

Économétrie non-linéaire

Chapitre 3: Modèles à changement de régime Markovien

Gilles de Truchis, Elena Dumitrescu

Master 2 BMM - EIPMC - GDA

Septembre 2016

Les chapitres du cours

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

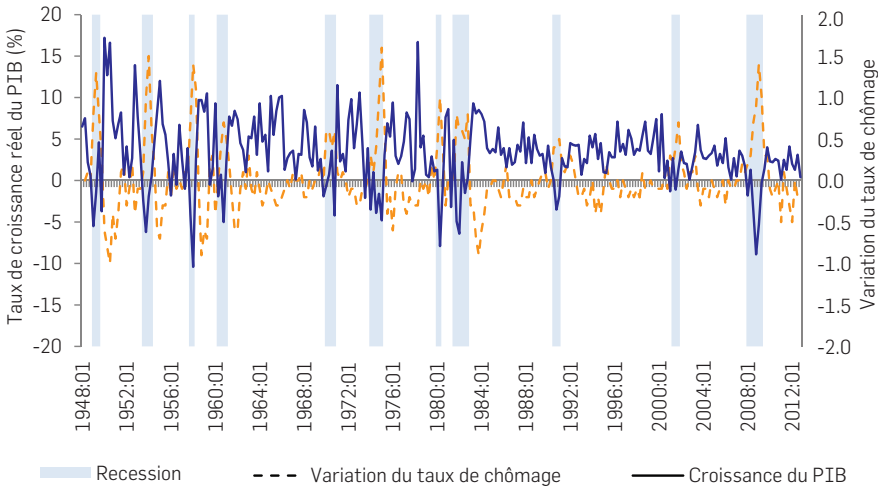
- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Non-linéarité et cycles

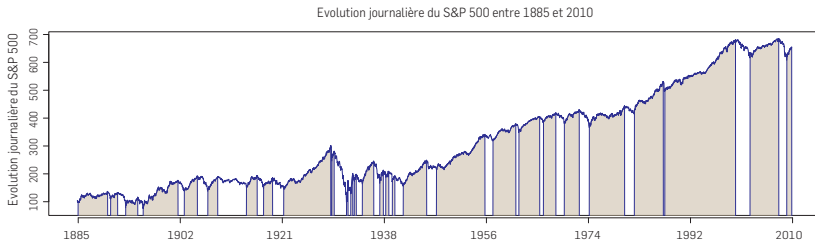
- Les phénomènes cycliques sont un sujet d'étude central
- Quelques exemples :
 - Cycles des affaires : l'économie fait face à des cycles expansionnistes et récessifs
 - Cycles financiers : les marchés boursiers sont animés par des phases haussières et baissières

Cycles des affaires

Données pour les USA, 1948:01-2012:04



Cycles des marchés financiers



Dynamique non-linéaire

- Ces dynamiques cycliques irrégulières échappent aux modèles linéaires
 - Les cycles sont persistants et de durées diverses
- ⇒ ajouter des retards dans un modèle autorégressif ne suffit pas
- Ces dynamiques impliquent des distributions multimodales
- ⇒ l'hypothèse Gaussienne ne tient pas

- En fait on verra plus tard qu'avec un ARMA non-Gaussien on peut approximer ces dynamiques
- On ne peut pas miser sur une inférence statistique basée sur l'hypothèse de normalité

Intuition

- Tout se passe comme si

- plusieurs **états de la nature** de la variable observée existaient

⇒ e.g. haussier / baissier

- une **variable cachée** déterminait l'état du processus observé

⇒ e.g. le sentiment du marché (pessimiste / optimiste)

- Hypothèses :

- soit une séquence aléatoire $\{Y_t\}_{t=1}^n$ observables

- les réalisations de Y_t sont notées y_t

- La distribution de Y_t dépend d'un processus sous-jacent S_t

- S_t est aléatoire et représente l'état de la nature en t

Les états de la nature

- S_t est caractérisé par un nombre **fini** d'états de la nature
 - On notera \mathbb{S} l'ensemble **dénombrable** des états de la nature
 - On notera $i \in \mathbb{S}$ un état particulier de la nature
- \Rightarrow e.g. pour un marché financier : $\mathbb{S} = \{bull, bear\}$
- Le processus S_t est généralement **inobservé**
 - Il s'agit donc d'une **variable latente** (voir Quandt , 1958)
- \Rightarrow question cruciale : comment inférer S_t à partir de y_t ?

Si la notion de variable latente ne leur parle pas voici une définition :

- Une variable latente est une variable que l'on observe pas directement mais que l'on peut inférer à partir d'autres variables observées via un modèle économétrique.

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Chaîne de Markov

- Goldfeld et Quandt (1973) proposent de modéliser S_t par une **chaîne de Markov**

- Intuition de la chaîne de Markov d'ordre k :

- On suppose l'ensemble des états de la nature \mathbb{S} dénombrable

- La probabilité d'être dans un état i à la période suivante dépend de l'état présent et des $k - 1$ états précédents

⇒ Une chaîne de Markov d'ordre k est une variable aléatoire notée $\{S_t\}_{t \in \mathbb{N}}$ à valeur dans \mathbb{S} telle que $\forall i_t \in \mathbb{S}$ et $t \geq 1$,

$$\mathbb{P}(S_{t+1} = i_{t+1} | S_t = i_t, \dots, S_0 = i_0) = \mathbb{P}(S_{t+1} = i_{t+1} | S_t = i_t, \dots, S_{t-k+1} = i_{t-k+1})$$

Chaîne de Markov

- Par la suite on considèra des chaînes de Markov d'ordre $k = 1$:
 - la probabilité d'observer l'état i à la période suivante ne dépend que de l'état présent et non de l'état passé

$$\mathbb{P}(S_{t+1} = i_{t+1} | S_t = i_t, \dots, S_0 = i_0) = \mathbb{P}(S_{t+1} = i_{t+1} | S_t = i_t)$$

- On parle de processus sans mémoire : propriété de Markov

Hypothèse

H1 : $\{S_t\}_{t \in \mathbb{N}}$ est une chaîne de Markov d'ordre 1 à valeur dans \mathbb{S} , l'ensemble dénombrable des états de la nature

Rappel pour les processus Markovien

- Propriété de Markov = absence de mémoire : toute l'information utile pour prédire le futur est contenue dans l'état présent du processus et les états antérieurs ne comptent pas

Matrice de transition

- Soit S_t un processus respectant **H1**

- soit deux états distincts : $i, j \in \mathbb{S}, j \neq i$

- la probabilité de transition de l'état i vers l'état j est alors

$$p_{ij} = \mathbb{P}(S_{t+1} = j | S_t = i)$$

- la probabilité de ne pas transiter d'un état à l'autre est

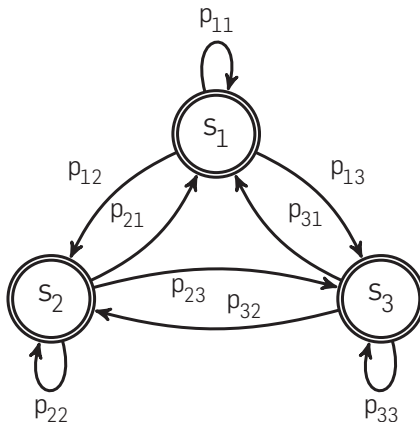
$$p_{ii} = \mathbb{P}(S_{t+1} = i | S_t = i)$$

- naturellement, $\forall i$ et $\forall j$ on a $p_{ij} \geq 0$

- Les paramètres de S_t sont donc les probabilités de transition

- elles forment une **matrice de transition** dont la taille dépend du nombre d'états

Chaîne de Markov



Représentation des probabilités de transition
d'une chaîne de Markov à trois états

Matrice de transition

- Pour N états on obtient alors une matrice $N \times N$ des probabilités de transition

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1j} & \cdots & p_{1N} \\ \vdots & \ddots & & & \vdots \\ p_{i1} & & p_{ij} & & p_{iN} \\ \vdots & & & \ddots & \vdots \\ p_{N1} & \cdots & p_{Nj} & \cdots & p_{NN} \end{pmatrix} \quad (1)$$

- $\forall t \in \mathbb{N}$, la i -ième ligne de P définit la distribution de S_t conditionnellement à l'information que $S_{t-1} = i$, ce qui implique

$$\sum_j p_{ij} = 1, \quad \{i, j\} \in \mathbb{S}$$

- Sous forme matricielle cela s'exprime ainsi :

$$P \times \mathbf{1}_{N \times 1} = \mathbf{1} \times \mathbf{1}_{N \times 1} \quad (2)$$

avec $\mathbf{1}_{N \times 1}$ un vecteur colonne composé de 1

Le résultat simple de l'équation (2) est important pour la suite car il exprime une propriété importante sur les vecteurs propres et valeurs propres des chaînes de Markov.

En effet, cela montre que $\lambda = 1$ est une valeur propre de P et que le vecteur propre unitaire $\mathbf{1}_{N \times 1}$ lui est associé.

On reviendra plus tard sur ce résultat.

Matrice de transition

- Pour inférer S_t à partir de y_t , **H1** ne suffit pas

⇒ il faut que S_t vérifie des propriétés supplémentaires

Hypothèse

H2 : $\{S_t\}_{t \in \mathbb{N}}$ est une chaîne de Markov d'ordre 1, homogène, irréductible, apériodique, de récurrence positive, de distribution ergodique $\pi = (\pi_1, \dots, \pi_N)$ et vérifiant

$$\mathbb{P}(S_0 = i | P) = \pi_i$$

- Les caractéristiques de P suffisent à garantir ces propriétés
- Nous allons définir chacune de ces propriétés

Selon la forme de la matrice P , il est possible de détecter si S_t possède de bonnes propriétés. Si ce n'est pas le cas, alors on ne peut pas faire confiance aux résultats de l'estimation...

Propriétés des chaînes de Markov

Definition

Homogénéité : $\{S_t\}_{t \in \mathbb{N}}$ est homogène si les probabilités de transition d'un état à l'autre sont indépendantes du temps

$$\mathbb{P}(S_{t+1} = j | S_t = i) = \mathbb{P}(S_1 = j | S_t = i), \quad \forall t$$

- Il existe des processus à changement Markovien non-homogènes :
 - Filardo (1994) propose de faire dépendre les probabilités de transition d'un processus exogène (e.g. une variable économique)

Propriétés des chaînes de Markov

Definition

Irréductibilité : $\{S_t\}_{t \in \mathbb{N}}$ est irréductible si tout état est accessible à partir de n'importe quel autre état. En d'autres termes, pour $\tau_{ij} \in \{0, 1, 2, \dots\}$,

$$\mathbb{P}(S_{\tau_{ij}} = j | S_0 = i) = p_{ij}^{(\tau_{ij})} > 0, \quad i \neq j$$

■ Exemple de réductibilité (absence d'irréductibilité) :

■ Soit S_t une chaîne de Markov à N états avec $p_{11} = 1$

\Rightarrow une fois dans l'état 1, on ne pourra plus quitter cet état

Si S_t est réductible, P dispose d'une unique valeur propre unitaire et les autres valeurs propres sont dans le cercle unitaire. En conséquence, les conditions de stationnarité ne sont pas respectées pour S_t . Il s'en suit que y_t ne sera pas non plus stationnaire.

