzig (1997) to the VAR case. Exemple of applications to GARCH models can be found in

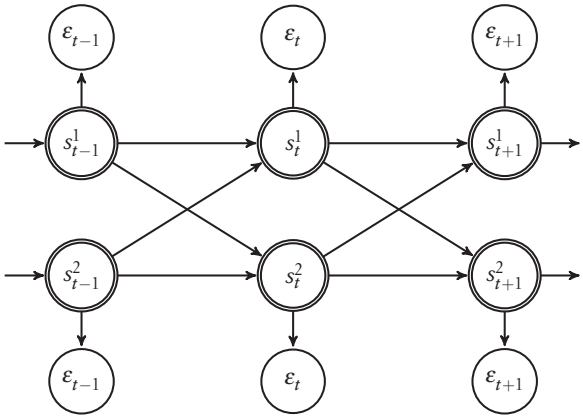
Hamilton and Susmel (1994), Gray (1996), Dueker (1997), Klaassen (2002) and Haas, Mittnik, and Paollela (2004b). Markov-switching based extensions of multivariate GARCH models with dynamic correlations have also been proposed. Billio and Caporin (2005) propose a DCC model with Markov switch in the unconditional correlation process. Application of this MS-DCC model to daily stock market indices from January 2000 to December 2003 highlights the advantage of introducing Markov jumps during turbulent periods. Pelletier (2006) also suggested a Markov-switching DCC. This regime switching for dynamic correlations (RSDC) model assumes that correlations are constant within regime but vary from one regime to another.
The HMDT can be seen as a hybrid model combining the two previous approach. The result gives a probabilistic decision tree where the decision at each time is conditional of the decision at the previous time.

A factorial and coupled HMM

In this paper, we adopt another look of the HMDT. Contrary to the previous approach where the HMDT is defined as a hierarchical mixture with Markovian dynamics, we can also consider this architecture as a pure HMM extension. Indeed, HMDT can be view as both a factorial HMM and a coupled HMM.
The factorial hidden Markov (FHMM) model has been proposed by Ghahramani and Jordan (1997). This extension of the basic HMM assumes the state variable is factored into various state variables. Indeed, each of these has its independent Markovian dynamic and the output is the combination of these several underlying processes (see figure 4.2(a)) . Because the model has various



(a) FACTORIAL HIDDEN MARKOV MODEL.



(b) COUPLED HIDDEN MARKOV MODEL.

FIG. 4.2: Factorial and Coupled HMM as graphical models.

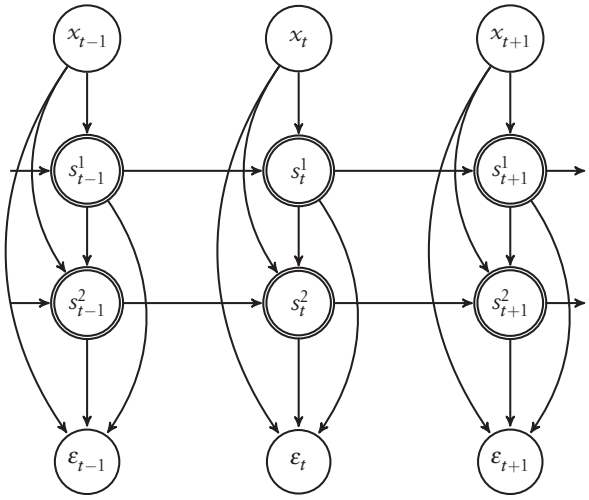


FIG. 4.3: The hidden Markov decision tree model views as a directed acyclic graph.

independent hidden Markov models in parallel, the resulting state space of the model is the cartesian product of the parallel sub-processes. The FHMM has been widely used in speech recognition because of its ability to model time series generated from various independent sources (see among others Roweis (2000) for an application to microphone source separation). The distributed state representation of the FHMM also provide an interesting framework for handwriting recognition (see Williams, Toussaint, and Storkey (2007) for the use of a primitive model based on FHMM to represent natural handwriting data).

Coupled hidden Markov models (CHMM) are also an extension of the classical HMM. It consists of a model with several HMMs whose Markov chains interact together (see Brand (1997)). There are two ways to allow the states to have interdependencies. We can qualify as *independent coupling* when various processes are coupled at the output. From this point of view, FHMM can be considered as an independent coupled HMM. Another kind of making interactions between chains is to assume that processes are dependent. In *dependent coupling*, the current state is dependent of the other states of its own chain but also depends on the previous state of another chain and so on (see figure 4.2(b)). Dependent coupled hidden Markov models include many architecture, as models with different lag, nested loop and non-linear dependancy relationship. Coupled HMM have been used in various fields like, without being exhaustive, facial event mining (see Ma, Zhou, Celenk, and Chelberg (2004)), video-realistic speech animation (see Xie and Liu (2007)), audio-visual speech recognition (see Nefian, Liang, Pi, Xiaoxiang, Mao, and Murphy (2002)).

In this framework, HMDT can be view as a HMM which is both factorial and dependent coupled. The factorial decomposition provides a factorised state space. This state space decomposition is done using constant or state dependent time-varying transition probabilities given an input variable. As described in Brand (1997), the top level of the tree can be seen as a *master* process and the following levels as *slave* processes. The constaint of a level on the following is done via a coupling transition matrix which produce the ordered hierarchy of the structure. As the links between decision states are driven with Markovian dynamics and the switch from one level to the following is done via a coupling transition matrix, this architecture gives a fully probabilistic decision tree.

4.2.2 Hidden Markov decision tree for correlations

Starting point

The model we propose in this paper is an extension of the DCC of the scalar DCC (see Engle and Sheppard (2001)). Given y_t an K dimensional time series of length T the DCC class models assumes :

$$y_t|\mathscr{F}_{t-1} \overset{iid}{\sim} \mathscr{N}(0, H_t) \tag{4.1}$$

where $\mathscr{F}_{t-1}$ refers to the information set at time $t-1$. Conditional variance-covariance matrix of returns y_t is expressed as follows :

$$H_t = D_t R_t D_t \tag{4.2}$$

where R_t is a $K \times K$ constant correlation matrix. the matrix D_t is a $K \times K$ diagonal matrix containing univariate time varying standard deviations :

$$D_t = \text{diag}\{h_{i,t}^{1/2}\} \tag{4.3}$$

for $i = 1, ..., K$. Getting the matrix D_t is generally refered as the so-called *degarching* filtration required to construct standized residuals expressed as :

$$\varepsilon_t = D_t^{-1} r_t \tag{4.4}$$

The conditional correlations are simply the expectation of the standardized residuals :

$$\mathbb{E}_{t-1}[\varepsilon_t \varepsilon_t'] = D_t^{-1} H_t D_t^{-1} = R_t \tag{4.5}$$

In the scalar DCC of Engle and Sheppard (2001), correlations are computed using a combination of the conditional covariance of the standardized residuals $Q-t$ which follows a scalar BEKK process :

$$Q_t = (1-a-b)\overline{V} + a\varepsilon_{t-1}\varepsilon_{t-1}' + bQ_{t-1} \tag{4.6}$$

where $\overline{Q}$ is the unconditional covariance matrix of the standardized residuals and a and b are two positive scalars such that $a+b<1$. The conditional correlation matrix is then obtained by :

$$R_t = \text{diag}\{Q_t\}^{-1/2} Q_t \text{diag}\{Q_t\}^{-1/2} \tag{4.7}$$

Our model has the DCC of Engle and Sheppard (2001) as starting point. We propose to improve this specification by introducing a hidden decision. To study the relationship between univariate volatility and correlation, this tree has two levels. The first established a splitting rule to discriminate low and high regime of variance while the second level discriminates low and high correlations regime. Thus, this construction is a suitable method to investigate the link between high (low) volatility and high (low) correlations. Specifically, it allows us to study the probability of being in a type regime of correlations given the regime of the univariate conditional variance.

Hidden tree structure

As point out in section 4.2.1, a HMDT can be view as a factorial and coupled HMM. The factorial HMM decomposition is used to divide the space of the time series into low/high conditional variance for each series and for low/high sequence of correlations. Transition matrix used in our model can be both static or time-varying and can depend of endogenous or exogenous variables. Formally, at the first level, each time series have its own conditonal variance driven by a transition matrix with time varying probabilities. Univariate high and low conditional variance follow a two

state Markov process of first order such. It means the state s_t is independant of state lagged state except s_{t-1} :

$$\mathbb{P}[s_{t+1} = i | s_t = j, ..., s_t = h] = \mathbb{P}[s_{t+1} = i | s_t = j] \tag{4.8}$$

To introduce time-varying transition probabilities, we can use the specification proposed by Diebold, Lee, and Weinbach (1994). In that way, transition probabilities are assumed to follow a logistic function of a endogenous or exogenous variable. Elements of the transition matrix $\mathrm{P}^{\mathrm{k}}_{\mathrm{vol}}$ for the conditional variance of the k^{th} with row i and column j is defined as :

$$p^k_{ij}(t) = \mathbb{P}[s_{t+1} = j | s_t = i; x_{t-1}, \gamma^k_n], \quad i, j = 1, .., N \tag{4.9}$$

Introducing time-variation, transition probabilities are then expressed as :

$$p^k_{ij}(t) = \frac{\exp(x'_{t-1}\gamma^k_n)}{1 + \exp(x'_{t-1}\gamma^k_n)} \tag{4.10}$$

with x_{t-1} is the conditionning vector refering as the input of figure 4.3 and γ its coefficient. As we partition the space of the univariate conditional variance in two subspaces, low and high variance, the transition matrix for the variance of the k^{th} time serie is of size 2×2 and can be expressed as follow :

$$\mathrm{P}^{\mathrm{k}}_{\mathrm{vol}} = \begin{bmatrix} p^k_{11}(t) & 1 - p^k_{22}(t) \\ 1 - p^k_{11}(t) & p^k_{22}(t) \end{bmatrix} \tag{4.11}$$

Since all univariate volatility processes have the same Markovian dynamic specification, individual transition matrix can be aggregate to constitute the first level of the decision tree. Individual HMMs are aggregated in a factorial HMM representation. The dynamic of the univariate volatility level can be resumed in a general transition $\mathrm{P}_{\mathrm{vol}}$ by the cross product of the transition matrix of univariate volatility models $\mathrm{P}^{\mathrm{k}}_{\mathrm{vol}}$:

$$\mathrm{P}_{\mathrm{vol}} = \bigotimes_{i=1}^{K} \mathrm{P}^{\mathrm{i}}_{\mathrm{vol}} \tag{4.12}$$

This factorial representation allows to represent all the dynamics of the K univariate volatilities containing 2 states with a single transition matrix of size $2^K \times 2^K$.
The same specification is used for the second level. This level discriminates between low and high correlations. Thus, the decision step is represented by a 2-by-2 transition matrix written as :

$$\mathrm{P}_{\mathrm{corr}} = \begin{bmatrix} p^c_{11}(t) & 1 - p^c_{22}(t) \\ 1 - p^c_{11}(t) & p^c_{22}(t) \end{bmatrix} \tag{4.13}$$

whose elements can be both static and time-varying. In that case, this is the same specification as defined in the equation (4.10). Given the transition matrix of the first and second level, the partition of the space is represented by transition matrix P expressed as :

$$\mathrm{P} = \mathrm{P}_{\mathrm{vol}} \otimes \mathrm{P}_{\mathrm{corr}} \tag{4.14}$$

of size $2^{K+1} \times 2^{K+1}$.

Given this space partition the first and the second level are then linked. This relation is fully probabilistic and attribute a weight to the decision related to the correlation given the decision of the univariate volatility. This relation is done using a *coupling* matrix. Coupling matrix are generally used in the context of coupled HMMs where a set of Markov chain influence each other. In our purpose, the matrix we use verifies the Markov property. Thus, the dynamic of the junction

between the two levels is done with an abstract Markov chain. This linking relation does not emit directly observation but plays the part of attributing a weight between volatility and correlations. This coupling matrix is of size 2-by2 and is written as :

$$\mathrm{P}_{\text{coupl}} = \begin{bmatrix} c_{11}^{(}t) & 1-c_{22}^{(}t) \\ 1-c_{11}^{(}t) & c_{22}^{(}t) \end{bmatrix} \tag{4.15}$$

The elements of this matrix can alos be time-varying given the input and are defined with the logistic specification of the equation (4.10).

Thereby, the dynamic of the probabilistic tree can be can be summarized as follow. All the ordered hierarchy is based on a set of HMMs which make the partition of the space. This partition is slightly different from that usually considered in binary tree models where the space is partitioned recurcively using binary rules. In our model, the partition is static and is defined a priori. The decision process occurs according to a cascade of ordered synchronous HMMs. The coupling process has not the Markovian property but can be interpretated as vector of probabilities, whose sum is equal to one. By making the link between the HMMs, the coupling matrix captures interprocess influences. Conditional time-varying transition and coupled probabilities allow the structure to adapt for each observation.
In our model, we use a set of HMMs to the coupling vector is used to weight influence of the univariate volatility on the correlations. The first level discriminates each time serie between low and high volatility while the second level ordered between high and low correlations. The input variable can be both endogenous or exogenous. A consensual choice could be time calendar.