Propriétés des chaînes de Markov

Definition

Apériodicité : $\{S_t\}_{t \in \mathbb{N}}$ est apériodique si un état i peut survenir de manière irrégulière sans périodicité dans son occurrence. En d'autres termes, il existe $\tau \in \mathbb{N}$ tel que $\tau' \geq \tau$ et

$$\mathbb{P}(S_{\tau'} = i | S_0 = i) > 0$$

- Astuce pour analyser l'apériodicité
 - S_t est apériodique si les coefficients diagonaux de la matrice de transition sont non-nuls (donc positifs)
- Exemple de matrice de transition périodique (mais irréductible)

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

Propriétés des chaînes de Markov

- La transience se caractérise par le temps de séjour dans un état à long terme, dénoté η_i et donné par

$$\eta_i = \sum_{t=0}^{n-1} \mathbb{1}_i S_t$$

avec $\mathbb{1}_i$ la fonction indicatrice qui vaut 1 si $S_t = i$ et zéro sinon

Definition

Transience : un état est transient s'il existe une probabilité unitaire que la chaîne de Markov passe un nombre fini de fois par cet état

$$\mathbb{P}(\lim_{n \rightarrow \infty} \eta_i < \infty) = 1$$

Propriétés des chaînes de Markov

Definition

Réurrence : un état est récurrent s'il existe une probabilité unitaire que la chaîne de Markov passe une infinité de fois par cet état

$$\mathbb{P}(\lim_{n \rightarrow \infty} \eta_i = \infty) = 1$$

- On remarque aisément qu'un état est soit transient, soit récurrent

Propriétés des chaînes de Markov

- On distingue 2 types de récurrence selon le comportement de η_i

Definition

Récurrence positive : Une état i est récurrent positif si

$$\lim_{n \rightarrow \infty} \frac{\eta_i}{n} = c_i > 0$$

ce qui indique que le processus passe une fraction importante de son temps dans l'état i

Definition

Récurrence nulle : Une état i est récurrent nul si

$$\lim_{n \rightarrow \infty} \frac{\eta_i}{n} = 0$$

ce qui indique que le processus passe une fraction négligeable de son temps dans l'état i

Propriétés des chaînes de Markov

Definition

***Ergodicité d'état** : un état i est ergodique s'il est apériodique et de récurrence positive.*

Definition

***Ergodicité** : Si $\{S_t\}_{t \in \mathbb{N}}$ est irréductible et que tous ses états sont ergodiques, alors $\{S_t\}_{t \in \mathbb{N}}$ est ergodique.*

- L'ergodicité de la chaîne de Markov est une propriété importante :
 - ⇒ elle assure que la distribution de S_t **converge** vers une **distribution invariante** quelque soit l'état initial S_0
- Nous allons donc définir la propriété d'invariance de S_t

Propriété d'invariance

- Toute distribution de probabilité $\pi = (\pi_1, \dots, \pi_N)$ qui respecte la propriété d'invariance

$$P'\pi = \pi \quad (3)$$

est dénommée distribution invariante de S_t

- Importance de l'invariance : soit S_t dont les états sont tirés d'une distribution invariante de P de sorte que $\mathbb{P}(S_t = i|P) = \pi_i$

$\Rightarrow \forall j = 1, \dots, N$ on peut écrire la probabilité non-conditionnelle d'observer l'état j en $t + 1$ ainsi

$$\mathbb{P}(S_{t+1} = j|P) = \sum_{i=1}^N \mathbb{P}(S_{t+1} = j|S_t = i)\mathbb{P}(S_t = i|P) = \sum_{i=1}^N p_{ij}\pi_i = \pi_j$$

- On constate que $\mathbb{P}(S_{t+1} = j|P)$ est également tirée de π

\Rightarrow cela est vrai pour S_{t+2}, S_{t+3}, \dots

La propriété d'invariance existe pour toute chaîne de Markov, MAIS la distribution de S_t converge vers la distribution invariante quelque soit l'état initial de S_0 uniquement si S_t est ergodique

Propriété d'invariance

- La propriété d'invariance existe pour toute chaîne de Markov
- En effet, rappelons que

$$\sum_j p_{ij} = 1, \quad \{i, j\} \in \mathbb{S}$$

et que sous forme matricielle cela s'exprime comme

$$P \times \mathbf{1}_{N \times 1} = 1 \times \mathbf{1}_{N \times 1}$$

- Cela montre que $\forall P$, une des valeurs propres est égale à 1
- En réécrivant la propriété d'invariance (3) ainsi

$$\pi' P = \pi' \times 1 \tag{4}$$

on voit que π est le vecteur propre (à gauche) associé à la valeur propre 1

Rappel : valeurs propres et vecteurs propres

- Soit M une matrice 2×2

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

- Une valeur propre λ de M est un réel tel que

$$\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

- $(v_1 \ v_2)'$ est le vecteur propre associé à la valeur propre λ
- En notation matricielle on a

$$Mv = \lambda v$$

- Dans le cas $\pi'P = \pi' \times 1$ on parle de vecteur propre à gauche car π' est à gauche de P
- On aurait pu travailler sur $P'\pi = \pi$ dans la mesure où les valeurs propres restent les mêmes mais les vecteurs propres de P' ne sont pas égaux aux vecteurs propres de P

Distribution invariante

- Cependant, la distribution invariante π n'est pas unique $\forall P$
- Pour **garantir l'unicité** de π , S_t doit être irréductible
- Sous cette hypothèse on peut trouver π à partir de la propriété d'invariance (3) en utilisant (4)
- Pour $N = 2$ on trouve les probabilités invariantes suivantes

$$\begin{aligned}\pi_1 &= \frac{1 - p_{22}}{(1 - p_{11}) + (1 - p_{22})} = \frac{p_{21}}{p_{12} + p_{21}} \\ \pi_2 &= \frac{1 - p_{11}}{(1 - p_{11}) + (1 - p_{22})} = \frac{p_{12}}{p_{12} + p_{21}}\end{aligned}\tag{5}$$

- Notons que si $p_{11} > p_{22}$, $\pi_1 > \pi_2$: plus p_{11} est grand, plus l'état 1 sera **persistant**

Dans quelques slides il y aura une application concrète pour mieux comprendre comment on obtient π_1 et π_2

Distribution invariante

- Pour $N > 2$, trouver le vecteur propre π est complexe
- Hamilton (1994) propose une solution en introduisant

$$M = \begin{pmatrix} I_N - P' \\ \mathbf{1}_{1 \times N} \end{pmatrix}$$

- Avec I_N une matrice identité et $\mathbf{1}_{1 \times N}$ un vecteur ligne de 1
- π est alors donné par

$$\pi = \left((M' M)^{-1} M' \right)_{., N+1} \quad (6)$$

où $., N + 1$ désigne la dernière colonne de la matrice $(M' M)^{-1} M'$

En gros ça reste la méthode du vecteur propre mais une version évoluée. On peut trouver l'explication détaillée dans [Hamilton \(1994\)](#) p 684.

Ergodicité et convergence

- Outre l'irréductibilité, l'ergodicité est également cruciale pour des questions d'inférence

Cas 1 S_0 est tiré de la distribution invariante π

$$\Rightarrow \mathbb{P}(S_t|P) = \pi \text{ pour } t \geq 1$$

Cas 2 S_0 est tiré d'une distribution différente de π

\Rightarrow Si S_t est ergodique, on peut montrer que $\mathbb{P}(S_t|P, S_0 = i)$ converge vers π quelque soit l'état initial i

\Rightarrow Si S_t n'est pas ergodique, pas de convergence vers π dans le Cas 2

C'est important lors de l'initialisation des algorithmes dans l'estimation. Sans l'ergodicité, on est pas certain de converger vers la distribution invariante de S_t

Distribution prédictive

- Cette convergence peut s'analyser via les puissances de P :

$$P^h = P \dots P$$

- P^h détermine la dynamique de long terme de S_t

\Rightarrow la distribution prédictive de S_t pour $t+h$ est donnée par

$$\mathbb{P}(S_{t+h} = l | S_t = k, P) = (P^h)_{kl}$$

- En effet, pour $h > 1$

$$\begin{aligned}\mathbb{P}(S_{t+h} = l | S_t = k, P) &= \sum_{j=1}^N \mathbb{P}(S_{t+h} = l | S_{t+h-1} = j) \mathbb{P}(S_{t+h-1} = j | S_t = k) \\ &= \sum_{j=1}^N p_{jl} (P^{h-1})_{kj} = (P^h)_{kl}\end{aligned}$$

Convergence et distribution invariante

- En termes de distribution prédictive, le Cas 2 s'analyse ainsi :
- Observons que $\mathbb{P}(S_t|S_0 = i, P) = (P^h)_i$.
 - la notation $(P^h)_i$ dénote la i -ième ligne de P^h
- $\forall i$ et S_t ergodique, on peut montrer que

$$\lim_{h \rightarrow \infty} (P^h)_i = \pi'$$

Vitesse de convergence

- Analyser la vitesse de convergence de P^h vers π est intéressant pour des questions d'estimation
- Considérons $\{S_t\}_{t \in \mathbb{N}}$ à valeur dans $\mathbb{S} = \{1, 2\}$ et

$$P = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}$$

- Supposons $p_{11} < 1$ et $p_{22} < 1$ pour garantir l'ergodicité et calculons les valeurs propres

$$\begin{vmatrix} p_{11} - \lambda & 1 - p_{11} \\ 1 - p_{22} & p_{22} - \lambda \end{vmatrix} = (\lambda - 1)(\lambda - (p_{11} + p_{22} - 1)) = 0$$

- On sait que $\lambda_1 = 1$ est une solution et que la seconde solution est donnée par

$$\lambda_2 = p_{11} + p_{22} - 1 = p_{11} - p_{21}$$

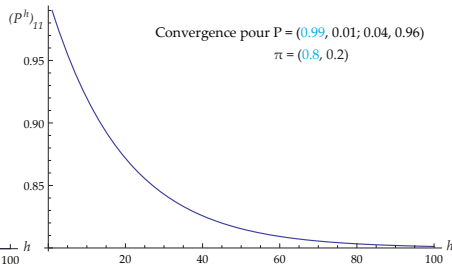
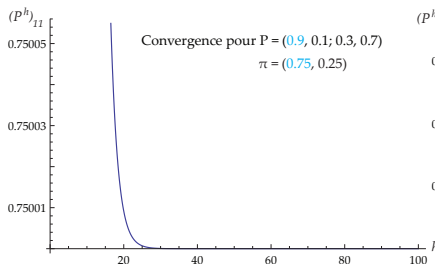
- On voit que plus S_t est persistante, plus λ_2 se rapproche de 1

Rappelons que S_t est stationnaire si $\lambda_2 < 1$

Vitesse de convergence

- Hamilton (1994) montre que pour $N = 2$, on peut représenter P^h en fonction des probabilités invariantes pour tout h

$$P^h = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix} + \lambda_2^h \begin{pmatrix} \pi_2 & -\pi_2 \\ -\pi_1 & \pi_1 \end{pmatrix}$$



On remarque que plus S_t est persistante, plus λ_2 se rapproche de 1 et plus la convergence de P^n vers π est lente (encore plus flagrant quand on regarde l'échelle).

Stationnarité

- S'il existe une mesure $\pi = (\pi_i)_{i \in \mathbb{S}}$ telle que $\pi P = \pi$, alors π est une **mesure stationnaire** de S_t
 - **mesure stationnaire** et **distribution invariante** sont des synonyme
- Si $\forall i \in \mathbb{S}, \pi_i \geq 0$ et $\sum_i \pi_i = 1$, π représente la **loi stationnaire** de S_t (découle de l'ergodicité de S_t)
- S_t est un **processus stationnaire** si sa loi initiale, i.e. la loi de S_0 , est stationnaire

En gros, S_t est stationnaire si la matrice de transition est telle qu'après n période on converge vers la distribution invariante du processus, i.e. $\pi = (\pi_i)_{i \in \mathbb{S}}$

Applications

- Soit un marché pouvant être dans 2 états de la nature

$$\mathbb{S} = \{1 = \textit{haussier}, 2 = \textit{baissier}\}$$

- Le comportement du marché est gouverné par une chaîne de Markov dont la matrice de transition est

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix}$$

- Cherchons à calculer π
- Le vecteur propre est composé des solutions du système

$$(P' - \lambda I_2)v$$

avec $\lambda = 1$ et sachant que $\pi_1 + \pi_2 = 1$ et donc $v_1 + v_2 = 1$

Applications

- Dans le cas présent on a

$$\left(\begin{pmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Avec $\lambda = 1$ et sous forme de système cela donne

$$\begin{cases} -0.1v_1 + 0.3v_2 = 0 \\ 0.1v_1 - 0.3v_2 = 0 \end{cases}$$

- Or on sait que $v_1 = 1 - v_2$ et donc $0.1(1 - v_2) - 0.3v_2 = 0$ ce qui donne

$$0.1 - 0.4v_2 = 0 \Leftrightarrow v_2 = 0.1/0.4 = 0.25 = \pi_2$$

- On en déduit alors $v_1 = 1 - v_2 = 0.75 = \pi_1$
- En utilisant les formules en (5) on trouve bien 0.75
- Conclusion, le marché est haussier 75% du temps

Applications

- Soit un marché pouvant être dans trois états de la nature

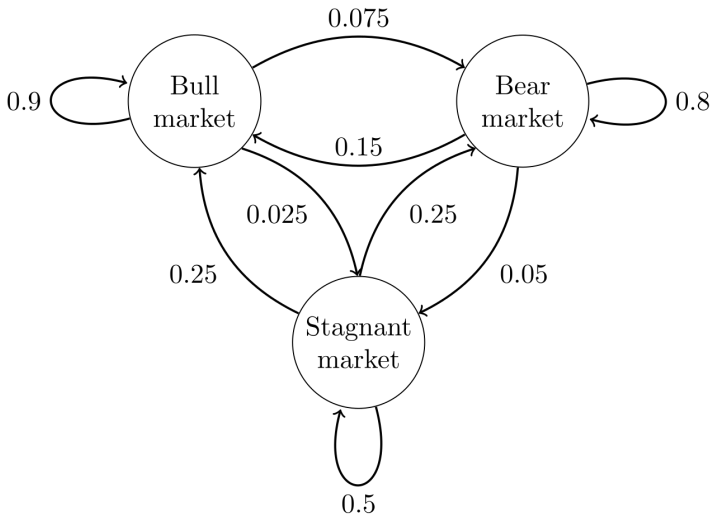
$$\mathbb{S} = \{1 = \textit{haussier}, 2 = \textit{baissier}, 3 = \textit{stagnant}\}$$

- Le comportement du marché est gouverné par une chaîne de Markov dont la matrice de transition est

$$P = \begin{pmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}$$

- 1 Représentons graphiquement le processus de Markov
- 2 Cherchons à calculer π

Chaîne de Markov



Applications

■ Calcul de π

■ Commençons par construire la matrice M

$$M = \begin{pmatrix} I_N - P' \\ \mathbf{1}_{1 \times N} \end{pmatrix} = \begin{pmatrix} 0.1 & -0.15 & -0.25 \\ -0.075 & 0.2 & -0.25 \\ -0.025 & -0.05 & 0.5 \\ 1 & 1 & 1 \end{pmatrix}$$

■ En utilisant un logiciel (ici Mathematica) on trouve aisément

$$\pi = \left(\left(M' M \right)^{-1} M' \right)_{.,N+1} = \begin{pmatrix} 2.38 & -1.55 & -0.83 & 0.6250 \\ -1.67 & 2.08 & -0.41 & 0.3125 \\ -0.71 & -0.54 & 1.25 & 0.0625 \end{pmatrix}_{.,N+1}$$

■ On a donc $\pi = (0.6250 \ 0.3125 \ 0.0625)$

\Rightarrow L'état baissier est observé 31.25% du temps

Distribution initiale de S_0

- S_0 joue un rôle crucial dans la spécification de S_t et donc de Y_t
- Sous **H2**, S_t débute tel que

$$\mathbf{p}_0 = (p_{0,1}, \dots, p_{0,N}) = \pi$$

avec $p_{0,k} = \mathbb{P}(S_0 = k)$

- Dans la pratique, π n'est généralement pas observé
- on peut relâcher **H2** et fixer arbitrairement \mathbf{p}_0
 - \Rightarrow dans ce cas la stationnarité de S_t n'est plus garantie
- on peut traiter \mathbf{p}_0 comme une variable inconnue
 - \Rightarrow dans ce cas on cherche à l'estimer

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Distribution de Y_t

- Soit S_t une chaîne de Markov respectant **H2**
- Il reste à étudier comment la distribution de Y_t dépend de S_t
- Soit $\mathcal{D}(\theta)$ une famille de distributions de paramètres θ et définie sur un espace d'échantillonnage discret
- On note $p(y|\theta)$ la densité de $\mathcal{D}(\theta)$ avec $\theta \in \Theta$ et Θ un sous-ensemble compact de $\mathbb{R}^{\dim(\theta)}$
- Sachant $\{S_t\}_{t=0}^n$, on suppose la séquence $\{Y_t\}_{t=1}^n$ indépendante

Hypothèse

H3 : Pour $t \geq 1$, la distribution de Y_t est issue des N distributions $\mathcal{D}(\theta_1), \dots, \mathcal{D}(\theta_N)$, selon l'état de S_t :

$$Y_t | S_t = k \sim \mathcal{D}(\theta_k)$$

Distribution de Y_t

- Sous **H3**, Y_t est un modèle de mélange

$\Rightarrow Y_t$ est un mélange de distribution dont la règle de mélange dépend d'une variable inobservée, la chaîne de Markov S_t

- Sous **H2** et **H3**, on peut dériver la distribution non-conditionnelle de Y_t

$$p(y_t|\theta) = \sum_{k=1}^N p(y_t|S_t = k, \theta) \mathbb{P}(S_t = k|\theta)$$

- Comme sous **H2** S_t est stationnaire et sous **H3** $p(y_t|S_t = k, \theta) = p(y_t|\theta_k)$, on peut également écrire la distribution non-conditionnelle de Y_t comme

$$p(y_t|\theta) = \sum_{k=1}^N p(y_t|\theta_k) \pi_k \quad (7)$$

- Sous **H2**, S_t est stationnaire, donc Y_t est stationnaire

Modèle à changement de régime markovien dans la moyenne

- Partons du modèle le plus simple possible sous **H2** et **H3**

$$Y_t = \mu_{S_t} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad S_t \in \mathbb{S} = \{1, 2\} \quad (8)$$

- La notation μ_{S_t} signifie que μ dépend de la chaîne de Markov

\Rightarrow donc ici $\theta_k = \mu_k$

- Il existe alors deux modèles possibles pour Y_t

$$Y_t = \begin{cases} \mu_{S_t=1} + \varepsilon_t, & S_t = 1 \\ \mu_{S_t=2} + \varepsilon_t, & S_t = 2 \end{cases}$$

\Rightarrow Cette classe de modèle dit **Markov-Switching (MS)** a été popularisée par [Hamilton \(1989\)](#)

- Y_t est donc un **mélange Markovien** de 2 distributions normales

Modèle à changement en moyenne-variance

- Des modèles MS plus sophistiqués sont également possibles
- Une **première généralisation** : changement dans la variance

$$Y_t = \mu_{S_t} + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \quad S_t \in \mathbb{S} = \{1, 2\} \quad (9)$$

⇒ ce modèle autorise donc une forme d'hétéroscédasticité et

$$\theta_k = (\mu_k, \sigma_k)'$$

- La distribution de Y_t conditionnelle à S_t est alors

$$(Y_t | S_t = k) \sim \mathcal{D}(\theta_k) \stackrel{d}{=} \mathcal{N}(\mu_k, \sigma_k^2)$$

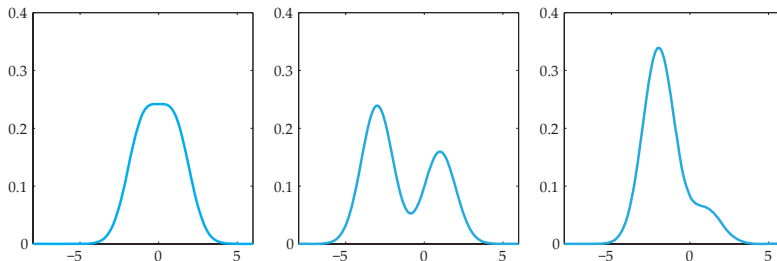
- La fonction de densité non-conditionnelle à S_t est alors

$$p(Y_t | \theta) = \pi_1 p(Y_t | \mu_1, \sigma_1) + \pi_2 p(Y_t | \mu_2, \sigma_2)$$

comme définie à l'équation (7)

Distribution du MS en moyenne-variance

- Pour $\mu_1 \neq \mu_2$ ou $\sigma_1 \neq \sigma_2$, $p(Y_t|\theta)$ est non-Gaussien



- Densité non-conditionnelle d'un mélange markovien pour $\sigma_1 = \sigma_2 = 1$
 - à gauche : $\mu_1 = -1$ et $\mu_2 = 1$, $\pi_1 = 0.5$
 - à droite : $\mu_1 = -2$ et $\mu_2 = 1$, $\pi_1 = 0.85$
 - au milieu : $\mu_1 = -3$ et $\mu_2 = 1$, $\pi_1 = 0.6$

Moments d'une mixture de distributions markovienne

- Soit Y_t un processus résultant d'un mélange markovien à N états
- Espérance de $h(Y_t)$ une fonction continue de Y_t :

$$\mathbb{E}(h(Y_t)|\theta) = \sum_{i=1}^N \mathbb{E}(h(Y_t)|\theta_i)\pi_i, \text{ avec}$$

$$\mathbb{E}(h(Y_t)|\theta_i) = \int h(Y_t)p(Y_t|\theta_i)dY_t$$

\Rightarrow application pour la moyenne : on a $h(Y_t) = Y_t$ et donc

$$\mu = \mathbb{E}(Y_t|\theta) = \sum_{i=1}^N \mu_i \pi_i$$

\Rightarrow application pour la variance : on a $h(Y_t) = (Y_t - \mu)^2$ et donc

$$\sigma^2 = \mathbb{V}(Y_t|\theta) = \sum_{i=1}^N (\mu_i^2 + \sigma_i^2)\pi_i - \mu^2$$