Specification for univarite volatilities

In our decision tree, the first level distinguishes between low and high volatility. Thus, this decision step parametrizes a univariate regime switching GARCH model. Among the litterature of time series, many specifications have been proposed. Most of them have the drawback to require approximation schemes to avoid the path dependency problem. This is the case of the models suggested by Gray (1996) and Klaassen (2002). Beyond this numerical aspect, specification with approximations can entail difficulties in the interpretation of the processes corresponding to each regime. This motivation led us to use the Markov-switching GARCH proposed by Haas, Mittnik, and Paollela (2004b). This model is expressed as follow :

$$\begin{pmatrix} h_{1,t} \\ \vdots \\ h_{N,t} \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_N \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} y_{t-1}^2 + \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \odot \begin{pmatrix} h_{1,t-1} \\ \vdots \\ h_{N,t-1} \end{pmatrix} \tag{4.16}$$

where $\odot$ means the element-by-element multiplication. Stationnarity condition implies $\alpha_n + \beta_n < 1$ for each $n = 1, ..., N$. Besides its computational advantages, this model has a clear cut interpretation. It assumes the conditional variance can switch between N separate GARCH models that evolve in parallel. Because we need two regimes, each time serie involves eight parameters.

Specification for correlations

The second level discriminates between low and high conditional correlations. Similarly with the first level, we use a specification which does not require approximation in order to facilitate the interpretation of the regimes. Haas and Mittnik (2008) have proposed an extension of the Markov-switching GARCH to the multivariate case. Since the conditonal covariances of the standardized

residuals has a BEKK dynamic, we can use this Markov-switching extension in the scalar case. It can be written as :

$$\begin{pmatrix} Q_{1,t} \\ \vdots \\ Q_{N,t} \end{pmatrix} = \begin{pmatrix} \Omega_1 \\ \vdots \\ \Omega_N \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \varepsilon_{t-1}\varepsilon'_{t-1} + \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} \odot \begin{pmatrix} Q_{1,t-1} \\ \vdots \\ Q_{N,t-1} \end{pmatrix} \tag{4.17}$$

Stationnarities conditions imply the intercept Ω_n to be a positive definite matrix and $a_n + b_n < 1$ for $n = 1, ...N$. This specification requires $K(K+1)/2+4$ parameters.

4.2.3 Estimation

Estimation of the model is done with maximum likelihood. One of the advantages of the DCC of Engle (2002) is the possibility to estimate the model in a two steps approach. The idea is to estimate the parameters related to the univariate volatiliites in a first step and in a second step whose related to the correlations. In our case, because of the relation between the degarching process and the correlation, we must estimate the model in one step. With the assumption of normality, the log-likelihood can be written :

$$L = -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi) + \log(|H_t|) + y'_t H_t^{-1} y_t\right) \tag{4.18}$$

To maximize the likelihood we need to make inference on the state of the various Markov chains of the model. The strategy developed in our paper is to convert the complex dynamic of the factorial and coupled HMM into a simple regular HMM so that we can use standarts tools on filtering and smoothing probabilities. A conversion relationship between coupled HMM and standart HMM has been proposed by Brand (1997). Given $\mathrm{P}_{\mathrm{S|S}}$ and $\mathrm{P}_{\mathrm{S'|S'}}$ two transition matrix, $\mathrm{P}_{\mathrm{S|S'}}$ and $\mathrm{P}_{\mathrm{S'|S}}$ two coupling matrix, the relationship of Brand (1997) is given by :

$$(\mathrm{P}_{\mathrm{S|S}} \otimes \mathrm{P}_{\mathrm{S'|S'}}).R(\mathrm{P}_{\mathrm{S'|S}} \otimes \mathrm{P}_{\mathrm{S|S'}}) \tag{4.19}$$

where R is a row permute operator swaping fast and slow indices. In our case, because we use only one coupling vector in a downward direction, the regular HMM representation $\mathrm{P}_{\mathrm{reg}}$ of our model can be expressed as :

$$\mathrm{P}_{\mathrm{reg}} = (\mathrm{P}_{\mathrm{vol}} \otimes \mathrm{P}_{\mathrm{corr}}) \odot (\mathrm{P}_{\mathrm{coupl}} \otimes (\iota\iota')) \tag{4.20}$$

with ι a vector of ones of length 2^{K+1}.
The convertion into regular HMM allows to use standard tools about the inference of the Markov chains. Let ξ_{jt} the probability to be in regime j given the information set avalaible in $t-1$ and η_{jt} the density under the regime j. The probability to be in each regime at time t given the observations set up to t, written $\hat{\xi}_{t|t}$, can be calculated using the following expression of the so-called Hamilton's filter :

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{\imath'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{4.21}$$

and :

$$\hat{\xi}_{t|t+1} = \mathrm{P}_{\mathrm{reg}} \times \hat{\xi}_{t|t} \tag{4.22}$$

where $\odot$ denotes element by-element multiplication. The conversion to regular HMM has the advantage to be easy to compute. Interpretation is also natural. Each state of the regular representation corresponds of a combination of possible cases, e.g. the first series in high or low

volatility period, the second in high or low, etc. and the correlation in low and high period. The main drawback of the conversion is the complexity of the maximum likelihood estimation. Optimization methods based on gradient methods using numerical derivatives can have difficulties to escape from local maxima. That therefore justifies to test several initial conditions.

4.3 Application

For the first application, we propose to illustrate our proposed model on a daily dataset. We consider futures prices on 10 year treasury bonds and the S&P 500 from september 1994 to february 2003 (2201 observations). The two series comes from DataStream and are plotted in figure 4.4. Returns are calculated by taking 100 times the first difference of the logarithm of each series minus the sample mean. Descriptives statistics of the log-returns are given in table 4.1. All the results reported in the analysis were generated with Matlab on Linux.

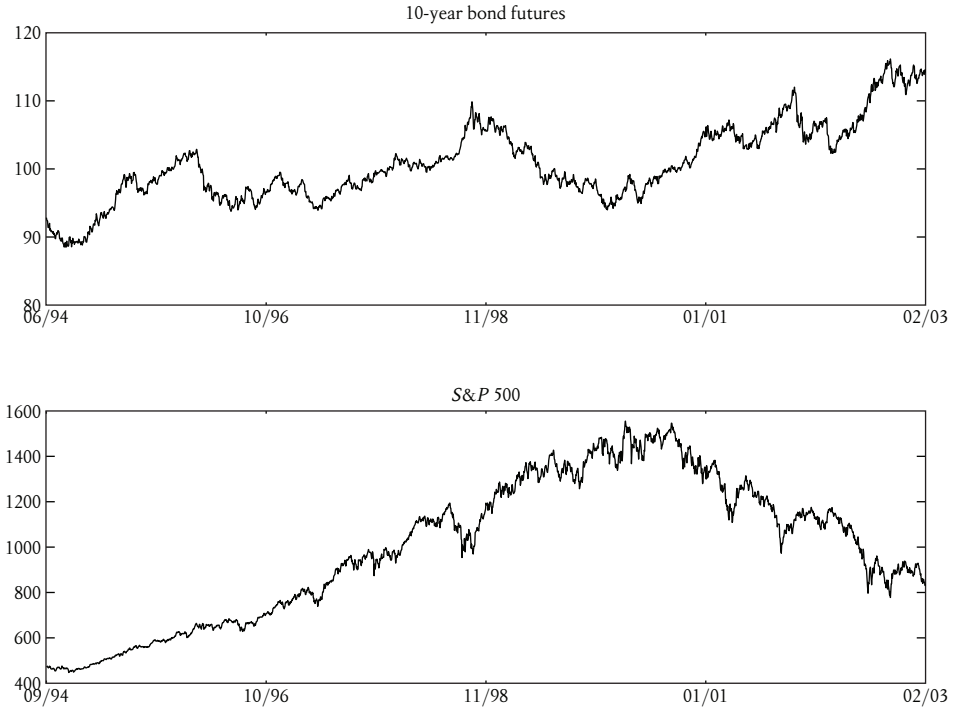


FIG. 4.4: Futures prices of 10-year bond and S&P500 1994 to july 2003.

With two series, the dynamic of the univariate volatility level can be summerized in a 4-by-4 transition matrix (see equation (4.12)). The second level has a 2-by-2 transition matrix. Thus, given equation (4.14), the space is partionned into eigh sub-spaces, each corresponding of a special case. Let $h_{t,Y}$ and $h_{t,P}$ the univariate volatilities of the bonds and the S&P 500. The superscript h denotes a high regime of volatility and l a low regime. A high regime of correlations is written by R_t^h and a low by R_t^l. The enumeration of the various cases corresponds to :

1. $\{h_{t,Y}^h, h_{t,P}^h\}, \{R_t^h\}$
2. $\{h_{t,Y}^h, h_{t,P}^h\}, \{R_t^l\}$
3. $\{h_{t,Y}^h, h_{t,P}^l\}, \{R_t^h\}$
4. $\{h_{t,Y}^h, h_{t,P}^l\}, \{R_t^l\}$

	Bond $	S&P500
Minimum	-2.2154	-7.7621
Maximum	1.4175	5.7549
Mean	0.0094	0.0258
Variance	0.1430	1.5191
Skewness	-0.5310	-0.1236
Kurtosis	5.4602	6.5324
Jarque-Bera	658.2178 (1.0000e−03)	1.1494e+03 (1.0000e−03)
Box-Ljung Q(25)	30.9139 (0.1919)	29.8141 (0.2313)
Box-Ljung Q(50)	65.4385 (0.0703)	64.6047 (0.0802)

TAB. 4.1: Descriptive statistics for the 10 year treasury bonds and the S&P 500 futures prices.

5. $\{h^l_{t,Y}, h^h_{t,P}\}, \{R^h_t\}$
6. $\{h^l_{t,Y}, h^h_{t,P}\}, \{R^l_t\}$
7. $\{h^l_{t,Y}, h^l_{t,P}\}, \{R^h_t\}$
8. $\{h^l_{t,Y}, h^l_{t,P}\}, \{R^l_t\}$

With two series, the decision can be summerized as in figure 4.5. The coupling matrix is then of size 2×2. This coupling matrix links each state $\{h^i_{t,Y}, h^j_{t,P}\}$, $i, j = h, l$ with a state $\{R^i_t\}$, $i = h, l$. The relation introduced by this vector can be view as a weighted function.
With two series, the number of parameters related to the first level is 8 and 12 for the correlations. With 2 parameters for the coupling matrix, the total number of parameters to estimate is 30. As explained in the previous section, all the decisions of the first level is resumed in a transition matrix which is the cross product of the two tansition matrix related to each serie.

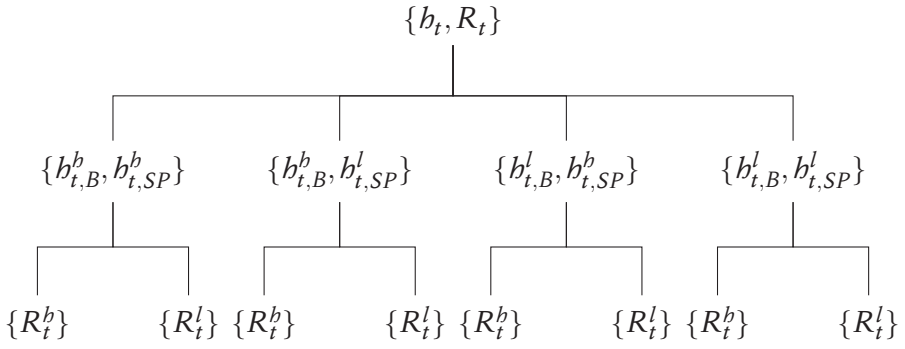


FIG. 4.5: Structure of the decision process for a bivariate dataset.

4.3.1 Results of the model

We first apply our model with constant transition probabilities. This configuration has no input. Estimated correlations are plotted on figure 4.6(a). Our results confirm those of Colacito and Engle (2006), who had previously studied this dataset. The correlation are positive in the first part of the sample and becomes negative at this end.

First level

	10-year bond futures		S&P500	
Parameters	Estimate	standard errors	Estimate	standart errors
ω_1	0.0027	4.07e-6	0.0082	5.50e-6
α_1	0.0434	1.66e-4	0.0694	1.92e-4
β_1	0.9394	5.45e-4	0.9263	1.91e-4
ω_2	0.0342	0.0101	0.0157	6.48e-5
α_2	0.1652	0.3214	0.0911	0.0013
β_2	0.8201	0.3342	0.8992	0.001
p_{11}	0.9981	1.50e-4	0.9998	5.08
p_{22}	0.0010	7.69e-8	0.6522	0.0154

Coupling matrix

Parameters	Estimate	standart errors
p_{11}	0.2522	0.0584
p_{22}	0.8629	0.0049

Second level

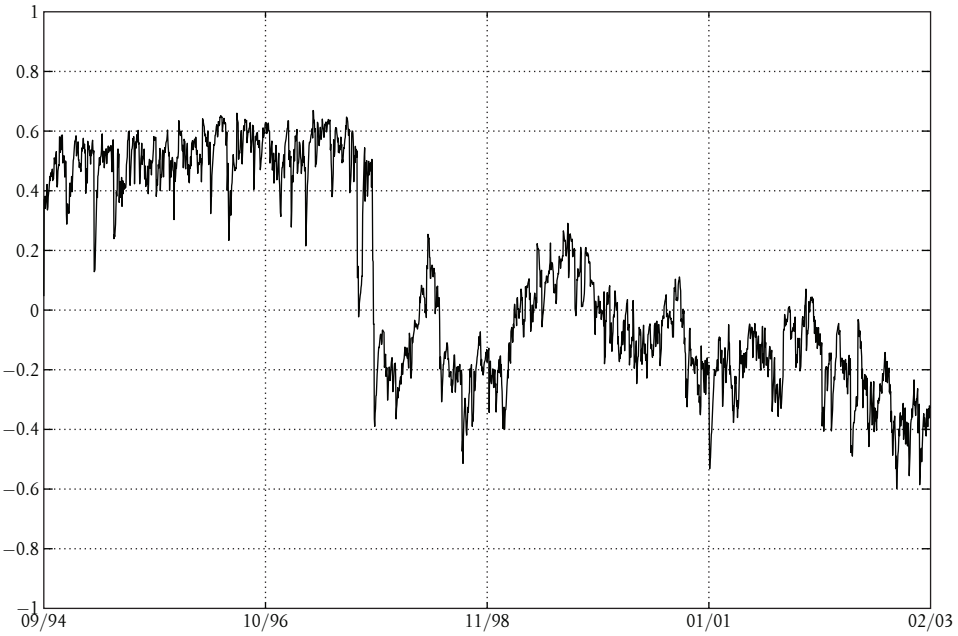
Parameters	Estimate	standart errors
a_1	0.1532	0.1372
b_1	0.7349	0.2126
const	$\begin{bmatrix} 0.1281 & 0.2852 \\ & 0.6880 \end{bmatrix}$	$\begin{bmatrix} 8.44e-4 & 0.0013 \\ & 0.0092 \end{bmatrix}$
a_2	0.0138	2.76e-5
b_2	0.9567	2.61e-4
const	$\begin{bmatrix} 0.0037 & -0.0087 \\ & 0.0164 \end{bmatrix}$	$\begin{bmatrix} 5.32e-6 & 1.53e-5 \\ & 0.0037 \end{bmatrix}$
p_{11}	0.8843	0.0154
p_{22}	0.2502	0.0305

TAB. 4.2: Model estimates.

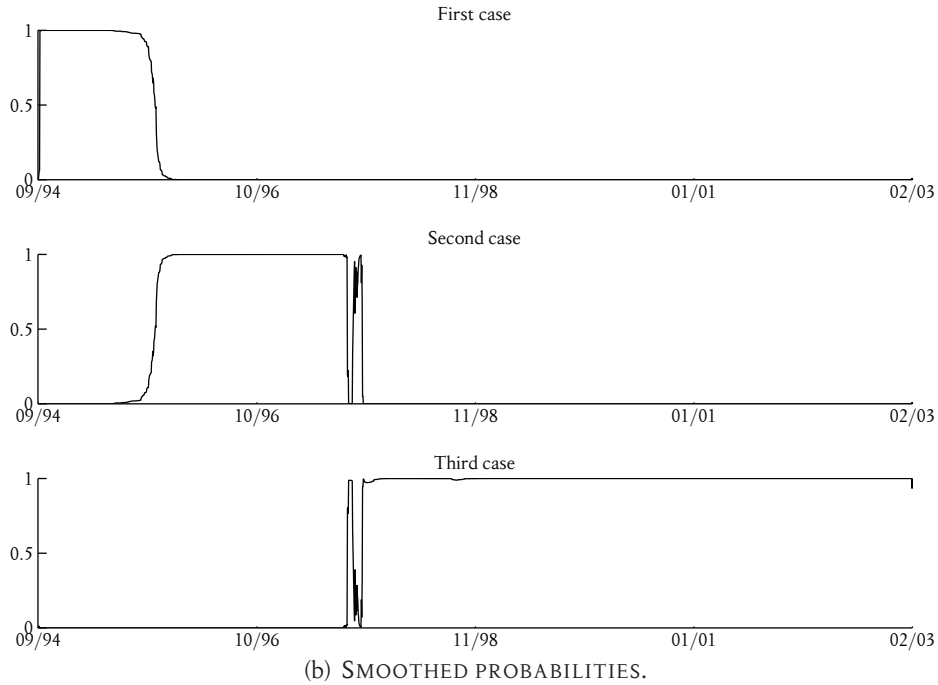
The figure 4.6(b) shows the smoothed probabilties of the regimes selected by the decision process. Contrary to deterministic decision tree, interpretation is far from obvious. This is due to the fact that deterministic tree attribute a model for each partition of the space. Because of the threshold classification, each partition is represented. In our case, we discriminate among various regime. That why the smoothed probabilities of the figure 4.6(b) represents the combination of regime that fit the data.

The model clearly identifies three combinations of regimes. The identification of the nature of regimes can be done using the table 4.2 of estimated paremeters. We first note that the 10-year bond futures serie does not seem to have a change of regime. The following interpretation of theses combinations are, by chronological order :

1. $\{h^2_{t,Y}, h^1_{t,P}, R^1_t\}$: normal volatility for bond, low volatility for S&P500 and positive correlations regime.
2. $\{h^2_{t,Y}, h^2_{t,P}, R^1_t\}$: normal volatility for bond, high volatility for S&P500 and positive correlations regime.
3. $\{h^2_{t,Y}, h^1_{t,P}, R^2_t\}$: normal volatility for bond, low volatility for S&P500 and negative correla-



(a) ESTIMATED CORRELATIONS.



(b) SMOOTHED PROBABILITIES.

FIG. 4.6: Estimated correlations and smoothed probabilities.

tions regime.

The results of the estimated transition probability matrices clearly shows that the first regime are persistent for the bond. Our model does not identify low and high volatility period. This is not the case of the index where the model indentifies two regimes. This is also the same for the correlations and the first part of the sample is clearly a regime of positive correlations while and the second part has negative correlations.

The estimated results of the coupling transition matrix gives us the relationship between the first

and the second level. This 2-by-2 transition matrix with constant probabilities is written as :

$$\mathrm{P}_{\text{coupl}} = \begin{bmatrix} c_{11} & 1-c_{22} \\ 1-c_{11} & c_{22} \end{bmatrix} \tag{4.23}$$

where the first state link the first state link the volatility level with the positive correlation regime and the second rely the first level with the negative correlation regime. Estimated parameters are :

$$\widehat{\mathrm{P}}_{\text{coupl}} = \begin{bmatrix} 0.2522 & 0.1371 \\ 0.7478 & 0.8629 \end{bmatrix} \tag{4.24}$$

This result means that in general, volatility is associated with negative correlations.

4.3.2 Deepening of the relationship between first and second level

In the later application, we use a very simple transition matrix to link the first and the second level. This specification has the advantage of being attractive from a computational point of vue but seems to be quite restrictive. Indeed the linking relation is common to all case of the first stage. In this subsection, we extend the later specification by introducing a specific relation for all the possible case à the first level. In that case, there are four two-by-two coupling matrix matrix P^i, $i = 1, ..., 4$. Each of these links a pair $\{h_{t,B}, h_{t,SP}\}$ with a case $\{R_t\}$. The decision of this extended case can be summerized as in figure 4.7.

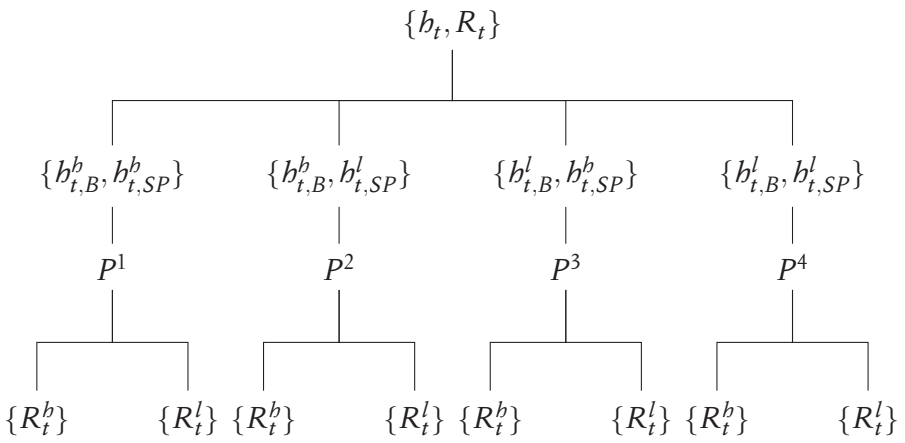
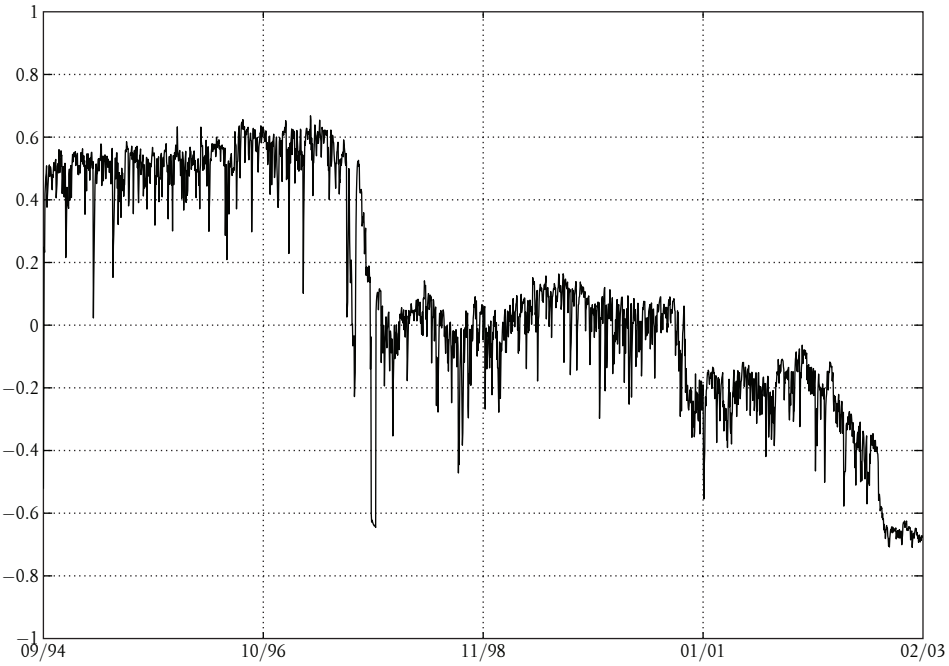


FIG. 4.7: Structure of the decision process for a bivariate dataset of the extended version.

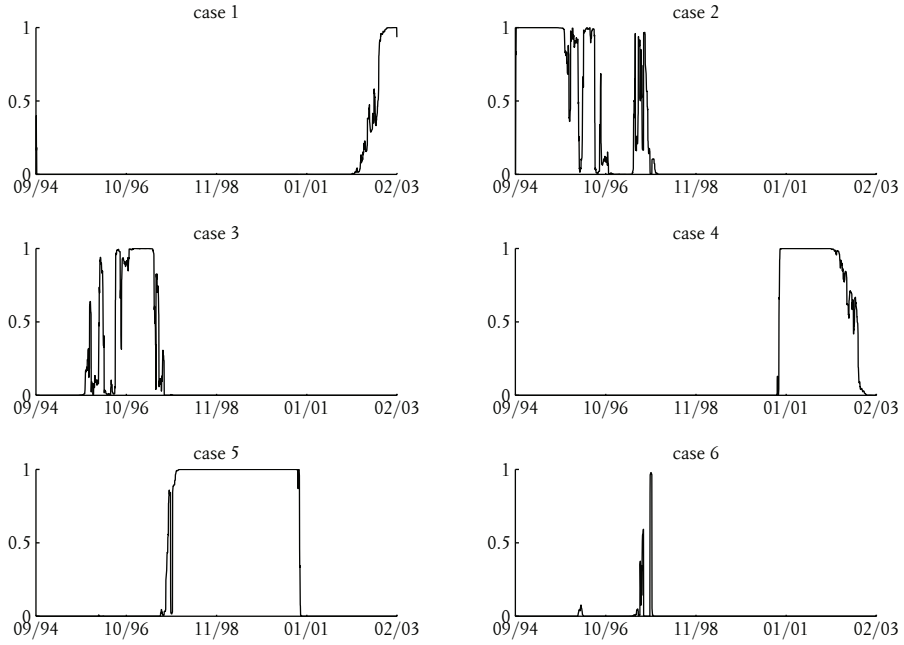
While the simple model needs only two parameters for the coupling matrix, the extended case needs eight parameters. Estimation is done using maximum likelihhod using a regular HMM representation. Estimated correlations are plotted in figure 4.8(a) and the smoothed probabilities of the relevant regimes are plotted in figure 4.8(b) ; estimated parameters are in table 4.3.