- Timmermann (2000) dérive les moments d'ordre plus élevés

- de manière générale, pour le moment centré d'ordre k on a

$$\mathbb{E}(h(Y_t)^m|\theta) = \mathbb{E}((Y_t - \mu)^m|\theta) = \sum_{i=1}^N \mathbb{E}((Y_t - \mu_i + \mu_i - \mu)^m|\theta_i)\pi_i$$

- A l'aide des coefficients binomiaux on obtient la formule suivante mais c'est trop moche pour les étudiants (voir Fruhwirth-Schnatter p. 11)

$$\mathbb{E}((Y_t - \mu)^m|\theta) = \sum_{i=1}^N \sum_{j=0}^m \binom{m}{i} (\mu_i - \mu)^{m-j} \mathbb{E}((Y_t - \mu_i)^j|\theta_i)\pi_i$$

Autocovariance d'une mixture de distributions markovienne

- Soit Y_t un processus résultant d'un mélange markovien à N états
- La fonction d'autocovariance est donnée par :

$$\begin{aligned} Cov(Y_t, Y_{t+h}) &= \mathbb{E}(Y_t Y_{t+h}) - \mu^2 \\ &= \int Y_t Y_{t+h} p(Y_t, Y_{t+h} | \theta) dY_t dY_{t+h} - \mu^2 \end{aligned}$$

- avec $p(Y_t, Y_{t+h} | \theta)$ la distribution non-conditionnelle jointe de Y_t et Y_{t+h} donnée par

$$\begin{aligned} p(Y_t, Y_{t+h} | \theta) &= \sum_{i,j=1}^N p(Y_t | S_t = i, \theta) p(Y_{t+h} | S_{t+h} = j, \theta) \\ &\quad \times \mathbb{P}(S_t = i) \mathbb{P}(S_{t+h} = j | S_t = i) \\ &= \sum_{i=1}^N p(Y_t | \theta_i) \pi_i \sum_{j=1}^N p(Y_{t+h} | \theta_j) (P^h)_{ij} \end{aligned}$$

Rappel slides sur les moments et l'autocovariance :

- Si Y_t est une variable aléatoire continue, elle admet une densité de probabilité $p(Y_t|\theta)$ et son espérance est donnée par

$$\mathbb{E}(Y_t) = \int Y_t p(Y_t|\theta) dY_t = \mu$$

- De même, pour une densité de probabilité jointe $p(Y_t Y_{t+h} | \cdot)$ on obtient

$$\mathbb{E}(Y_t Y_{t+h}) = \int Y_t Y_{t+h} p(Y_t, Y_{t+h} | \theta) dY_t dY_{t+h}$$

Autocovariance d'une mixture de distributions markovienne

- L'expression de $p(Y_t, Y_{t+h}|\theta)$ nous permet de réécrire $\mathbb{E}(Y_t Y_{t+h})$

$$\mathbb{E}(Y_t Y_{t+h}) = \sum_{i=1}^N \mu_i \pi_i \sum_{j=1}^N \mu_j (P^h)_{ij} \quad (10)$$

- La fonction d'autocovariance devient alors

$$\text{Cov}(Y_t, Y_{t+h}) = \sum_{i=1}^N \mu_i \pi_i \sum_{j=1}^N \mu_j (P^h)_{ij} - \mu^2 = \gamma_Y(h) \quad (11)$$

- $\gamma_Y(h)$ ne dépend que de h et son indépendance au temps est claire

- La fonction d'autocorrélation de Y_t est donnée par

$$\rho_Y(h) = \frac{\text{Cov}(Y_t, Y_{t+h})}{\sigma_t \sigma_{t+h}} = \frac{\gamma_Y(h)}{\sigma^2} \quad (12)$$

Calcul de l'autocovariance (à vérifier) :

- En partant de l'expression de $\mathbb{E}(Y_t Y_{t+h})$ et en appliquant le Théorème de Fubini on obtient

$$\begin{aligned}\mathbb{E}(Y_t Y_{t+h}) &= \int Y_t Y_{t+h} p(Y_t, Y_{t+h} | \theta) dY_t dY_{t+h} \\&= \int Y_t Y_{t+h} \sum_{i=1}^N p(Y_t | \theta_i) \pi_i \sum_{j=1}^N p(Y_{t+h} | \theta_j) (P^h)_{ij} dY_t dY_{t+h} \\&= \sum_{i=1}^N \sum_{j=1}^N \int Y_t p(Y_t | \theta_i) \pi_i Y_{t+h} p(Y_{t+h} | \theta_j) (P^h)_{ij} dY_t dY_{t+h} \\&= \sum_{i=1}^N \sum_{j=1}^N \mu_i \pi_i \mu_j (P^h)_{ij} \\Cov(Y_t, Y_{t+h}) &= \mathbb{E}(Y_t Y_{t+h}) - \mu^2 = \sum_{i=1}^N \mu_i \pi_i \sum_{j=1}^N \mu_j (P^h)_{ij} - \mu^2\end{aligned}$$

Moments du MS en moyenne-variance

- Concernant le modèle MS décrit par l'Equation (9) avec $N = 2$

- pour la moyenne on obtient

$$\mu = \mathbb{E}(Y_t|\theta) = \mu_1\pi_1 + \mu_2\pi_2$$

\Rightarrow Notons que $\mu < \infty$ si $\mu_i < \infty$ et que μ ne dépend pas du temps

- pour la variance on obtient

$$\sigma^2 = \mathbb{V}(Y_t|\theta) = \sigma_1^2\pi_1 + \sigma_2^2\pi_2 + \pi_1\pi_2(\mu_2 - \mu_1)^2$$

\Rightarrow Notons que σ^2 ne dépend pas du temps

\Rightarrow On sait également que $\gamma_Y(h)$ ne dépend que de h

$\Rightarrow Y_t$ est **stationnaire** car la chaîne de Markov est stationnaire sous

H2

- On peut avoir un mélange de distributions stationnaires dans chaque régime mais avoir Y_t globalement non-stationnaire si la chaîne de Markov est non-ergodique (e.g. si elle réductible)

Autocorrelation du MS en moyenne-variance

- La fonction d'autocorrelation du modèle :

$$\rho_Y(h; \theta) = \frac{\gamma_Y(h)}{\sigma^2} = \frac{\pi_1 \pi_2 (\mu_1 - \mu_2)^2}{\sigma^2} \lambda_2^h$$

avec $\lambda_2 = p_{11} - p_{21}$

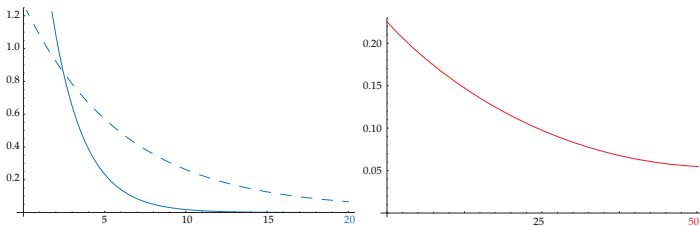
⇒ En rappelant que $p_{21} = 1 - p_{22}$, on constate que

- dès lors que $\mu_1 \neq \mu_2$ l'autocorrélation peut apparaître
- si les régimes sont suffisamment persistents ($p_{11} + p_{22} > 1$), de l'autocorrélation **positive** apparaît
- dès lors que $p_{11} + p_{22} < 1$, de l'autocorrélation **négative** apparaît

⇒ Notons qu'au sein de chaque état, Y_t n'est pas autocorrélé

$$Y_t = \begin{cases} \mu_{S_t=1} + \sigma_{S_t=1} \varepsilon_t, & S_t = 1 \\ \mu_{S_t=2} + \sigma_{S_t=2} \varepsilon_t, & S_t = 2 \end{cases}$$

Propriété du MS en moyenne-variance



■ Fonction d'autocovariance $\gamma_Y(h)$, du MS moyenne-variance pour

■ à gauche (ligne) : $p_{12} = 0.3$, $p_{21} = 0.1$ et $\mu_1 = 2$, $\mu_2 = -2$

■ à gauche (pointillés) : $p_{12} = 0.05$, $p_{21} = 0.1$ et $\mu_1 = 2$, $\mu_2 = -2$

■ à droite : $p_{12} = 0.01$, $p_{21} = 0.01$ et $\mu_1 = 2$, $\mu_2 = -2$

■ On voit que si les régimes sont très persistents, il est possible de **confondre** le MS avec un processus non-stationnaire

En fait plutôt un processus à mémoire longue mais vu ce type de processus qu'en cours de non-stationnaire...

Modèle de mélange Markovien et ARMA

- Poskitt et Chung (1996) montrent qu'il existe un lien entre MS et ARMA
- Exemple pour un modèle MS en moyenne à 2 états

- La fonction d'autocorrélation d'un modèle MS peut s'écrire

$$\rho_Y(h|\theta) = \lambda_2 \rho_Y(h-1|\theta)$$

ce qui correspond à la fonction d'autocorrélation d'un ARMA(1,1)

- Plus généralement un modèle MS en moyenne à N états

⇒ il existe une représentation ARMA($N-1, N-1$) du processus

⇒ Attention il s'agit d'un ARMA avec erreurs non-Gaussiennes

Modèle de mélange Markovien en finance

- Timmermann (2000) : les modèles MS sont intéressants pour modéliser les rendements financiers
- Soit Y_t le rendement d'un actif
 - $\Rightarrow Y_t^2$ est un proxy de la volatilité
 - \Rightarrow Fait stylisé : la volatilité est autocorrélée
- Les modèles MS rendent compte de ce fait stylisé via :

$$\rho_{Y^2}(h) = \frac{\mathbb{E}(Y_t^2 Y_{t+h}^2 | \theta) - \mathbb{E}(Y_t^2 | \theta)^2}{\mathbb{E}(Y_t^4 | \theta) - \mathbb{E}(Y_t^2 | \theta)^2} \quad (13)$$

avec $\mathbb{E}(Y_t^k | \theta) = \sum_{i=1}^N \mathbb{E}(Y_t^k | \theta_i) \pi_i$ et $\mathbb{E}(Y_t^2 Y_{t+h}^2 | \theta)$ donné par Eq. (10)

Modèle de mélange Markovien à deux états en finance

- Illustrons le résultat de Timmermann (2000)
- Soit un modèle MS moyenne-variance à deux états

⇒ Pour $\mathbb{E}(Y_t^2 Y_{t+h}^2 | \theta)$ on obtient

$$\mathbb{E}(Y_t^2 Y_{t+h}^2 | \theta) = \mathbb{E}(Y_t^2 | \theta) + \pi_1 \pi_2 (\mu_1^2 - \mu_2^2 + \sigma_1^2 - \sigma_2^2)^2 \lambda_2^h$$

⇒ On en déduit la fonction d'autocorrélation de Y_t^2

$$\rho_{Y^2}(h) = \frac{\pi_1 \pi_2 (\mu_1^2 - \mu_2^2 + \sigma_1^2 - \sigma_2^2)^2}{\mathbb{E}(Y_t^4 | \theta) - \mathbb{E}(Y_t^2 | \theta)^2} \lambda_2^h$$

- Y_t^2 est autocorrélé si $p_{11} > p_{21}$, i.e. si $p_{11} + p_{22} > 1$
- Sous des conditions plus complexes on peut montrer que Y_t^2 peut générer de l'hétéroscédasticité (même si $\sigma_1^2 = \sigma_2^2$)

⇒ Intéressant pour tenir compte de “l'integrated quarticity”