Comparing to the previous specification, the introduction of specific coupling matrix increases the explanatory power of the model. Estimation of the model exhibits six cases :

1. $\{h^1_{t,Y}, h^1_{t,P}, R^1_t\}$: low volatility for bond, low volatility for S&P500 and negative correlations regime.
2. $\{h^1_{t,Y}, h^1_{t,P}, R^2_t\}$: low volatility for bond, low volatility for S&P500 and positive correlations regime.



(a) ESTIMATED CORRELATIONS.



(b) SMOOTHED PROBABILITIES.

FIG. 4.8: Estimated correlations and smoothed probabilities.

3. $\{h^1_{t,Y}, h^2_{t,P}, R^2_t\}$: low volatility for bond, high volatility for S&P500 and positive correlations regime.
4. $\{h^2_{t,Y}, h^1_{t,P}, R^1_t\}$: high volatility for bond, low volatility for S&P500 and negative correlations regime.
5. $\{h^2_{t,Y}, h^1_{t,P}, R^2_t\}$: high volatility for bond, low volatility for S&P500 and positive correlations regime.
6. $\{h^2_{t,Y}, h^2_{t,P}, R^2_t\}$: high volatility for bond, high volatility for S&P500 and positive correlations.

First level

	10-year bond futures		S&P500	
Parameters	Estimate	standard errors	Estimate	standart errors
ω_1	0.0017	1.03e-06	0.0115	0.00015
α_1	0.0406	0.00029	0.0867	0.00129
β_1	0.9488	7.55e-05	0.8391	0.00933
ω_2	0.0137	6.35e-05	0.0831	0.00013
α_2	0.0706	0.00050	0.3582	1.11e-05
β_2	0.9195	0.00078	0.6417	1.06e-05
p_{11}	0.9987	0.00227	0.9999	1.02e-28
p_{22}	0.0010	6.272e-07	0.0001	1.70e-07

Coupling matrix

Parameters	Estimate	standart errors
p_{11}^1	0.7558	0,002485
p_{22}^1	0.09319	1,28e-05
p_{11}^2	0.7661	0,00066
p_{22}^2	0.11846	1,66e-05
p_{11}^3	0.2874	0,00045
p_{22}^3	0.0267	8,71e-07
p_{11}^4	0.5686	0,00026
p_{22}^4	1.00e-05	6,43e-08

Second level

Parameters	Estimate	standart errors
a_1	8.8334e-04	8.473e-06
b_1	0.9716	0.00086
const	$\begin{bmatrix} 0.0008 & -0.0031 \\ & 0.0127 \end{bmatrix}$	$\begin{bmatrix} 1.12e-05 & 0.00014 \\ & 0,00211 \end{bmatrix}$
a_2	0.0344	0.00044
b_2	0.4325	0.04093
const	$\begin{bmatrix} 0.0834 & 0.1107 \\ & 0.1836 \end{bmatrix}$	$\begin{bmatrix} 0,00111 & 0,01074 \\ & 0,11735 \end{bmatrix}$
p_{11}	0.8223	0.00144
p_{22}	0.9714	6.76e-06

TAB. 4.3: Model estimates.

While the previous specification only shows us that in general volatility is associated with negative correlations, our extended model gives a more precise information about the relasionship between volatilty and correlations. Given the estimated parameters of the four coupling matrix, we can deduce that there is a strong probability to have postive correlation when bond and S&P500 are in low volatility regime and when bond has low volatility and S&P500 high has high volatility. On the other hand, negative correlations seems to be associated with a high volatility of the bond. This particularity is more significative when the S&P500 is in low volatility regime.

4.4 Conclusion

The model we have presented in this paper has a natural and easy interpretation. Based on the Hidden Markov Decision Tree introduced by Jordan, Ghahramani, and Saul (1997), we develop a stochastic decision tree liking the dynamics of univariate volatility with the dynamics of the correlations. Unlike conventional approaches based on deterministic decision tree, our model allows a probabilistic point of view of the relationship between univariate volatility and correlations. We have restricted ourselves the complexity of the tree-structure. More complex architecture remains to be done. It also remains to develop estimation tools more suitable when the dimension of the problem increases.

CHAPITRE

5

Conclusions et perspectives

L'étude des propriétés statistiques observées sur les séries financières a engendré une littérature volumineuse. Plusieurs classes de modèles ont ainsi émergé au fil des années (GARCH, volatilité stochastique, multifractal) afin de modéliser l'évolution des rendements. Plus récemment, l'intérêt des chercheurs s'est porté sur l'étude de la volatilité des séries vectorielles de rendements afin de mieux comprendre les liens reliant diverses séries de rendements.

Les modèles GARCH multivariés à corrélations conditionnelles, développés avec l'article de Bollerslev (1990) et le modèle CCC, constituent un cadre d'analyse des variances et corrélations conditionnelles des séries financières. Cependant, si la formulation originale du CCC a constitué une avancée majeure dans la modélisation des covariances conditionnelles, son hypothèse de corrélations conditionnelles constantes dans le temps s'est avérée très réductrice empiriquement. L'introduction de corrélations conditionnelles variant dans le temps est proposée presque simultanément par Tse and Tsui (2002) et Engle and Sheppard (2001) (modèles DCC). Les modèles à corrélations conditionnelles dynamiques constituent désormais un axe de recherche très actif de l'économétrie financière.

Dans cette thèse, nous nous sommes focalisés sur les modèles DCC à changements de régime. Ces modèles sont un outil précieux pour les praticiens afin de comprendre la dynamique de séries financières présentant des ruptures structurelles. Plusieurs approches classiques ont déjà été proposées dans la littérature, basées sur un mécanisme de saut qui peut être déterministe ou stochastique.

Notre travail s'intéresse à l'écriture du mécanisme permettant de passer d'un régime à l'autre en se focalisant sur les mécanismes probabilistiques. L'approche classique en économétrie consiste à supposer que le mécanisme de passage d'un régime à un autre est déterminé selon une variable aléatoire cachée suivant une chaîne de Markov au premier ordre. Cette modélisation par modèle de Markov caché (HMM) a déjà été appliquée à diverses extensions du modèle DCC. Les contributions de cette thèse ont pour point commun d'être toutes basées sur des extensions du modèle HMM de base. Ses extensions ont été développées dans une autre discipline que l'économétrie financière : les modèles graphiques probabilistes (MGP). Les MGP sont une formulation hybride combinant la théorie des graphes et celle des probabilités dont les applications concernent la reconnaissance d'écriture, de la parole etc. Ainsi, nous proposons des modèles DCC à changements de régime dans lesquels nous avons substitué au modèle HMM classique des extensions issues du formalisme des MGP.

Dans le chapitre 2 nous proposons un cas particulier du modèle RSDC de Pelletier (2006) basé sur le modèle HMM hiérarchique de Fine, Singer, and Tishby (1998). Le modèle RSDC suppose que les corrélations évoluent entre des matrices de corrélations constantes dans le temps, dont le passage de l'une à l'autre s'effectue via une chaîne de Markov cachée. Dans notre modèle, la chaîne de Markov est remplacée par une structure hiérarchique Markovienne cachée. Ainsi, dans le HRSDC à deux régimes, les deux régimes principaux sont une combinaison de sous régimes. La matrice de corrélation conditionnelle évolue alors entre quatre matrices constantes dans le temps. Cette spécification apporte une meilleurs définition des régimes en augmentant leur granularité.

Le modèle du chapitre 3 utilise une autre extension du modèle HMM afin d'apporter davantage de flexibilité à l'approche classique. En effet, dans un modèle DCC avec changements de régime selon une chaîne de Markov cachée classique (comme le Markov-switching), tous les éléments de la matrice des corrélations conditionnelles ont la même dynamique de saut. Cette hypothèse peut s'avérer restrictive pour l'étude de séries financières en période de crise puisque cette spécification impose un changement de régime commun à toutes les corrélations. Pour relaxer cette hypothèse, nous utilisons le modèle HMM factorisé (FHMM) de Ghahramani and Jordan (1997) dans lequel la variable d'état est décomposée en plusieurs variables d'état. Il y a donc plusieurs chaînes de Markov en parallèle et l'état du modèle est déterminé par la combinaison de ces chaînes. A la différence du modèle HMM classique, cette approche permet d'attribuer une dynamique de saut à

chaque élément de la matrice des corrélations.
La contribution du chapitre 4 développe un arbre de décision afin de relier le niveau des volatilités individuelles aux corrélations. L'approche usuelle des arbres de décision est généralement construite dans un cadre déterministe. Les décisions sont prises de façon binaire et l'arbre se construit selon une procédure séquentielle. L'arbre de décision Markovien caché que nous utilisons a été proposé par Jordan, Ghahramani, and Saul (1997). Il se situe dans un cadre purement probabiliste et peut s'interprêter comme un HMM à la fois factorisé et couplé. Appliqué au modèle DCC de Engle and Sheppard (2001), le premier niveau de l'arbre discrimine entre un régime faible et fort de volatilité, tandis que le second niveau discrimine entre un régime faible et fort des corrélations.

Ainsi, les contributions de cette thèse se situent à l'interface de deux disciplines : l'économétrie financière et les modèles graphiques probabilistes. Les résultats obtenus montrent que les extensions HMM provenant des MGP peuvent s'appliquer facilement aux modèles à changements de régime utilisés en économie. Ils apportent par ailleurs un gain en terme de pouvoir explicatif pour le praticien. Cependant, l'introduction des extensions laisse beaucoup de questions sans réponses. Tout d'abord, au niveau théorique, les propriétés probabilistes des extensions du modèle HMM restent à confirmer. S'il existe en effet de nombreuses études dans le cadre du modèle standard HMM, il n'y a, à notre connaissance, pas d'analyses précises du comportement asymptotique des chaînes de Markov structurées. Cette lacune théorique rend également délicate la recherche des propriétés asymptotiques des modèles de type DCC construits à partir d'extensions de HMM.
Par ailleurs, dans un registre plus pratique, il apparaît nécessaire de développer des méthodes d'estimation adaptées à ces processus. Dans cette thèse, nous avons fait le choix de convertir les HMM structurés en HMM de base. Cette stratégie avait pour but de permettre l'utilisation d'outils classiques d'inférence et d'estimation de l'économétrie financière. Ainsi, les modèles présentés dans cette thèse s'estiment par maximum de vraisemblance et l'inférence sur l'état de la chaîne de Markov est réalisée avec le filtre d'Hamilton. Néanmoins la stratégie consistant à convertir une structure Markovienne structurée en écriture HMM équivalente présente une malédiction de la dimension. En effet, les représentations équivalentes HMM présentent alors généralement une matrice de transition de grande taille.

Au final, les résultats présentés dans cette thèse montrent que les extensions HMM provenant des modèles graphiques probabilistes permettent d'améliorer le mécanisme de changement de régime par rapport à l'approche Markov-switching classique. Si dans cette thèse nous nous sommes restreint à l'étude des corrélations de séries financières, ces extensions trouveront aussi leur utilité pour raffiner les approches de type Markov-switching VAR, mais aussi pour l'analyse de données hautes fréquences.

ANNEXE

A

A.1 An example of HHMM

Figure A.1 is an example of HHMM with four levels and six emitting states.

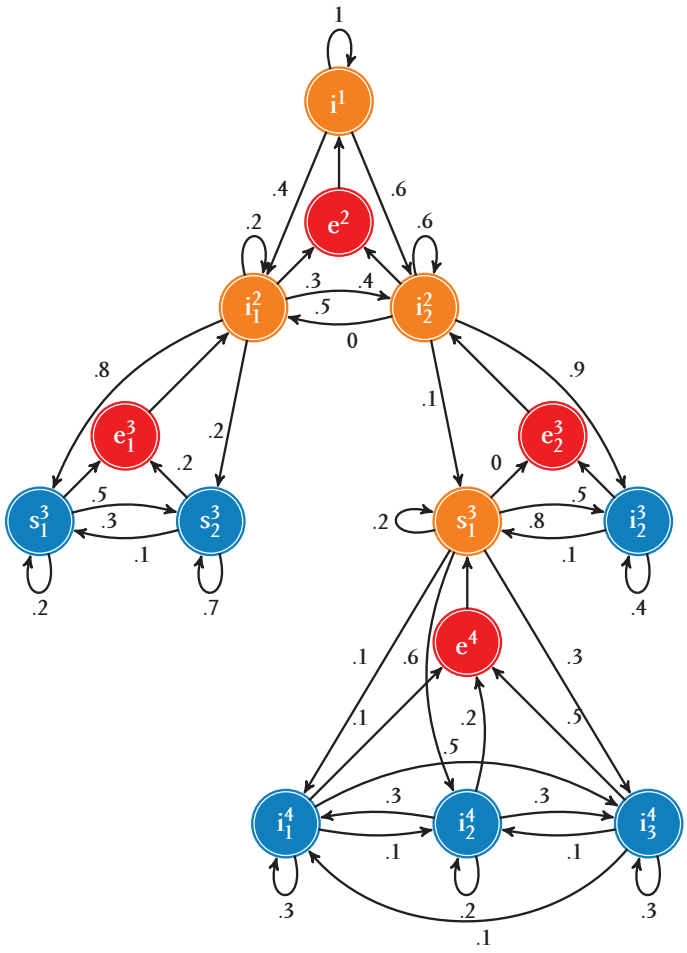


FIG. A.1: An example of Hierarchical Hidden structure.

A.2 EM algorithm

The origin of the Expectation-Maximization (EM) algorithm comes from the difficulties to estimate the parameters of models based on unobserved latent variables. More specially, it is an alternate approach face to usual iterative method to reach the optimum of an objective function. In the case of HMM/Markov-Switching models, the difficulties comes generally from two levels : (i) the likelihood function has numerous optimums (ii) optimizing the likelihood function is analytically intractable.

The firsts major contributions to elaborate this technique are due to Baum and T. (1966); Baum, T., Soules, and Weiss (1970). They developed the two steps methodology which consist to first compute the expectation of the likelihood function with the latent variables (E step) and then maximizing this expected likelihood (M step). From a theoretical point if view, the condition of convergence of the algorithm were proved by Dempster, Laird, and Rubin (1977) and improved few years later by Wu (1983). The literature on the subject was later expanded considerably and

many variants have emerged (see for example Meng and van Dyk (1997) and also Geoffrey and Thriyambakam (1997)).
More formally, the mechanic of the EM algorithm is to proceed by successive maximization based on observable data and information that we have every iterations on the latent variables. Instead of trying to directly calculate the conditional expectation $\mathbb{E}[\log f(\varepsilon, s|\theta)|\varepsilon, \theta]$, which is impossible because of s, the problem is being circumvented by focusing on the calculation of the quantity :

$$\mathscr{Q}(\theta|\theta_p) \overset{def}{=} \mathbb{E}_{\theta_p}[\log f(\varepsilon_{1:T}, s_{1:T}; \theta')|\varepsilon_{1:T}]$$

that it is calculable. Because the log-likelihood of the complete data is convex, it is possible to use Jensen inequality to deduce that, for all $(\theta', \theta) : \mathscr{Q}(\theta'|\theta) \geq \mathscr{Q}(\theta|\theta) \Longrightarrow \ell(\theta') \geq \ell(\theta)$. The two steps methodology for the p^{th} is then as follow :

$$\begin{aligned} &① \textit{ E-step} : \mathscr{Q}(\theta|\theta_p) = \mathbb{E}_{\theta_p}[\log f(\varepsilon, s; \theta')|\varepsilon] \\ &② \textit{ M-step} : \hat{\theta}_{p+1} = \arg\max_{\theta} \mathscr{Q}(\theta|\theta_p) \end{aligned}$$

By iterating ① and ②, the sequence $(\theta_p)_{p\in\mathbb{N}}$ ensures the convergence of the sequence $(\ell(\theta_p))_{p\in\mathbb{N}}$ to the true vector of parameters.

A.3 Scaling factor

Computing the quantities $\alpha_t(q)$ and $\beta_t(q)$ can lead to underflow problem. A easy way to circumvent this issue is to introduce a *scaling factor*, written as ς. In that case, initialization of Forward variable, written $\breve{\alpha}_t(q)$ is done by taking $\varsigma(1) = \sum_q \alpha_1(q)$:

$$\breve{\alpha}_1(q) = \frac{\alpha_1(q)}{\varsigma(1)}$$

We can then compute the *intermediary Forward variable*, written $\bar{\alpha}_{t+1}(q)$:

$$\bar{\alpha}_{t+1}(q) = f_q(\varepsilon_{t+1}) \sum_{q'} \sum_{d} \breve{\alpha}_t(q_t) \tilde{a}^d(q', q)$$

which lead to the scaling factor $\varsigma(t+1)$:

$$\varsigma(t+1) = \sum_q \bar{\alpha}_{t+1}(q)$$

The adjusted Forward variable can then be written as follow :

$$\breve{\alpha}_{t+1}(q) = \frac{\bar{\alpha}_{t+1}(q)}{\varsigma(t+1)}$$

Similar calculus can be apply to obtain the Backward variable. By initializate with $\breve{\beta}_T(q) = 1$, the intermediary quantity is expressed as :

$$\bar{\beta}_t(q) = \sum_{q_t} \sum_{d} \breve{\beta}_{t+1}(q') f_{q'}(\mathbf{y}_{t+1}) \tilde{a}^d(q, q')$$

and the adjusted Backward variable is obtained by computing for each $t = T-1, ..., 1$:

$$\breve{\beta}_t(q) = \frac{\bar{\beta}_t(q)}{\varsigma(t)}$$

The two quantities $\breve{\alpha}_t(q)$ and $\breve{\beta}_t(q)$ allow to rewriting the probabilities $\gamma_t^p(q)$ and $\xi_t^p(q', q, d)$ as :

$$\begin{aligned} \gamma_t^p(q) &= \breve{\alpha}_t(k)\breve{\beta}_t(k) \\ \xi_t^p(q', q, d) &= \breve{\alpha}_t(q')\tilde{a}^d(q', q) f_{q'}(\mathbf{y}_{t+1})\breve{\beta}_{t+1}(q) \end{aligned}$$

A.4 Exchange rate data

The exchange rate database is that of Harvey, Ruiz, and Shephard (1994) and was also used by Kim, Shephard, and Chib (1998) and Pelletier (2006). It is freely downloadable at the url :

`http://www.nuffield.ox.ac.uk/users/shephard/pub.aspx`

It contains week-days close exchange rates against US dollar for Pound, Deutschmark, Yen and Swiss-Franc over the period 1/10/81 to 28/6/85. This series are plotted on figure A.2.

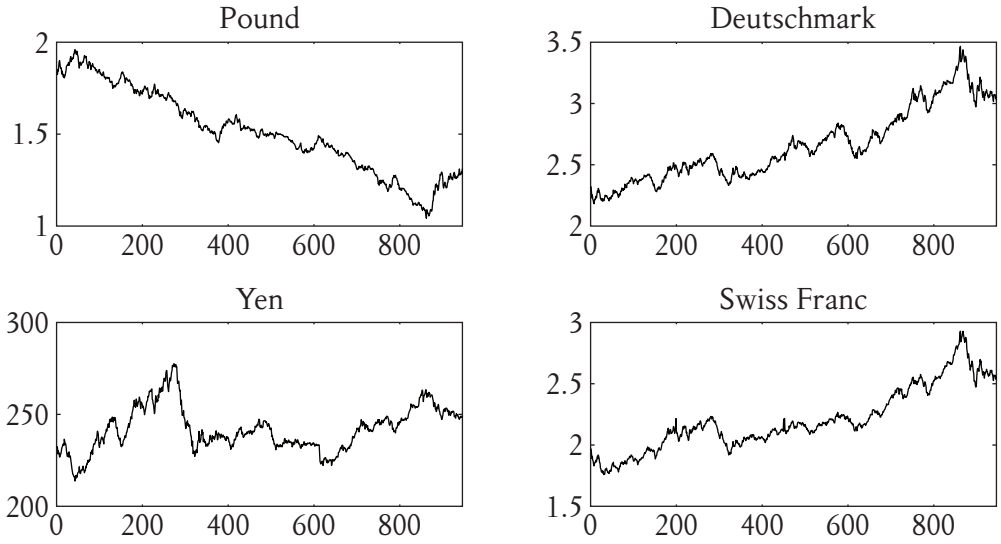


FIG. A.2: Exchange rate database.

A.5 S&P500 index futures and 10-year bond futures

This database is that of Colacito and Engle (2006) (plotted on figure A.3) and available at the url :

`http://www.unc.edu/~colacitr/Research/Files/EC2006.zip`

It contains daily returns of S&P500 index futures and 10-year bond futures from January 1990 to August 2003 (3912 observations).

A.6 Estimated parameters of the HRSDC for the real database applications

Table A.2 shows the parameters estimations of the correlations process for the database of exchange rate data ; table A.1 for the database of Engle and Colacito.

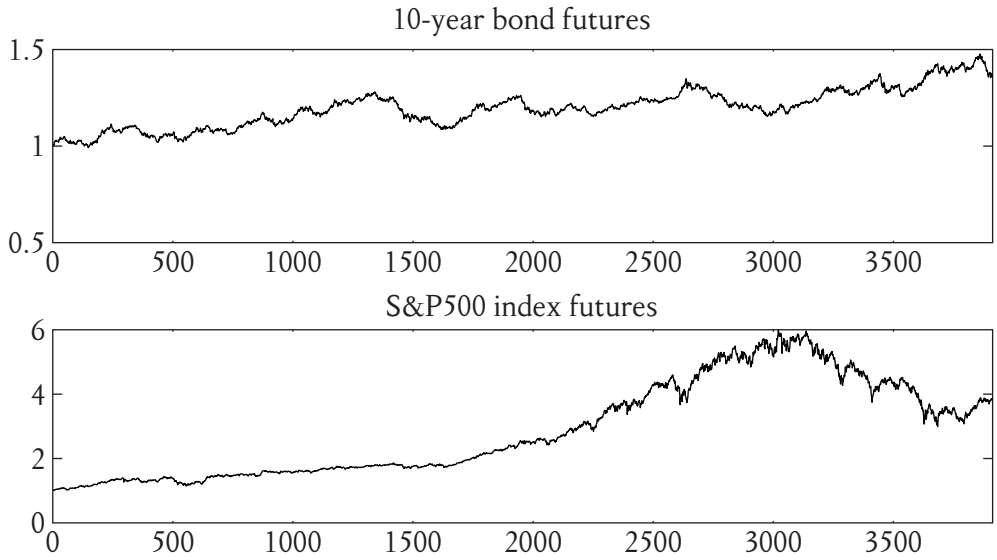


FIG. A.3: S&P500 index futures and 10-year bond futures.

correlation matrix :

$$R_1 = \begin{bmatrix} 1 & \\ 0,5189 & 1 \end{bmatrix} \qquad R_2 = \begin{bmatrix} 1 & \\ 0.2924 & 1 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 1 & \\ -0.1853 & 1 \end{bmatrix} \qquad R_4 = \begin{bmatrix} 1 & \\ -0.1597 & 1 \end{bmatrix}$$

transition probabilities

$$A^1 = \begin{bmatrix} 0.9797 & 0.9365 \\ 0.0203 & 0.0635 \end{bmatrix}$$

$$A_1^2 = \begin{bmatrix} 0.9919 & 1.4e^{-4} \\ 0.0055 & 0.7876 \end{bmatrix} \quad A_1^2 = \begin{bmatrix} 0.8139 & 0.0618 \\ 0.1853 & 0.9299 \end{bmatrix}$$

$$e_1^2 = 5.85e-4,\ e_2^2 = 0.2123,\ e_3^2 = 7.7e-4,\ e_4^2 = 0.0084$$

$$\pi_1^2 = 0.0067,\ \pi_2^2 = 0.9933,\ \pi_3^2 = 0.9933,\ \pi_4^2 = 0.0067$$

TAB. A.1: Estimated parameters for the correlations of the second real data application.

correlation matrix :

$$R_1 = \begin{bmatrix} 1 & & & \\ -0,4860 & 1 & & \\ -0,3943 & 0,8125 & 1 & \\ -0,5004 & 0,9488 & 0,8032 & 1 \end{bmatrix} \quad R_2 = \begin{bmatrix} 1 & & & \\ -4e-5 & 1 & & \\ -3e-5 & 8e-05 & 1 & \\ -5e-5 & 9e-5 & 8e-5 & 1 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 1 & & & \\ -0,8746 & 1 & & \\ -0,6912 & 0,7720 & 1 & \\ -0,8299 & 0,9214 & 0,7587 & 1 \end{bmatrix} \quad R_4 = \begin{bmatrix} 1 & & & \\ -0,7337 & 1 & & \\ -0,5798 & 0,6476 & 1 & \\ -0,6962 & 0,7730 & 0,6365 & 1 \end{bmatrix}$$

transition probabilities

$$A^1 = \begin{bmatrix} 0,9933 & 0,4167 \\ 0,0066 & 0,5832 \end{bmatrix}$$

$$A_1^2 = \begin{bmatrix} 0,6711 & 0,0084 \\ 0,0115 & 0,9915 \end{bmatrix} \quad A_1^2 = \begin{bmatrix} 0,9835 & 0,0002 \\ 0,0157 & 0,9571 \end{bmatrix}$$

$$e_1^2 = 0.3172,\ e_2^2 = 8.21e-05,\ e_3^2 = 7.37e-04,\ e_4^2 = 0.0425$$

$$\pi_1^2 = 0.9933,\ \pi_2^2 = 0.0066,\ \pi_3^2 = 0.9933,\ \pi_4^2 = 0.0067$$

TAB. A.2: Estimated parameters for the correlations of the second real data application.