- Si $\mu_1 = \mu_2$ et $p_{11} + p_{22} > 1$, Y_t on a $\rho_Y(h) = 0$ mais $\rho_{Y^2}(h) \neq 0$ si $\sigma_1 \neq \sigma_2$
- Pour l'IQ cela provient de l'expression du moment d'ordre 4 mais trop compliqué

Modèle MS-régression

- Goldfeld et Quandt (1973)

$$Y_t = \beta_{S_t} \mathbf{x}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_{S_t}) \quad (14)$$

avec $\beta_{S_t} = (\beta_{1S_t}, \dots, \beta_{kS_t})$ et $\mathbf{x}_t = (x_{1t}, \dots, x_{kt})$ un vecteur de k variables exogènes

- Il s'agit d'une **seconde généralisation** car c'est ici le coefficient de régression qui dépend de S_t
- Pour des raisons encore inexplorées, ce modèle engendre un problème d'identification si les variables \mathbf{x}_t sont discrètes

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Modèle MS autorégressif

- Une **troisième généralisation** : la composante autorégressive
- Soit $\mathbf{S} = \{S_t\}_{t=0}^n$ une chaîne de Markov définie sous **H2**
- Soit $\mathbf{Y} = \{Y_t\}_{t=1}^n$ une séquence de variables aléatoires
- La densité $p(\mathbf{S}, \mathbf{y}|\theta)$ de la distribution jointe de \mathbf{S} et \mathbf{Y} est

$$p(\mathbf{S}, \mathbf{y}|\theta) = p(S_0|\theta) \prod_{t=1}^n p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \theta) p(S_t|\mathbf{S}^{t-1}, \mathbf{y}^{t-1}, \theta) \quad (15)$$

- $p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \theta)$ est la densité prédictive de la distribution conditionnelle de Y_t sachant l'information
 - des réalisations passées de Y_t : $\mathbf{y}^{t-1} = (y_1, \dots, y_{t-1})$
 - des états passés et présent : $\mathbf{S}^t = (S_0, \dots, S_t)$

Modèle MS autorégressif

- Sous **H3** la densité $p(y_t|\theta)$ n'intègre pas \mathbf{y}^{t-1}

Hypothèse

H4 : Seule la valeur présente de S_t influence la densité $p(y_t|\mathbf{y}^{t-1}, \mathbf{s}^t; \theta)$ qui peut dépendre de variables exogènes \mathbf{x}_t

$$p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t; \theta) = p(y_t|\mathbf{y}^{t-1}, S_t; \theta)$$

- Sous **H4** le modèle de McCulloch et Tsay (1994) est valide (nous allons présenter ce modèle dans quelques slides)

Modèle MS autorégressif

Hypothèse

H5 : Seule la valeur présente et un nombre fini de valeurs passées de S_t influencent la densité $p(y_t|\mathbf{y}^{t-1}, \mathbf{s}^t; \theta)$ qui peut dépendre de variables exogènes \mathbf{x}_t

$$p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t; \theta) = p(y_t|\mathbf{y}^{t-1}, S_t, \dots, S_{t-p}; \theta), \quad p < \infty$$

- Sous **H5** le modèle de Hamilton (1989) est valide (nous allons présenter ce modèle dans quelques slides)

Le modèle d'Hamilton permet de clarifier pourquoi on a besoin des p retards sur S_t à la différence des modèles sous H4

Modèle MS autorégressif

Hypothèse

H6 : S_t est une chaîne de Markov *inhomogène*. La distribution conditionnelle de S_t dépend de S_{t-1} et d'une variable exogène d'information $\mathbf{z}^t = (z_t, z_{t-1}, \dots, z_1)'$

$$\mathbb{P}(S_t = k | \mathbf{S}^{t-1}, \mathbf{y}^{t-1}) = \mathbb{P}(S_t = k | S_{t-1}, \mathbf{z}^t), \quad \forall k \in \mathbb{S}$$

- Il est possible de considérer $\mathbf{z}^t = \mathbf{y}^{t-1}$
- Sous **H6** le MS-TVTP de Filardo (1994) est valide (nous allons présenter ce modèle dans quelques slides)

- Sous **H3**, on peut spécifier la densité de la distribution jointe pour les modèles simples du type MS-mean ou MS-mean-variance
 - Sous **H4**, on peut spécifier la densité de la distribution jointe pour les modèles à la McCulloch du type MS-AR
 - Sous **H5**, on peut spécifier la densité de la distribution jointe pour les modèles à la Hamilton du type MS-AR
 - Sous **H6**, on peut spécifier la densité de la distribution jointe pour les modèles plus sophistiqués comme les MS-ARMA ou les MS-GARCH
- Sous **H6** la densité dépend de \mathbf{y}^{t-1} et tous les états passés. Cela crée un problème de path-dependence

Modèle MS autorégressif

- Sous **H5**, un premier type de modèle MS-AR est développé

- Hamilton (1989)

$$Y_t - \mu_{S_t} = \sum_{k=1}^p \varphi_k (Y_{t-k} - \mu_{S_{t-k}}) + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (16)$$

\Rightarrow un changement de S_{t-1} vers $S_t \neq S_{t-1}$ induit un changement **abrupt** de la moyenne $\mu_{S_{t-1}}$ vers μ_{S_t}

- Sous **H4**, un second type de modèle MS-AR est développé

- McCulloch et Tsay (1994)

$$Y_t = \mu_{S_t} + \sum_{k=1}^p \varphi_k Y_{t-k} + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (17)$$

\Rightarrow un changement de S_{t-1} vers $S_t \neq S_{t-1}$ induit un changement **progressif** de la moyenne $\mu_{S_{t-1}}$ vers μ_{S_t}

- Chez Hamilton (1989), le changement est abrupt car au sein de chaque régime, la variable est centrée et se comporte comme un bruit... Cela implique que le switch se fait ressentir via μ_{S_t} et que la transition d'un régime à l'autre est abrupte
- Chez McCulloch et Tsay (1994), à l'inverse, le régime précédent peut impacter la dynamique de Y_t car il est présent dans Y_{t-1}
- On peut également réécrire le modèle de McCulloch et Tsay (1994) ainsi

$$Y_t = \mu_0 + \mu_1 S_t + \sum_{k=1}^p \varphi_k Y_{t-k} + (\sigma_0 + \sigma_1 S_t) \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1)$$

avec $S_t = \{0, 1\}$

Modèle MS autorégressif

- Le modèle proposé par Hamilton (1989) est complexe
 - ⇒ la densité de Y_t dépend des valeurs passées de Y_t et S_t
 - d'où la nécessité de supposer **H5**
- Le modèle proposé par McCulloch et Tsay (1994) est plus simple
 - ⇒ la densité de Y_t dépend des valeurs passée de Y_t seulement
 - d'où la nécessité de supposer uniquement **H4**

Modèle MS autorégressif

- Une **ultime généralisation** : changement de régime dans la composante autorégressive
 - variante à la Hamilton (1989)

$$Y_t - \mu_{S_t} = \sum_{k=1}^p \varphi_{k,S_t} (Y_{t-k} - \mu_{S_{t-k}}) + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (18)$$

- variante à la McCulloch et Tsay (1994)

$$Y_t = \mu_{S_t} + \sum_{k=1}^p \varphi_{k,S_t} Y_{t-k} + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (19)$$

Propriétés du MS-AR

- On notera MS(N)-AR(p) un processus à N états et p retards
- Timmermann (2000) dérive les moments de Hamilton (1989)
- e.g. pour un MS(2)-AR(1) à variance et coefficients AR constants
 - Variance non-conditionnelle

$$\mathbb{V}(Y_t|\theta) = (\mu_1 - \mu_2)^2 \pi_1 \pi_2 + \frac{\sigma_\varepsilon^2}{1 - \varphi_1^2}$$

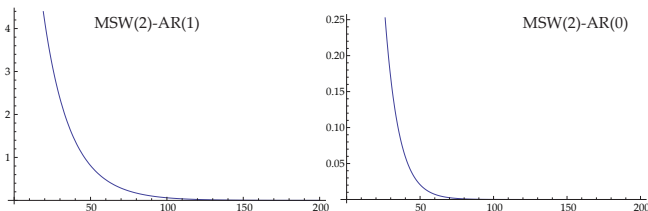
- Autocovariance non-conditionnelle

$$\gamma_Y(h) = (p_{11} - p_{21})^h (\mu_1 - \mu_2)^2 \pi_1 \pi_2 + \varphi_1^h \frac{\sigma_\varepsilon^2}{1 - \varphi_1^2}$$

- Autocorrelation non-conditionnelle

$$\rho_Y(h; \theta) = \frac{(p_{11} - p_{21})^h (\mu_1 - \mu_2)^2 \pi_1 \pi_2 + \varphi_1^h \frac{\sigma_\varepsilon^2}{1 - \varphi_1^2}}{(\mu_1 - \mu_2)^2 \pi_1 \pi_2 + \frac{\sigma_\varepsilon^2}{1 - \varphi_1^2}}$$

Propriété du MS en moyenne-variance



- Fonction $\gamma_Y(h)$, du MS(2)-AR(1) et du MS(2)-AR(0)
 - à gauche : $p_{12} = 0.05$, $p_{21} = 0.05$ et $\mu_1 = 2$, $\mu_2 = -2$ et $\varphi_1 = 0.95$
 - à droite : $p_{12} = 0.05$, $p_{21} = 0.05$ et $\mu_1 = 2$, $\mu_2 = -2$
- Dans le MS(2)-AR(1) il y a deux sources de persistance :
 - les probabilités π_1 et π_2
 - le coefficient autorégressif φ_1

Parler de nonstationnarité locale ? Que ce passe-t-il si $\varphi_{1,S_t} > 1$? Tant que la chaîne de Markov est stationnaire, le processus Y_t est stationnaire même si dans l'un des régimes $\varphi_{1,S_t} > 1$.