ANNEXE

B

B.1 Stability of stochastic matrix under kronecker product

A matrix $A = \big((a_{ij})\big)_{1 \leqslant i,j \leqslant n}$ of $M_n(\mathbb{R})$ with a_{ij} the i^{th} row and j^{th} column element is said stochastic if $\forall (i,j) \in [0,1]^2$:

$$a_{ij} \in (0,1) \tag{B.1}$$

and $\forall i \in \{1,...,n\}$:

$$\sum_{i=1}^{n} a_{ij} = 1 \tag{B.2}$$

Let P_n the set of stochastic matrix of order n. To show that P_n is stable under kronecker product, we consider two stochastic matrix $A = \big((a_{ij})$ and $B = \big((b_{ij})$ with $1 \leqslant i,j \leqslant n$. Let C such that :

$$C = A \otimes B \tag{B.3}$$

Then by definition, the kronecker product, denoting by $\otimes$, of A and B is equal to :

$$C = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{11}b_{1n} & a_{12}b_{1n} & \cdots & a_{1n}b_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{n1} & \cdots & a_{11}b_{nn} & a_{12}b_{nn} & \cdots & a_{1n}b_{nn} \\ a_{21}b_{11} & \cdots & a_{21}b_{1n} & a_{22}b_{1n} & \cdots & a_{2n}b_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1}b_{n1} & \cdots & a_{n1}b_{nn} & a_{n2}b_{nn} & \cdots & a_{nn}b_{nn} \end{bmatrix} \tag{B.4}$$

which is a matrix of size $n^n \times n^n$. Since a_{ij} and b_{ij} are in $[0,1]$, we have for each $0 \leqslant k,l \leqslant n^n$:

$$0 \leqslant c_{kl} \leqslant 1 \tag{B.5}$$

If we calculate the sum of the first row, we have :

$$\sum_{k=1}^{n^n} c_{k1} = \sum_{j=1}^{n} \big(a_{j1} \sum_{i=1}^{n} b_{i1}\big) \tag{B.6}$$

$$= \sum_{j=1}^{n} a_{j1} \times 1 \tag{B.7}$$

$$= 1 \tag{B.8}$$

and similarly for the following columns. Thus C is indeed a stochastic matrix. This intersting result allows to have a representation of any factorial hidden Markov model with M parallel chains with N states into a regular hidden Markov model with a transition matrix such $\Upsilon = \bigotimes_{i=1}^{M} \mathrm{P^i}$.

B.2 Smoothed probabilities

Apply Kim's filter to the regular representation Markov-switching representation is an easy way to compute smoothed probabilities. But one drawback is they refer to a combination of state since the transition matrix is the kronecker product of the transition matrix of the parallel chains. One element of the resulting smoothing probabilities vector can then de decomposed as :

$$\hat{\xi}_{t|T} = \mathbb{P}[\upsilon_t = j | \mathscr{F}_T, \theta] \tag{B.9}$$

$$= \mathbb{P}[s_t^1 = j | \mathscr{F}_T, \theta] \times ... \times \mathbb{P}[s_t^M = j | \mathscr{F}_T, \theta] \tag{B.10}$$

Making a correspondance between the smoothed probabilities of independant chains and regular representation is simple using basic linear algebra. Regimes of each chains are represented by an identity matrix $\mathbf{I}_N$ of size N, where each diagonal element represents a regime. To have an make the computation of the smoothed probabilities of the i^{th} regime of the chain m, this chains is represented by a matrix of zeros except the except the diagonal element of the i^{th} regime whose value is one. Then, the resulting smoothing probabilites are computed as :

$$\mathbb{P}[s_t^1 = j|\mathscr{F}_T, \theta] = \mathbf{Tr}\Big[\big[[\mathbf{I}_N^1 \otimes ... \otimes \mathbf{I}_N^m \otimes ... \otimes \mathbf{I}_N^M]\hat{\xi}_{t|T}\big]\mathtt{1}_{M^N}\Big] \tag{B.11}$$

where $\mathbf{Tr}[\cdot]$ is the trace operator and $\mathtt{1}_{M^N}$ a vector of ones with length M^N.
As an exemple, we consider a three chains FHM containing two regimes each. The transition matrix for each chain are expressed as :

$$\mathrm{P^a} = \begin{bmatrix} a11 & a21 \\ a12 & a22 \end{bmatrix}, \mathrm{P^b} \begin{bmatrix} b11 & b21 \\ b12 & b22 \end{bmatrix}, \mathrm{P^c} = \begin{bmatrix} c11 & c21 \\ c12 & c22 \end{bmatrix} \tag{B.12}$$

The transition matrix of the regular Markov-switching formulation is then of size $2^3 \times 2^3$ and is expressed as :

$$\Upsilon = \mathrm{P^a} \otimes \mathrm{P^b} \otimes \mathrm{P^c} \tag{B.13}$$

which gives :

$$\Upsilon = \begin{bmatrix} a11b11c11 & a11b11c21 & a11b21c11 & a11b21c21 & a21b11c11 & a21b11c21 & a21b21c11 & a21b21c21 \\ a11b11c12 & a11b11c22 & a11b21c12 & a11b21c22 & a21b11c12 & a21b11c22 & a21b21c12 & a21b21c22 \\ a11b12c11 & a11b12c21 & a11b22c11 & a11b22c21 & a21b12c11 & a21b12c21 & a21b22c11 & a21b22c21 \\ a11b12c12 & a11b12c22 & a11b22c12 & a11b22c22 & a21b12c12 & a21b12c22 & a21b22c12 & a21b22c22 \\ a12b11c11 & a12b11c21 & a12b21c11 & a12b21c21 & a22b11c11 & a22b11c21 & a22b21c11 & a22b21c21 \\ a12b11c12 & a12b11c22 & a12b21c12 & a12b21c22 & a22b11c12 & a22b11c22 & a22b21c12 & a22b21c22 \\ a12b12c11 & a12b12c21 & a12b22c11 & a12b22c21 & a22b12c11 & a22b12c21 & a22b22c11 & a22b22c21 \\ a12b12c12 & a12b12c22 & a12b22c12 & a12b22c22 & a22b12c12 & a22b12c22 & a22b22c12 & a22b22c22 \end{bmatrix}$$

This a transition matrix with eight regimes. Thus the resulting smoothed probabilities using Kim's filter, defining by equation (3.33), can be decomposed as :

$$\begin{bmatrix} \mathbb{P}[\upsilon_t = 1|\mathscr{F}_T, \theta] \\ \mathbb{P}[\upsilon_t = 2|\mathscr{F}_T, \theta] \\ \mathbb{P}[\upsilon_t = 3|\mathscr{F}_T, \theta] \\ \mathbb{P}[\upsilon_t = 4|\mathscr{F}_T, \theta] \\ \mathbb{P}[\upsilon_t = 5|\mathscr{F}_T, \theta] \\ \mathbb{P}[\upsilon_t = 6|\mathscr{F}_T, \theta] \\ \mathbb{P}[\upsilon_t = 7|\mathscr{F}_T, \theta] \\ \mathbb{P}[\upsilon_t = 8|\mathscr{F}_T, \theta] \end{bmatrix} = \begin{bmatrix} \mathbb{P}[s_t^a = 1, s_t^b = 1, s_t^c = 1|\mathscr{F}_T, \theta] \\ \mathbb{P}[s_t^a = 1, s_t^b = 1, s_t^c = 2|\mathscr{F}_T, \theta] \\ \mathbb{P}[s_t^a = 1, s_t^b = 2, s_t^c = 1|\mathscr{F}_T, \theta] \\ \mathbb{P}[s_t^a = 1, s_t^b = 2, s_t^c = 2|\mathscr{F}_T, \theta] \\ \mathbb{P}[s_t^a = 2, s_t^b = 1, s_t^c = 1|\mathscr{F}_T, \theta] \\ \mathbb{P}[s_t^a = 2, s_t^b = 1, s_t^c = 2|\mathscr{F}_T, \theta] \\ \mathbb{P}[s_t^a = 2, s_t^b = 2, s_t^c = 1|\mathscr{F}_T, \theta] \\ \mathbb{P}[s_t^a = 2, s_t^b = 2, s_t^c = 2|\mathscr{F}_T, \theta] \end{bmatrix} \tag{B.14}$$

where $\mathbb{P}[\upsilon_t = i|\mathscr{F}_T, \theta]$, $i = 1, ..., 8$, denotes the smoothed probability of the of thje regular MS representation. To have the smoothed probability of the second regime of the second chain, we first write the identy matrix for each chains using the rule explained in the last paragraph :

$$\mathbf{I}_2^a = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{I}_2^b \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{I}_2^c = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{B.15}$$

Using equation (B.11), we then have :

$$\mathbb{P}[s_t^b = 2|\mathscr{F}_T, \theta] = \mathbf{Tr}\Big[\big[[\mathbf{I}_2^a \otimes \mathbf{I}_2^b \otimes \mathbf{I}_2^c]\hat{\xi}_{t|T}\big]\mathtt{1}_8\Big] \tag{B.16}$$

With :

$$\mathbf{I}_2^a \otimes \mathbf{I}_2^b \otimes \mathbf{I}_2^c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{B.17}$$

it follows that :

$$\begin{aligned} \mathbb{P}[s_t^b = 2|\mathscr{F}_T, \theta] = \mathbb{P}[s_t^a = 1, s_t^b = 2, s_t^c = 1|\mathscr{F}_T, \theta] + \mathbb{P}[s_t^a = 1, s_t^b = 2, s_t^c = 2|\mathscr{F}_T, \theta] \\ + \mathbb{P}[s_t^a = 2, s_t^b = 2, s_t^c = 1|\mathscr{F}_T, \theta] + \mathbb{P}[s_t^a = 2, s_t^b = 2, s_t^c = 2|\mathscr{F}_T, \theta] \end{aligned} \tag{B.18}$$

B.3 Constant matrix construction

We first define the matrix of the two extrem cases, i.e. when all covariances are in regime 1 :

$$\overline{Q}_{s_t^1=1,s_t^2=1,s_t^3=1} = \begin{bmatrix} q_{11,s_t^1=1} & q_{12,s_t^1=1,s_t^2=1} & q_{13,s_t^1=1,s_t^3=1} \\ & q_{22,s_t^2=1} & q_{23,s_t^2=1,s_t^3=1} \\ & & q_{33,s_t^3=1} \end{bmatrix} \tag{B.19}$$

and the intercept corresponding at the case where all the covariances are in regime 2 :

$$\overline{Q}_{s_t^1=2,s_t^2=2,s_t^3=2} = \begin{bmatrix} q_{11,s_t^1=2} & q_{12,s_t^1=2,s_t^2=2} & q_{13,s_t^1=2,s_t^3=2} \\ & q_{22,s_t^2=2} & q_{23,s_t^2=2,s_t^3=2} \\ & & q_{33,s_t^3=2} \end{bmatrix} \tag{B.20}$$

In the next matrix, non-boxed elements are coming from the constants of the two extrem cases while boxed elements corresponds to the additional elements needed to define the intercept in that case. The six others intercepts can be expressed as :

$$\overline{Q}_{s_t^1=1,s_t^2=1,s_t^3=2} = \begin{bmatrix} q_{11,s_t^1=1} & q_{12,s_t^1=2,s_t^2=1} & \boxed{q_{13,s_t^1=2,s_t^3=2}} \\ & q_{22,s_t^2=1} & \boxed{q_{23,s_t^2=2,s_t^3=2}} \\ & & q_{33,s_t^3=2} \end{bmatrix} \tag{B.21}$$

$$\overline{Q}_{s_t^1=1,s_t^2=2,s_t^3=1} = \begin{bmatrix} q_{11,s_t^1=1} & \boxed{q_{12,s_t^1=1,s_t^2=2}} & q_{13,s_t^1=1,s_t^3=1} \\ & q_{22,s_t^2=2} & \boxed{q_{23,s_t^2=2,s_t^3=1}} \\ & & q_{33,s_t^3=1} \end{bmatrix} \tag{B.22}$$

$$\overline{Q}_{s_t^1=1,s_t^2=2,s_t^3=2} = \begin{bmatrix} q_{11,s_t^1=1} & \boxed{q_{12,s_t^1=2,s_t^2=1}} & \boxed{q_{13,s_t^1=1,s_t^3=2}} \\ & q_{22,s_t^2=2} & q_{23,s_t^2=2,s_t^3=2} \\ & & q_{33,s_t^3=2} \end{bmatrix} \tag{B.23}$$

$$\overline{Q}_{s_t^1=2,s_t^2=1,s_t^3=1} = \begin{bmatrix} q_{11,s_t^1=2} & \boxed{q_{12,s_t^1=2,s_t^2=1}} & \boxed{q_{13,s_t^1=2,s_t^3=1}} \\ & q_{22,s_t^2=1} & q_{23,s_t^2=1,s_t^3=1} \\ & & q_{33,s_t^3=1} \end{bmatrix} \tag{B.24}$$

$$\overline{Q}_{s_t^1=2,s_t^2=1,s_t^3=2} = \begin{bmatrix} q_{11,s_t^1=2} & \boxed{q_{12,s_t^1=2,s_t^2=1}} & q_{13,s_t^1=2,s_t^3=2} \\ & q_{22,s_t^2=1} & \boxed{q_{23,s_t^2=1,s_t^3=2}} \\ & & q_{33,s_t^3=2} \end{bmatrix} \tag{B.25}$$

$$\overline{Q}_{s_t^1=2,s_t^2=2,s_t^3=1} = \begin{bmatrix} q_{11,s_t^1=2} & q_{12,s_t^1=2,s_t^2=2} & \boxed{q_{13,s_t^2=1,s_t^3=1}} \\ & q_{22,s_t^2=2} & \boxed{q_{23,s_t^2=2,s_t^3=1}} \\ & & q_{33,s_t^3=1} \end{bmatrix} \tag{B.26}$$

Bibliographie

ALEXANDER, C. (2002) : "Principal Component Models for Generating Large GARCH Covariance Matrices," *Economic Notes*, 31, 337–359.

ANG, A., AND J. CHEN (2002) : "Asymmetric correlations of equity portfolios," *Journal of Financial Economics*, 63(3), 443–494.

AUDRINO, F. (2006) : "The impact of general non-parametric volatility functions in multivariate GARCH models," *Computational Statistics & Data Analysis*, 50(11), 3032–3052.

AUDRINO, F. T., AND P. BUHLMANN (2001) : "Tree-structured GARCH models," *Journal of the Royal Statistical Society, Series B*, 63(4), 727–744.

AUDRINO, F. T., AND A. TROJANI (2006) : "Estimating and Predicting Multivariate Volatility Thresholds in Global Stock Markets," *Journal of Applied Econometrics*, 21(3), 345–369.

BACHELIER, L. (1900) : "Théorie de la spéculation," *Annales scientifiques de l'Ecole Normale Supérieure*, 17, 21–86.

BARONE-ADESI, G., AND F. AUDRINO (2006) : "Average conditional correlation and tree structures for multivariate GARCH models," *Journal of Forecasting*, 25(8), 579–600.

BAUM, L., AND P. T. (1966) : "Inference for Probabilistic Functions of Finite State Markov Chains," *The Annals of Mathematical Statistics*, 37, 1554–1563.

BAUM, L., P. T., G. SOULES, AND N. WEISS (1970) : "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical Statistics*, 41, 164–171.

BAUWENS, L., S. LAURENT, AND J. ROMBOUTS (2006) : "Multivariate GARCH Models : a Survey," *Journal of Applied Econometrics*, 21(1), 79–109.

BERBEN, R., AND W. JANSEN (2005) : "Comovement in International Equity Markets : a Sectoral View," *Journal of International Money and Finance*, 24, 832–857.

BILLINGSLEY, P. (1986) : *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics. Wiley–Interscience, New–York, 2nd edn.

BILLIO, M., AND M. CAPORIN (2005) : "Multivariate Markov switching dynamic conditional correlation GARCH representations for contagion analysis.," *Statistical Methods and Applications*, 14(2), 145–161.

BILLIO, M., AND M. CAPORIN (2006) : "A Generalized Dynamic Conditional Correlation model for Portfolio Risk Evaluation," Working Paper No. 53/WP/2006, Department of Economics, Ca Foscari University of Venice.

BILLIO, M., AND M. CAPORIN (2009) : "A generalized Dynamic Conditional Correlation model for portfolio risk evaluation," *Mathematics and Computers in Simulation*, 79(8).

BILLIO, M., M. CAPORIN, AND M. GOBBO (2003) : "Block Dynamic Conditional Correlation Multivariate GARCH models," Working Paper 03.03, GRETA.

——— (2006) : "Flexible Dynamic Conditional Correlation multivariate GARCH models for asset allocation," *Applied Financial Economics Letters*, 2(2), 123–130.

BISHOP, C. M. (2006) : *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

BLACK, F. (1976) : "Studies of Stock Price Volatility Changes," in *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

BÖHNING, D., W. SEIDEL, M. ALFO, B. GAREL, V. PATILEA, AND G. WALTHER (2007) : "Advances in Mixture Models," *Computational Statistics & Data Analysis*, 51, 5205–5210.

BOLLERSLEV, T. (1986) : "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), 307–327.

BOLLERSLEV, T. (1990) : "Modelling the Coherence in Short-run Nominal Exchange Rates : A Multivariate Generalized ARCH model," *Review of Economics and Statistics*, 72, 498–505.

BOLLERSLEV, T. (2008) : "Glossary to ARCH (GARCH)," CREATES Research Papers 2008-49, School of Economics and Management, University of Aarhus.

BOLLERSLEV, T., R. F. ENGLE, AND J. M. WOOLDRIDGE (1988) : "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96(1), 116–31.

BRAND, M. (1997) : "Coupled hidden Markov models for modeling interacting processes," Technical report 405, MIT Media Lab Perceptual Computing.

BREIMAN, L., J. FRIEDMAN, R. OLSHEN, AND C. STONE (1984) : *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.

CAI, J. (1994) : "A Markov Model of Unconditional Variance in ARCH," *Journal of Business and Economic Statistics*, 12(3), 309–316.

CAJIGAS, M., AND G. URGA (2006) : "Dynamic Conditional Correlation Models with Asymmetric Multivariate Laplace Innovations," Discussion paper, Centre for Econometric Analysis, Cass Business School.

CALVET, L., AND A. FISHER (2001) : "Forecasting multifractal volatility," *Journal of Econometrics*, 105(1), 27–58.

——— (2002) : "Multifractality In Asset Returns : Theory And Evidence," *The Review of Economics and Statistics*, 84(3), 381–406.

——— (2008) : *Multifractal Volatility : Theory, Forecasting, and Pricing*. Academic Press.

CALVET, L., A. FISHER, AND B. MANDELBROT (1997) : "A Multifractal Model of Asset Returns," Cowles Foundation Discussion Papers 1164, Cowles Foundation, Yale University.

CAMPBELL, R., K. KOEDIJK, AND P. KOFMAN (2002) : "Increased Correlation in Bear Markets," *Financial Analysts Journal*, 58(1), 87–94.

CAPORIN, M., AND P. PARUOLO (2009) : "Structured Multivariate Volatility Models," "Marco Fanno" Working Papers 0091, Dipartimento di Scienze Economiche "Marco Fanno".

CAPPÉ, O., E. MOULINES, AND T. RYDEN (2005) : *Inference in Hidden Markov Models*. Springer–Verlag, Berlin.

CAPPIELLO, L., R. F. ENGLE, AND K. SHEPPARD (2006) : "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," *Journal of Financial Econometrics*, 4(4), 537–572.

CARVALHO, A. X., AND M. A. TANNER (2006) : "Modeling nonlinearities with mixtures-of-experts of time series models," *International journal of mathematics and mathematical sciences*, 9, 1–22.

COLACITO, R., AND R. ENGLE (2006) : "Testing and Valuing Dynamic Correlations for Asset Allocation," *Journal of Business and Economic Statistics*, 24, 238–253.

CONT, R. (2001) : "Empirical properties of asset returns : stylized facts and statistical issues," *Quantitative Finance*, 1, 223–236.

DELLAPORTAS, P., AND I. D. VRONTOS (2007) : "Modelling volatility asymmetries : a Bayesian analysis of a class of tree structured multivariate GARCH models," *Econometrics Journal*, 10(3), 503–520.

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977) : "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society Series B*, 39(1), 1–38.

DIEBOLD, F., J. LEE, AND G. WEINBACH (1994) : "Regime Switching with Time-Varying Transition Probabilities," in *Non-Stationary Time Series Analysis and Co-Integration*, ed. by C. Hargreaves. The Clarendon Press & Oxford University Press, advanced texts in econometrics edn.

DUEKER, M. J. (1997) : "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility," *Journal of Business & Economic Statistics*, 15(1), 26–34.

ENGLE, R. (1995) : *ARCH : Selected Readings*. Oxford University Press, Oxford.

——— (2002) : "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

——— (2008) : "High dimensional dynamic correlations," in *The Methodology and Practice of Econometrics : Papers in Honour of David F Hendry*, ed. by J. L. Castle, and N. Shephard. Oxford University Press.

——— (2009) : *Anticipating correlations : a new paradigm for risk management*, The econometrics institute lectures. Princeton University Press.

ENGLE, R., AND B. KELLY (2009) : "Dynamic Equicorrelation," Working Paper FIN-08-038, New York University.

ENGLE, R., AND S. MANGANELLI (2004) : "CAViaR : Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, 22, 367–381.

ENGLE, R., AND J. MEZRICH (1996) : "GARCH for Groups," *Risk*, 9(8), 36–40.

ENGLE, R., AND K. SHEPPARD (2001) : "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," Working Paper 2001–15, UCSD.

ENGLE, R. F. (1982) : "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

ENGLE, R. F., AND K. F. KRONER (1995) : "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11(1), 122–150.

ENGLE, R. F., V. K. NG, AND M. ROTHSCHILD (1990) : "Asset pricing with a factor-arch covariance structure : Empirical estimates for treasury bills," *Journal of Econometrics*, 45(1–2), 213–237.

ENGLE, R. F., N. SHEPHARD, AND K. SHEPPARD (2008) : "Fitting vast dimensional time-varying covariance models," Discussion paper.

FAMA, E. F. (1965) : "The bahavior of stock market prices," *Journal of business*, 38, 34–105.

FENG, Y. (2007) : "A Local Dynamic Conditional Correlation Model," Working Paper 1592, MPRA.

FINE, S., Y. SINGER, AND N. TISHBY (1998) : "The Hierarchical Hidden Markov Model : Analysis and Applications," *Machine Learning*, 32(1), 41–62.

FRANCQ, C., AND J. ZAKOIAN (2009) : *Modèles GARCH*. Economica, Paris.

FRANQ, C., M. ROUSSIGNIL, AND J. ZAKOIAN (2001) : "Conditional Heteroskedasticity Driven by Hidden Markov Chains," *Journal of Time Series Analysis*, 22(2), 197–220.

FRANSES, P., C. HAFNER, AND D. VAN DIJK (2005) : "Semi–Parametric Modelling of Correlation Dynamics," *Advances in Econometrics*, 20, 53–103.

FRANSES, P., AND D. VAN DIJK (2004) : *Non–linear Time Series Models in Empirical Finance*. Cambridge University Press.

FRENCH, K. R. (1980) : "Stock returns and the weekend effect," *Journal of Financial Economics*, 8(1), 55–69.

FRIEDMAN, J. (1991) : "Multivariate Adaptive Regression Splines," *Annals of Statistics*, 19(1), 181–214.

GALEANO, P., AND M. C. AUSÍN (2009) : "The Gaussian Mixture Dynamic Conditional Correlation Model : Parameter Estimation, Value at Risk Calculation, and Portfolio Selection," *Journal of Business and Economic Statistics*, forthcomming.

GARCÍA-FERRER, A., E. GONZÁLEZ-PRIETO, AND D. PEÑA (2008) : "A multivariate generalized independent factor GARCH model with an application to financial stock returns," Statistics and Econometrics Working Papers ws087528, Universidad Carlos III, Departamento de Estadística y Econometría.

GEOFFREY, J., AND K. THRIYAMBAKAM (1997) : *The EM Algorithm and Extensions*, Wiley Series in Probability and Mathematical Statistics. Wiley–Interscience, New–York.

GHAHRAMANI, Z. (1998) : "Learning Dynamic Bayesian Networks," in *Adaptive Processing of Sequences and Data Structures, International Summer School on Neural Networks, "E.R. Caianiello"-Tutorial Lectures*, pp. 168–197, London, UK. Springer-Verlag.

GHAHRAMANI, Z., AND M. I. JORDAN (1997) : "Factorial hidden Markov models," *Machine Learning*, 29, 245–273.

GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1993) : "Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

GONZALES-RIVIERA, G. (1998) : "Smooth Transition GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 3, 61–78.

GRAY, S. (1996) : "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

GREEN, P. (1995) : "Reversible Jump MCMC Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

HAAS, M., AND S. MITTNIK (2008) : "Multivariate Regime-Switching GARCH with an Application to International Stock Markets," CFS Working Paper Series 2008-08, Center for Financial Studies.

HAAS, M., S. MITTNIK, AND M. PAOLLELA (2004a) : "Mixed Normal Conditional Heteroskedasticity," *Journal of Financial Econometrics*, 2(2), 211–250.