Donc, ergodicité de S_t fondamentale

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Modèle MS(N)-AR(p)-TVTP(k)

- Sous **H6**, Filardo (1994) suggère que les probabilités de transition peuvent varier dans le temps
- En partant du modèle de Hamilton (1989)

$$Y_t - \mu_{S_t} = \sum_{k=1}^p \varphi_{k,S_t} (Y_{t-k} - \mu_{S_{t-k}}) + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (20)$$

Filardo (1994) spécifie une nouvelle matrice de transition

$$P(\mathbf{z}_t) = \begin{pmatrix} p_{11}(\mathbf{z}_t) & 1 - p_{11}(\mathbf{z}_t) \\ 1 - p_{22}(\mathbf{z}_t) & p_{22}(\mathbf{z}_t) \end{pmatrix}$$

- $\mathbf{z}_t = z_t, \dots, z_{t-k}$ représente l'historique d'une variable d'information qui va guider les probabilités de transition

$\Rightarrow P(\mathbf{z}_t)$ va varier dans le temps car \mathbf{z}_t varie dans le temps

Modèle MS(N)-AR(p)-TVTP(k)

- Spécifications possibles pour $p_{ij}(\mathbf{z}_t)$

⇒ toute fonction monotone (généralement sigmoïde)

⇒ la fonction logistique est fréquemment utilisée

$$p_{ij}(\mathbf{z}_t) = \frac{\exp\left(\psi_{ij,0} + \sum_{l=1}^k \psi_{ij,l} z_{t-l}\right)}{1 + \exp\left(\psi_{ij,0} + \sum_{l=1}^k \psi_{ij,l} z_{t-l}\right)} \quad (21)$$

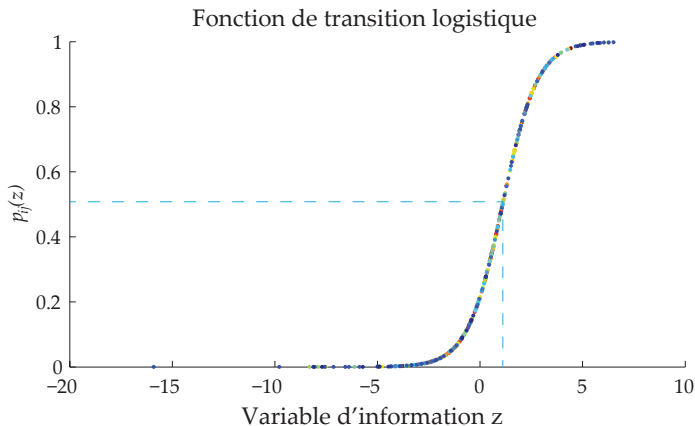
⇒ Si $\mathbf{z}_t = \mathbf{y}_{t-l}$, $l \geq 1$, on parle de Markov-switching endogène (voir Kim et al. , 2008)

- Notons que $\mathbf{z} = \mathbf{0}$ nous ramène au modèle de Hamilton (1989)

$$p_{ij}(\mathbf{z} = \mathbf{0}) = \frac{\exp\left(\psi_{ij,0}\right)}{1 + \exp\left(\psi_{ij,0}\right)} = p_{ij}, \quad \forall t \quad (22)$$

⇒ on peut spécifier p_{ij} avec une fonction logistique : lisse la transition

Modèle MS(N)-AR(p)-TVTP



- Selon les valeurs de z_t , la probabilité p_{ij} est plus ou moins grande

Modèle MS-VAR

- Phillips (1991) suggère une généralisation du MS-AR vers des modèles multivariés : MS-VAR
- e.g. d'un MS(N)-VAR(1) à k équations

$$\mathbf{Y}_t - \boldsymbol{\mu}_{\mathbf{S}_t} = \boldsymbol{\Phi}(\mathbf{Y}_{t-1} - \boldsymbol{\mu}_{\mathbf{S}_{t-1}}) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}) \quad (23)$$

avec

- \mathbf{S}_t un vecteur composé de k chaîne de Markov indépendantes
 - $\boldsymbol{\Phi}$ et $\boldsymbol{\Sigma}$ des matrices $k \times k$
 - $\boldsymbol{\mu}$ un vecteur $k \times 1$
- Krolzig (1997) suggère un MS(N)-VAR(1) bien plus complexe

$$\mathbf{Y}_t = \boldsymbol{\mu}_{\mathbf{S}_t} + \boldsymbol{\Phi}_{\mathbf{S}_t} \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{S}_t}) \quad (24)$$

où la chaîne de Markov \mathbf{S}_t affecte $\boldsymbol{\Phi}_{\mathbf{S}_t}$, $\boldsymbol{\mu}_{\mathbf{S}_t}$ et $\boldsymbol{\Sigma}_{\mathbf{S}_t}$

Il existe bien d'autres extensions :

- le MS-VECM pour tenir compte de nonlinéarité dans l'analyse de la cointégration
- le MS de [Otranto \(2005\)](#) où les probabilités de transition dépendent du régime d'une autre variable
- le MS-GARCH
- ...

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Méthodes d'estimation

- L'estimation des modèles MS(N)-AR(p) est complexe
- On distingue trois types d'approche pour l'estimation
 - L'estimation par **maximum de vraisemblance (ML) exact**
 - L'estimation via l'algorithme espérance-maximisation (EM)
 - L'approche bayésienne

⇒ Nous n'aborderons que l'approche par **ML exact**

Note Par la suite, θ représentera

- soit les paramètres du modèle
- soit les paramètres de la matrice de transition
- soit les paramètres du modèle et de la matrice de transition

Rappel des notations

- Soit $\mathbf{S} = \{S_t\}_{t=0}^n$ un processus de Markov à N états
- Soit $\mathbf{Y} = \{Y_t\}_{t=1}^n$ une variable aléatoire
- Les réalisations de \mathbf{Y} sont notés $\mathbf{y} = (y_1, \dots, y_n)$
- $\mathbf{y}^{t-1} = (y_1, \dots, y_{t-1})$ désigne les réalisations passées de Y_t
- $p(\mathbf{y}, \mathbf{S}|\theta)$ désigne la densité jointe de $\mathbf{y} = \{y_t\}_{t=1}^n$ et $\mathbf{S} = \{S_t\}_{t=0}^n$
- $p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \theta)$ désigne la fonction de densité de la distribution de Y_t conditionnée sur les réalisations passées \mathbf{y}^{t-1} et sachant \mathbf{S}^t
- $p(S_t|\mathbf{S}^{t-1}, \mathbf{y}^{t-1}, \theta)$ désigne la fonction de densité de la distribution de S_t conditionnée sur les réalisations passées \mathbf{y}^{t-1} et sachant \mathbf{S}^{t-1}

Inférence statistique

- On suppose S_t sous **H2** et Y_t sous **H5**
 - Sous **H6** les formules se complexifient mais restent valides
- Rappelons également la fonction de densité jointe établie en (15)

$$p(\mathbf{S}, \mathbf{y}|\theta) = p(S_0|\theta) \prod_{t=1}^n p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \theta) p(S_t|\mathbf{S}^{t-1}, \mathbf{y}^{t-1}, \theta)$$

avec $p(S_t|.)$ la densité de la **distribution conditionnelle** de S_t

- Plusieurs cas de figures se présentent :

Cas 1 Inférer \mathbf{S} si θ connu

Cas 2 Estimation de θ si \mathbf{S} connu

Cas 3 Estimation de \mathbf{S} et θ

- Le cas 1 est théorique car il est peut probable de rencontrer une situation où les paramètres du modèle sont connus
- Le cas 2 est parfois envisageable en pratique : e.g. avec les régimes de change que l'on observe si on décide de faire confiance aux déclarations de jure
- Le cas 3 est le plus fréquent car dans la pratique on observe souvent y_t mais ni les paramètres du modèle, ni la chaîne de Markov

Cas 1 : Inférer \mathbf{S} si θ connu

- Il s'agit d'inférer la distribution de probabilité du processus inobservé S_t en t
- On note $\mathbb{P}(S_t = j | \mathbf{y}^\tau, \theta)$ la **distribution de S_t conditionnelle** à \mathbf{y}^τ
- $\mathbf{y}^\tau = (y_1, \dots, y_\tau)$ est observé et révèle de l'information si $t \leq \tau$
- $\mathbb{P}(S_t = j | \mathbf{y}^\tau, \theta)$ est une distribution prédictive de S_t pour $t > \tau$
- Concernant $t \leq \tau$ on distingue deux cas :
 - $t = \tau$: on parle de probabilités filtrées $\mathbb{P}(S_t = j | \mathbf{y}^t, \theta)$
 - $t < \tau$: on parle de probabilités lissées $\mathbb{P}(S_t = j | \mathbf{y}^\tau, \theta)$

Filtrage des probabilités

- Pour évaluer $\mathbb{P}(S_t = j | \mathbf{y}^t, \theta)$ on va appliquer l'algorithme suivant

step 1 Prédiction de S_t (one step ahead)

$$\mathbb{P}(S_t = j | \mathbf{y}^{t-1}, \theta) = \sum_{i=1}^N P_{ij}(t-1) \mathbb{P}(S_{t-1} = i | \mathbf{y}^{t-1}, \theta) \quad (25)$$

avec $P_{ij}(t-1) = \mathbb{P}(S_t = j | S_{t-1} = i, \mathbf{y}^{t-1}, \theta) = P_{ij}$ si S_t est homogène

step 2 Filtration de S_t

$$\mathbb{P}(S_t = j | \mathbf{y}^t, \theta) = \frac{p(y_t | S_t = j, \mathbf{y}^{t-1}, \theta) \mathbb{P}(S_t = j | \mathbf{y}^{t-1}, \theta)}{p(y_t | \mathbf{y}^{t-1}, \theta)} \quad (26)$$

avec $p(y_t | \mathbf{y}^{t-1}, \theta) = \sum_{i=1}^N p(y_t | S_t = i, \mathbf{y}^{t-1}, \theta) \mathbb{P}(S_t = i | \mathbf{y}^{t-1}, \theta)$

- En $t = 1$, on initialise le filtre à $\mathbb{P}(S_0 = k | P)$ et donc

$$\mathbb{P}(S_1 = l | \mathbf{y}^0, \theta) = \sum_{k=1}^N P_{kl}(0) \mathbb{P}(S_0 = k | P)$$

L'équation (26) exprime la probabilité d'observer le numérateur sur l'ensemble des possibilités

Cela relève aussi d'un théorème de Bayes : $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Filtrage des probabilités

■ Résumé de l'algorithme

filtrage en $t - 1 : \mathbb{P}(S_{t-1} = i | \mathbf{y}^{t-1}, \theta)$

\Downarrow

prédiction pour $t : \mathbb{P}(S_t = i | \mathbf{y}^{t-1}, \theta)$

\Downarrow

données observées en $t : y_t$

\Downarrow

filtrage en $t : \mathbb{P}(S_t = i | \mathbf{y}^t, \theta)$

En gros il s'agit d'un outil d'inférence adaptatif. En $t-1$ et à θ fixé, $\mathbb{P}(S_{t-1} = i | \mathbf{y}^{t-1}, \theta)$ résume l'information des observations apportées par y_1, \dots, y_{t-1} sur S_{t-1} . On peut voir $\mathbb{P}(S_t = j | \mathbf{y}^{t-1}, \theta)$ comme une distribution à priori (prior) et $\mathbb{P}(S_t = j | \mathbf{y}^t, \theta)$ comme la distribution à posteriori qui sert à corriger la prédiction $\mathbb{P}(S_t = j | \mathbf{y}^{t-1}, \theta)$.

Lissage des probabilités

- Le calcul des probabilités par filtrage s'opère en t sachant l'information \mathbf{y}^t
 - On utilise l'information jusqu'à t (i.e. l'information connue)
- Mais il est préférable d'intégrer l'information $\mathbf{y} = (y_1, \dots, y_n)$
 - Possible via un algorithme de lissage qu'on applique sur \mathbf{y}
- Hamilton (1989) propose d'utiliser la distribution jointe de S_t et S_n pour calculer la probabilité lissée

$$\mathbb{P}(S_t = j | \mathbf{y}, \theta) = \sum_{i=1}^N \mathbb{P}(S_t = j, S_n = i | \mathbf{y}, \theta)$$

où $\mathbb{P}(S_t = j, S_n = i | \mathbf{y}, \theta)$ s'obtient récursivement via l'algorithme de filtrage

Lissage efficient des probabilités

- Le lissage proposé par Hamilton (1989) n'est pas très efficace et on lui préférera le lissage suivant

step 1 Récupérer les probabilités filtrées : $\mathbb{P}(S_t = i | \mathbf{y}^t, \theta)$

step 2 Initialiser l'algorithme de lissage en partant de n : $\mathbb{P}(S_n = i | \mathbf{y}, \theta)$

step 3 Pour chaque $t = n - 1, \dots, 0$, la distribution de probabilité lissée est calculée ainsi

$$\mathbb{P}(S_t = i | \mathbf{y}, \theta) = \sum_{j=1}^N \frac{P_{ij} \mathbb{P}(S_t = i | \mathbf{y}^t, \theta) \mathbb{P}(S_{t+1} = j | \mathbf{y}, \theta)}{\sum_{k=1}^N P_{kj} \mathbb{P}(S_t = k | \mathbf{y}^t, \theta)}$$

Cas 2 : Estimation de θ si \mathbf{S} connu

- Dans ce cas de figure, on observe $\mathbf{y} = (y_1, \dots, y_n)$ et $\{S_t\}_{t=0}^n$
- En revanche, θ est inobservé
 - θ regroupe les paramètres de la fonction de densité et de P
- La fonction de vraisemblance découle de la densité jointe

$$p(\mathbf{y}, \mathbf{S}|\theta) = p(\mathbf{y}|\mathbf{S}, \theta)p(\mathbf{S}|\theta)$$

- En développant $p(\mathbf{y}|\mathbf{S}, \theta)$ et $p(\mathbf{S}|\theta)$ on obtient

$$p(\mathbf{y}, \mathbf{S}|\theta) = \prod_{k=1}^N \left(\prod_{t:S_t=k} p(y_t|\mathbf{y}^{t-1}, \theta_k) \right) \prod_{j=1}^N \prod_{k=1}^N P_{jk}^{C_{jk}(\mathbf{S})} p(S_0|\theta) \quad (27)$$

- Démontrons ce résultat...

Cas 2 : Estimation de θ si \mathbf{S} connu

- Sous **H2**, la densité $p(\mathbf{S}|\theta)$ est donnée par

$$\begin{aligned} p(\mathbf{S}|\theta) &= \prod_{t=1}^n p(S_t|S_{t-1}, \theta) p(S_0|\theta) \\ &= \prod_{t=1}^n (P_{S_{t-1}S_t}) p(S_0|\theta) = \prod_{j=1}^N \prod_{k=1}^N P_{jk}^{C_{jk}(\mathbf{S})} p(S_0|\theta) \end{aligned}$$

- $C_{jk}(\mathbf{S})$ compte le nombre de transition de j vers k

$$C_{jk}(\mathbf{S}) = \#\{S_{t-1} = j, S_t = k\}, \quad \forall j, k \in \{1, \dots, N\}$$

- La densité de \mathbf{y} sachant que \mathbf{S} est observé est donnée par

$$p(\mathbf{y}|\mathbf{S}, \theta) = \prod_{t=1}^n p(y_t|S_t, \mathbf{y}^{t-1}, \theta) = \prod_{k=1}^N \left(\prod_{t: S_t=k} p(y_t|\mathbf{y}^{t-1}, \theta_k) \right)$$

- L'estimateur du maximum de vraisemblance suit :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L_n(\theta; \mathbf{y}, \mathbf{S}), \quad L_n(\theta; \mathbf{y}, \mathbf{S}) = \log p(\mathbf{y}, \mathbf{S}|\theta)$$

L'écriture de $\prod_{t:S_t=k} p(y_t|\mathbf{y}^{t-1}, \theta_k)$ s'interprète ainsi : on fait le produit sur t tant que $S_t = k$...

On voit alors que l'expression $\prod_{k=1}^N \left(\prod_{t:S_t=k} p(y_t|\mathbf{y}^{t-1}, \theta_k) \right)$ fait cela pour tout $k \in \mathbb{S}$ et permet d'obtenir $p(\mathbf{y}|\mathbf{S}, \theta)$

Cas 3 : Estimation de θ et \mathbf{S}

- Dans ce cas de figure, on observe uniquement $\mathbf{y} = (y_1, \dots, y_n)$
- $\{S_t\}_{t=0}^n$ et θ sont inobservés
- Une solution triviale serait d'étendre la vraisemblance du Cas 2 pour couvrir l'ensemble des réalisations possibles de \mathbf{S} :

$$\mathbb{S}^{n+1} = \{1, \dots, N\}^{n+1}$$

- Il en découlerait la vraisemblance exacte suivante

$$\begin{aligned} p(\mathbf{y}|\theta) &= \sum_{\mathbf{S} \in \mathbb{S}^{n+1}} p(\mathbf{y}|\mathbf{S}, \theta_1, \dots, \theta_N) p(\mathbf{S}|\theta) \\ &= \sum_{\mathbf{S} \in \mathbb{S}^{n+1}} \prod_{t=1}^n p(y_t | \mathbf{y}^{t-1}, \theta_{S_t}) \prod_{j=1}^N \prod_{k=1}^N P_{jk}^{C_{jk}(\mathbf{S})} p(S_0|\theta) \quad (28) \end{aligned}$$

- Numériquement cette solution n'est pas viable car \mathbb{S}^{n+1} est gigantesque

- Attention, dans le cas 2 nous avons $\prod_{t=1}^n p(y_t|S_t, \mathbf{y}^{t-1}, \theta)$ qui maintenant est remplacé par $\prod_{t=1}^n p(y_t|\mathbf{y}^{t-1}, \theta_{S_t})$ car S_t n'est plus observé...
- L'expression $\prod_{t:S_t=k} p(y_t|\mathbf{y}^{t-1}, \theta_k)$ qui intervient dans $p(y_t|\mathbf{y}^{t-1}, \theta_{S_t})$ doit à présent être explorée pour toutes réalisations de \mathbf{S} via $\sum_{\mathbf{S} \in \mathbb{S}^{n+1}}$ afin de couvrir toutes les trajectoires possibles...
- Impossible d'un point de vue calculatoire (sauf via l'algo EM dont on ne parle pas dans le cours)...

Cas 3 : Estimation de θ et S

- La solution est de repartir de la fonction de vraisemblance suggérée par Quandt (1972)

$$p(\mathbf{y}|\theta) = \prod_{t=1}^n \left(\sum_{k=1}^N p(y_t|\theta, S_t = k) \mathbb{P}(S_t = k|\theta) \right)$$

- Cependant cette fonction traite S_t comme indépendant dans le temps et ne s'applique pas pour notre classe de modèle
- On peut résoudre le problème en remplaçant $\mathbb{P}(S_t = k|\theta)$ par la probabilité prédictive $\mathbb{P}(S_t = k|\mathbf{y}^{t-1}, \theta)$ détaillée en (25) :

$$\begin{aligned} p(\mathbf{y}|\theta) &= \prod_{t=1}^n \left(\sum_{k=1}^N p(y_t|S_t = k, \theta) \mathbb{P}(S_t = k|\mathbf{y}^{t-1}, \theta) \right) \\ &= \prod_{t=1}^n p(y_t|\mathbf{y}^{t-1}, \theta) \end{aligned} \quad (29)$$

Cas 3 : Estimation de θ et \mathbf{S}

- Si le modèle possède une composante $\text{AR}(p)$ on aura l'approximation suivante

$$p(\mathbf{y}|\theta) \approx p(y_{p+1}, \dots, y_n | \mathbf{y}^p, \theta) = \prod_{t=p+1}^n p(y_t | \mathbf{y}^{t-1}, S_t, \theta) \quad (30)$$

- On voit alors que pour chaque $t = p+1, \dots, n$, la densité prédictive

$$p(y_t | S_t = k, \mathbf{y}^{t-1}, \theta)$$

apparaît alors comme un facteur constant normalisant de la distribution des probabilités filtrées $\mathbb{P}(S_t = k | \mathbf{y}^{t-1}, \theta)$

- En effet, l'équation (29) devient

$$p(\mathbf{y}|\theta) = \prod_{t=p+1}^n \left(\sum_{k=1}^N p(y_t | S_t = k, \mathbf{y}^{t-1}, \theta) \mathbb{P}(S_t = k | \mathbf{y}^{t-1}, \theta) \right)$$

Cas 3 : Estimation de θ et S

- $\mathbb{P}(S_t = k | \mathbf{y}^{t-1}, \theta)$ s'obtient simplement en appliquant l'algorithme de filtration
- Comme il s'agit d'une **approximation** de la vraisemblance exacte définie en (28), on parle de **quasi-vraisemblance**
- L'estimateur du quasi-maximum de vraisemblance (QML) suit :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L_n(\theta; \mathbf{y}), \quad L_n(\theta; \mathbf{y}) = \log p(\mathbf{y} | \theta)$$

Application au cas du QML Gaussien

- Soit un modèle Markov-Switching à la McCulloch et Tsay (1994)

$$Y_t = \alpha_{S_t} + \sum_{k=1}^p \beta_k Y_{t-k} + \sigma_\varepsilon \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \quad S_t \in \mathbb{S} = \{1, 2\}$$

- Le vecteur des paramètres à estimer est donc

$$\theta = (\alpha_1, \alpha_2, \beta_1, \dots, \beta_p, \sigma_\varepsilon^2, p_{11}, p_{22})'$$

- $\mathbf{y}^t = (y_1, \dots, y_t)$ désigne les réalisations passées et présente de Y_t

Application au cas du QML Gaussien

- Sous $\varepsilon_t \sim \mathcal{N}(0, 1)$, la densité de y_t conditionnelle à S_t et \mathbf{y}^{t-1} est

$$p(y_t | S_t = i, \mathbf{y}^{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(\frac{-(y_t - \alpha_i - \sum_{k=1}^p \beta_k y_{t-k})^2}{2\sigma_\varepsilon^2}\right) \quad (31)$$

- Sachant la probabilité prédictive $\mathbb{P}(S_t = i | \mathbf{y}^{t-1}, \theta)$, on obtient

$$\begin{aligned} p(y_t | \mathbf{y}^{t-1}, \theta) &= \sum_{i=1}^N p(y_t | S_t = i, \mathbf{y}^{t-1}, \theta) \mathbb{P}(S_t = i | \mathbf{y}^{t-1}, \theta) \\ &= p(y_t | S_t = 1, \mathbf{y}^{t-1}, \theta) \mathbb{P}(S_t = 1 | \mathbf{y}^{t-1}, \theta) \\ &\quad + p(y_t | S_t = 2, \mathbf{y}^{t-1}, \theta) \mathbb{P}(S_t = 2 | \mathbf{y}^{t-1}, \theta) \end{aligned} \quad (32)$$

- Les probabilités filtrées de S_t sont

$$\mathbb{P}(S_t = i | \mathbf{y}^t, \theta) = \frac{p(y_t | S_t = i, \mathbf{y}^{t-1}, \theta) \mathbb{P}(S_t = i | \mathbf{y}^{t-1}, \theta)}{p(y_t | \mathbf{y}^{t-1}, \theta)} \quad (33)$$

- Grâce au théorème de Bayes on vient boucler la récursion

$$\mathbb{P}(S_{t+1} = i | \mathbf{y}^t, \theta) = p_{1i} \mathbb{P}(S_t = 1 | \mathbf{y}^t, \theta) + p_{2i} \mathbb{P}(S_t = 2 | \mathbf{y}^t, \theta) \quad (34)$$

avec $p_{1i} = \mathbb{P}(S_{t+1} = i | S_t = 1)$ et $p_{2i} = \mathbb{P}(S_{t+1} = i | S_t = 2)$

Application au cas du QML Gaussien

- En $t = p$ on initialise $\mathbb{P}(S_p = i | \mathbf{y}^{t-1}, \theta)$
 - Hamilton (1994) suggère d'initialiser à l'aide de π (voir 6)
- Ensuite on itère sur les équations (31),(32),(33) et (34) pour obtenir à chaque $t = p, \dots, n$
 - les probabilités filtrées $\mathbb{P}(S_t = 1 | \mathbf{y}^t, \theta)$
 - les densités conditionnelles $p(y_t | \mathbf{y}^{t-1}, \theta)$
- La fonction de quasi-log-vraisemblance est la suivante

$$L_n(\theta; \mathbf{y}) = \frac{1}{n} \sum_{t=p+1}^n \log p(y_t | \mathbf{y}^{t-1}, \theta)$$

- L'estimateur du QML est alors

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L_n(\theta; \mathbf{y})$$

- Pour résumer

- 1 On choisit des valeurs initiales pour θ
- 2 On calcule (31) à l'aide des valeurs initiales du modèle
- 3 On calcule $\mathbb{P}(S_p = i | \mathbf{y}^{p-1}, \theta)$ à l'aide de (25) et des valeurs initiales choisies pour P
- 4 On calcule (32) à l'aide de l'étape 2 et 3
- 5 On calcule (33) à l'aide de l'étape 2, 3 et 4
- 6 On calcule (34) à l'aide de l'étape 5
- 7 On maximise la quasi-log-vraisemblance
- 8 On calcule (31) à l'aide des valeurs issues de la maximisation (remplace l'étape 2)
- 9 On utilise le résultat de l'étape 6 en remplacement de l'étape 3

...