——— (2004b) : "A New Approach to Markov-Switching GARCH Models," *Journal of Financial Econometrics*, 2(4), 493–530.

HAFNER, C., AND P. FRANSES (2003) : "A generalized dynamic conditional correlation model for many asset returns," Discussion paper.

HAFNER, C., AND P. H. FRANSES (2009) : "A Generalized Dynamic Conditional Correlation Model : Simulation and Application to Many Assets," *Econometric Reviews*, 28(6), 612–631.

HAFNER, C. M., AND J. V. ROMBOUTS (2007) : "Semiparametric Multivariate Volatility Models," *Econometric Theory*, 23(02), 251–280.

HAGERUD, G. (1996) : "A Smooth Transition ARCH Model for Asset Returns," Working Paper 162, Stockholm School of Economics.

HAMILTON, J. (1989) : "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

——— (1990) : "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39–70.

——— (1994) : *Time Series Analysis*. Princeton University Press.

HAMILTON, J., AND R. SUSMEL (1994) : "Autoregressive Conditional Heteroskedasticity and Change in Regime," *Journal of Econometrics*, 64, 307–333.

HARVEY, C., E. RUIZ, AND N. SHEPHARD (1994) : "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

HUERTA, G., W. JIANG, AND M. A. TANNER (2003) : "Time series modeling via hierarchical mixtures," *Statistica Sinica*, 13(4), 1097–1118.

HUNG, H., D. PHUNG, AND S. VENKATESH (2004) : "Learning Hierarchical Hidden Markov Models with General State Hierarchy," in *Nineteenth National Conference on Artificial Intelligence, San Jose, California, USA*, ed. by A. Press, pp. 324–329.

JACOBS, R. A., M. I. JORDAN, S. J. NOWLAN, AND G. E. HINTON (1991) : "Adaptive mixtures of local experts," *Neural Computation*, 3, 79–87.

JORDAN, M., F. GHAHRAMANI, AND L. SAUL (1997) : "Hidden Markov decision trees," in *Advances in Neural Information Processing Systems (NIPS) 9*, ed. by M. C. Mozer, M. I. Jordan, and T. Petsche. Cambridge MA : MIT Press.

JORDAN, M. I. (ed.) (1999) : *Learning in graphical models*. MIT Press, Cambridge, MA, USA.

JORDAN, M. I., AND R. A. JACOBS (1994) : "Hierarchical Mixtures of Experts and the EM Algorithm," *Neural Computation*, 6(2), 181–214.

KAWAKATSU, H. (2003) : "Cholesky factor GARCH," Discussion paper, Queen's University, Belfast., mimeo.

KEIM, D. B. (1983) : "Size-related anomalies and stock return seasonality : Further empirical evidence," *Journal of Financial Economics*, 12(1), 13–32.

KIM, C. (1994) : "Dynamic Linear Models with Markov-Swithcing," *Journal of Econometrics*, 60, 1–22.

KIM, S., N. SHEPHARD, AND S. CHIB (1998) : "Stochastic Volatility : Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65(3), 361–93.

KLAASSEN, F. (2002) : "Improving GARCH Volatility Forecasts with Regime-Switching GARCH," *Empirical Economics*, 27(2), 363–394.

KOUTMOS, G. (1998) : "Asymmetries in the Conditional Mean and the Conditional Variance : Evidence from Nine Stock Markets," *Journal of Economics and Business*, 50(3), 277–290.

KROLZIG, H. (1997) : *Markov–switching vector autoregressions*vol. 454. Lecture notes in economics and mathematical systems edn.

KWAN, W., W. K. LI, AND K. W. NG (2010) : "A Multivariate Threshold Varying Conditional Correlations Model," *Econometric Reviews*, 29(1), 20–38.

LANNE, M., AND P. SAIKKONEN (2005) : "Non–Linear GARCH Models for Highly Persistent Volatility," *The Econometrics Journal*, 8(2), 251–276.

LANNE, M., AND P. SAIKKONEN (2007) : "A Multivariate Generalized Orthogonal Factor GARCH Model," *Journal of Business & Economic Statistics*, 25, 61–75.

LEE, H., AND J. YODER (2007) : "A Bivariate Markov Regime Switching GARCH Approach to Estimate Time Varying Minimum Variance Hedge Ratios," *Applied Economics*, 39(10), 1253–1265.

LI, C., AND W. LI (1996) : "On a Double–Threshold Autoregressive Heteroscedastic Time Series Model," *Journal of Applied Econometrics*, 11(3), 253–274.

LINDGREN, G. (1978) : "Markov Regime Models for Mixed Distributions and Switching Regressions," *Scandinavian Journal of Statistics*, 5, 81–91.

LIU, J. (2006) : "Stationarity of a Markov-Switching GARCH Model," *Journal of Financial Econometrics*, 4(4), 573–593.

LOH, W. (2002) : "Regression trees with unbiased variable selection and interaction detection," *Statistica Sinica*, (12), 361–386.

LONG, X., AND A. ULLAH (2005) : "Nonparametric and Semiparametric Multivariate GARCH Model," .

LONGIN, F., AND B. SOLNIK (1995) : "Is the Correlation in International Equity Returns Constant : 1960–1990 ?," *Journal of International Money and Finance*, 14(1), 3–26.

——— (1996) : "The Asymptotic Distribution of Extreme Stock Market Returns," *Journal of Business*, 69(3), 383–408.

——— (2001) : "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56, 651–678.

LUNDBERGH, S., AND T. TERÄSVIRTA (1998) : "Modelling Economic Highfrequency Time Series with STAR-STGARCH Models," Working Paper Series in Economics and Finance 291, Stockholm School of Economics.

MA, L., Q. ZHOU, M. CELENK, AND D. M. CHELBERG (2004) : "Facial event mining using coupled hidden markov models," in *ICIP*, pp. 1405–1408.

MANDELBROT, B. (1963a) : "New Methods in Statistical Economics," *Journal of Political Economy*, 71, 421.

——— (1963b) : "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

——— (1967) : "How long is the coast of Britain ? Statistical self-similarity and fractional dimension," *Science*, 156, 636–638.

——— (1974) : "Intermittent turbulence in self-similar cascades : divergence of high moments and dimension of the carrier," *Journal Fluid Mechanics*, 62, 331–358.

——— (1982) : *The Fractal Geometry of Nature*. Freeman, San Francisco, 2nd edn.

——— (1988) : "An introduction to multifractal distribution functions," in *Random fluctuations and pattern growth*, ed. by H. E. Stanley, and N. Ostrowsky, pp. 279–291. Kluwer Academic, Dordrecht.

MCLACHLAN, G., AND D. PEEL (2000) : *Finite Mixture Models*, Wiley Series in Probability and Mathematical Statistics.

MEDEIROS, M., AND A. VEIGA (2004) : "Modelling multiple regimes in financial volatility with a flexible coefficient GARCH model," Textos para discussão 486, Department of Economics PUC–Rio (Brazil).

MENG, X., AND D. VAN DYK (1997) : "The EM Algorithm – An Old Folk-Song Sung to a Fast New Tune," *Journal of the Royal Statistical Society Series B*, 59(3), 511–567.

MURPHY, K. (2002) : "Dynamic Bayesian Networks : Representation, Inference and Learning," Ph.D. thesis, UC Berkeley, Computer Science Division.

MURPHY, K., AND M. PASKIN (2001) : "Linear time inference in hierarchical HMMs," in *In Proceedings of Neural Information Processing Systems*.

NAM, K., C. PYUN, AND A. ARIZE (2002) : "Asymmetric Mean–Reversion and Contrarian Profits : Anst-garch Approach," *Journal of Empirical Finance*, 9(5), 563–588.

NEFIAN, A. V., L. LIANG, X. PI, L. XIAOXIANG, C. MAO, AND K. MURPHY (2002) : "A coupled HMM for audio-visual speech recognition," in *International Conference on Acoustics, Speech and Signal Processing (CASSP'02)*, pp. 2013–2016.

PALANDRI, A. (2009) : "Sequential conditional correlations : Inference and evaluation," *Journal of Econometrics*, 153(2), 122–132.

PELLETIER, D. (2006) : "Regime switching for dynamic correlations," *Journal of Econometrics*, 131(1–2), 445–473.

PUTINTSEVA, M. (2010) : "Mixture Dynamic Conditional Correlation Model," *SSRN eLibrary*.

RABEMANANJARA, R., AND J. ZAKOIAN (1993) : "Threshold Arch Models and Asymmetries in Volatility," *Journal of Applied Econometrics*, 8(1), 31–49.

RABINER, L. R. (1989) : "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77(2), 257–286.

ROWEIS, S. T. (2000) : "One Microphone Source Separation," in *Proceedings of NIPS*, pp. 793–799.

SENTANA, E. (1995) : "Quadratic ARCH Models," *Review of Economic Studies*, 62(4), 639–61.

SHEPHARD, N. (2005) : *Stochastic Volatility : Selected Readings*. Oxford University Press, Oxford.

SHEPHARD, N., AND T. G. ANDERSEN (2009) : "Stochastic Volatility : Origins and Overview," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch. Springer, New York.

SHIRYAEV, A. (1996) : *Probability*. Springer-Verlag, Berlin, 2nd edn.

SILVENNOINEN, A., AND T. TERÄSVIRTA (2005) : "Multivariate Autoregressive Conditional Heteroskedasticity with Smooth Transitions in Conditional Correlations," Working Paper Series in Economics and Finance 0577, Stockholm School of Economics.

——— (2007) : "Modelling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH model," Working Paper Series in Economics and Finance 0652, Stockholm School of Economics.

——— (2009) : "Multivariate GARCH models," in *Handbook of Financial Time Series*, ed. by T. Andersen, R. Davis, J. Kreiss, and T. Mikosch. Springer, New York.

SILVENNOINEN, A., AND T. TERÄSVIRTA (2009) : "Modeling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model," *Journal of Financial Econometrics*, 7(4), 373–411.

SO, M. K., AND I. W. YIP (2009) : "Multivariate GARCH Models with Correlation Clustering," Discussion paper, The Hong Kong University of Science and Technology, mimeo.

——— (2010) : "A Multivariate GARCH Model With Aggregate Threshold Dynamics," Discussion paper, The Hong Kong University of Science and Technology, mimeo.

TAYLOR, S. (1982) : "Financial returns modeled by the product of two stochastic processes. A study of daily sugar prices 1961-79," in *Time Series Analysis : Theory and Practice 1*, ed. by O. Anderson, pp. 203–226. North-Holland Amsterdam.

TERÄSVIRTA, T. (1994) : "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218.

TONG, L., AND K. LIM (1980) : "Threshold Autoregressive, Limit Cycles and Cyclical Data," *Journal of the Royal Statistical Society B*, 42(3), 245–292.

TROJANI, F., AND F. AUDRINO (2005) : "A general multivariate threshold GARCH model with dynamic conditional correlations," University of St. Gallen Department of Economics working paper series 2005 2005-04, Department of Economics, University of St. Gallen.

TSE, Y., AND A. TSUI (2002) : "A Multivariate GARCH Model with Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20(3), 351–362.

VAN DER WEIDE, R. (2002) : "GO-GARCH : a multivariate generalized orthogonal GARCH model," *Journal of Applied Econometrics*, 17(5), 549–564.

VAN DIJK, D., AND P. FRANSES (1999) : "Modelling multiple regimes in the business cycle," *Macroeconomic Dynamics*, 3, 311–340.

VAN DIJK, D., T. TERÄSVIRTA, AND P. FRANSES (2000) : "Smooth Transition Autoregressive Models – A Survey of Recent Developments," *Econometric Reviews*, 20(1), 1–47.

VARGAS, G. A. (2006) : "An Asymmetric Block Dynamic Conditional Correlation Multivariate GARCH Model," *The Philippine Statistician*, 55(1–2), 83–102.

——— (2008) : "What Drives the Dynamic Conditional Correlation of Foreign Exchange and Equity Returns ?," MPRA Paper 7174, University Library of Munich, Germany.

VRONTOS, I. D., P. DELLAPORTAS, AND D. N. POLITIS (2003) : "A full-factor multivariate GARCH model," *Econometrics Journal*, 6(2), 312–334.

WIERSTRA, D. (2004) : "Statistical Models for Non-Markovian Control Tasks," Master's thesis, Utrecht University.

WILLIAMS, B., M. TOUSSAINT, AND A. J. STORKEY (2007) : "Modelling motion primitives and their timing in biologically executed movements.," in *NIPS*, ed. by J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis. MIT Press.

WU, C. (1983) : "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11(1), 95–103.

XIE, L. (2005) : "Unsupervised Pattern Discovery for Multimedia Sequences," Phd thesis, Columbia University.

XIE, L., AND Z.-Q. LIU (2007) : "A coupled HMM approach to video-realistic speech animation," *Pattern Recognition*, 40(8), 2325–2340.

ZHOU, T., AND L. CHAN (2008) : "Clustered Dynamic Conditional Correlation Multivariate GARCH Model," in *DaWaK '08 : Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery*.