- Pour le MLE Gaussien on pourrait calculer à la main les solutions mais c'est un peu lourd pour eux... les calculs ils sont dans le Hamilton p. 699

- Petite astuce dans la pratique : il y a souvent des problèmes numériques sur les MS que l'on peut pallier en changeant l'échelle des données par un facteur 100 où 1000.

Théorie asymptotique

- Les résultats concernant la théorie limite des MS sont morcelés
- La consistance et la normalité asymptotique est prouvée pour :
 - les modèles MS-simple (8) et (9) sous **H2** par Leroux (1992) pour la consistance et Bickel et al. (1998) pour la normalité
 - le modèle MS-régression (14) sous **H2** par Lindgren (1978)
 - le modèle MS-AR (16) sous **H2** par Douc et al. (2004)
 - le modèle MS-AR (20) sous **H6** par Ailliot et Pene (2015) pour la consistance
 - le modèle MS-VAR (24) sous **H2** par Krolzig (1997)

Problèmes d'estimation

- Concernant les modèles MS avec variance dépendante (i.e. $\sigma_{S_t}^2$), Hamilton (1988) souligne que la vraisemblance est non-bornée
 - e.g. si dans un état i , $\mu_{t,i} = y_t$ on a σ_i qui tend vers 0 et comme σ_i est au dénominateur, $L_n(\theta; \mathbf{y})$ diverge vers ∞
- \Rightarrow pour éviter cela, il faut borner σ_i^2 loin de 0
- On peut montrer que le MLE est confronté à l'existence d'un optimum local au moins
 - cet optimum survient lorsque les états sont numériquement indistinguables (e.g. $\mu_1 = \mu_2$)
- \Rightarrow pour éviter cela, on peut considérer plusieurs set de valeurs initiales

Prévisions

- La prédiction des modèles MS se décompose en deux éléments
 - La prédiction de y_{t+h} sachant le régime en $t+h$ (i.e. S_{t+h})
 - La prédiction des proba d'apparition de chaque régime en $t+h$
- Exemple de prévision de y_{t+1} avec $S_t = \{1, 2\}$
 - Pour un modèle à la McCulloch et Tsay (1994) (voir 19) et $p = 1$

$$\begin{aligned}\hat{y}_{t+1|t} &= \mathbb{E}(y_{t+1}|s_{t+1} = 1, \mathbf{y}^t, \theta) \mathbb{P}(S_{t+1} = 1|\mathbf{y}^t, \theta) \\ &\quad \times \mathbb{E}(y_{t+1}|s_{t+1} = 2, \mathbf{y}^t, \theta) \mathbb{P}(S_{t+1} = 2|\mathbf{y}^t, \theta)\end{aligned}$$

avec $\mathbb{P}(S_{t+1} = i|\mathbf{y}^t, \theta)$ donné par (34) et

$$\mathbb{E}(y_{t+1}|s_{t+1} = i, \mathbf{y}^t, \theta) = (\mu_i + \phi_{1,i}y_t)/\sigma_i^2$$

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Procédures de tests

- Plusieurs tests peuvent être intéressants pour l'économètre
 - 1 tests de structure : e.g. la structure non-linéaire du modèle est-elle plus appropriée qu'un modèle linéaire ?
 - 2 tests de diagnostic : e.g. variable omise ?
- Le premier type de test implique des difficultés importantes
 - pour détailler cela considérons le modèle suivant

$$Y_t - \mu_{S_t} = \sum_{k=1}^p \varphi_k (Y_{t-k} - \mu_{S_{t-k}}) + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (35)$$

avec $S_t = \{1, 2\}$, sans perte de généralité

Problèmes liés aux tests

- Si on souhaite tester la linéarité du modèle, sous H_0 on a
 - soit $\mu_1 = \mu_2$ et $\sigma_1 = \sigma_2$
 - soit $p_{11} = 1$ et $p_{21} = 1$
 - soit $p_{12} = 1$ et $p_{22} = 1$
- Or,
 - si $\mu_1 = \mu_2$ et $\sigma_1 = \sigma_2$, la matrice P n'est pas identifiée
 - si $p_{11} = 1$ et $p_{21} = 1$, μ_2 et σ_2 ne sont pas identifiées
 - si $p_{12} = 1$ et $p_{22} = 1$, μ_1 et σ_1 ne sont pas identifiées

Problèmes liés aux tests

- Cela pose un problème car la théorie limite du MLE repose notamment sur l'hypothèse que
 - la matrice d'information de Fisher est non-singulière
- Or la matrice d'information dépend (de la variance) du score
 - en l'absence d'identification du paramètre, le score par rapport à ce paramètre est nul et renvoie un optimum local
- ⇒ l'hypothèse de non-singularité est violée
- Un autre problème est que si $\mu_i = 0$, pour des raisons techniques, la matrice d'information est également singulière
 - ⇒ même sous l'alternative (modèle non-linéaire), des optimum locaux peuvent apparaître

Les tests de type Likelihood Ratio : Garcia (1998)

- Le test LR de Garcia (1998) est utile pour tester
 - ⇒ linéaire v.s. non-linéaires
- Il repose sur un ratio entre les vraisemblances des modèles non-contraint et contraint
 - ⇒ sous la nulle le test LR suit une distribution χ^2
- La théorie asymptotique des tests LR ne s'applique pas ici
 - ⇒ repose notamment sur la non-singularité du score
- Hansen (1992) : théorie limite avec conditions non-standards
 - ⇒ Hansen (1992) propose un test LR valable pour les modèles MS mais lourd à mettre en place
 - ⇒ Garcia (1998) améliore le test LR de Hansen (1992)

Les tests de type LR : Garcia (1998)

- Partons d'un modèle à la McCulloch et Tsay (1994)

$$Y_t = \mu_0 + \mu_1 S_t + \sum_{k=1}^p \varphi_k Y_{t-k} + (\sigma_0 + \sigma_1 S_t) \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (36)$$

avec $S_t = \{0, 1\}$, sans perte de généralités

- Garcia (1998) considère :

- $\gamma = (p_{12}, p_{22})'$ comme des paramètres d'intérêts

$$\Rightarrow H_0 : \gamma = (0, 0)'$$

- $\theta = (\mu_0, \sigma_0, \mu_1, \sigma_1, \varphi_1, \dots, \varphi_p)'$ comme des paramètres de nuisance

Les tests de type LR : Garcia (1998)

- Soit $\{\Gamma \in \mathbb{R} | 0 < \Gamma < 1\} = (0, 1)$ de sorte que $\gamma \in \Gamma$
- Sur Γ , Garcia (1998) dérive la statistique LR :

$$LR_n(\gamma) = 2n \left(L_n(\hat{\theta}(\gamma), \gamma) - L_n(\hat{\theta}) \right)$$

- La statistique de test est donnée par le supremum de $LR_n(\gamma)$:

$$LR_n = \sup_{\gamma \in \Gamma} LR_n(\gamma)$$

avec $LR_n(\gamma)$ évaluée pour différentes valeurs de $\gamma \in \Gamma$

- LR_n suit en limite un **processus** du χ^2 de covariance Ξ

- Un processus du χ^2 est un produit de processus Gaussiens

Les tests de type LR : Garcia (1998)

- Garcia (1998) dérive Ξ pour différentes spécifications
 - les calculs sont lourds et ne s'appliquent pas au TVTP
 - Garcia (1998) propose des simulations en échantillons finis
 - tabulation de valeurs critiques pour différents modèles
 - Cependant, Garcia (1998) ignore le problème $\mu_i = 0$
 - le score est nul dans cette situation
- ⇒ Il est préférable d'utiliser le test de Carrasco et al. (2014)

Les tests optimaux : Carrasco et al. (2014)

- Carrasco et al. (2014) se placent dans un cadre très général

- Soit les réalisations d'un processus : y_1, \dots, y_n

- Soit $p_t(\cdot)$ la fonction de densité de y_t sachant y_{t-1}, \dots, y_1

- Soit θ les paramètres dont $p_t(\cdot)$ dépend

- Carrasco et al. (2014) test l'hypothèse suivante :

- $H_0 : \theta_t = \theta_0$

- $H_1 : \theta_t = \theta_0 + \eta_t$ avec η_t une variable guidant les régimes

⇒ ici η_t est une chaîne de Markov latente de matrice de transition P

⇒ la distribution de η_t dépend des paramètres affectés par le changement de régime

Par exemple, dans le modèle $Y_t = \mu_0 + \mu_1 S_t + \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, on a

$$\eta_t = (\mu_1 S_t, 0, 0)'$$

avec les deux 0 qui tiennent compte de μ_0 et σ afin que η_t et θ_0 aient la bonne dimension

Les tests optimaux : Carrasco et al. (2014)

■ Hypothèses du test

- η_t est ergodique et γ n'est pas identifié sous H_0
- $\gamma \in \Gamma$ avec Γ un ensemble compact
- $\theta_t \in \Theta$ avec Θ un sous ensemble compact de $\mathbb{R}^{\dim(\theta)}$ contenant θ_0 , le vecteur des vrais valeurs des paramètres
- la vraisemblance jointe de \mathbf{y}^t et $\boldsymbol{\eta}^t$ est décomposable

$$p(\mathbf{y}|\theta) = \prod_{t=1}^n p(y_t|\mathbf{y}^{t-1}, \theta) p(\eta_t|\boldsymbol{\eta}^{t-1}, \gamma)$$

- la log-densité de y_t , notée $l_t = l_t(\theta) = \log(p(y_t|\theta))$ satisfait

$$\mathbb{E} \left(\sup_{\theta \in \mathcal{N}} \left(\|l_t^{(k)}(\theta)\|^2 \right) \right) < \infty, \quad k = 1, \dots, 5$$

avec $l_t^{(k)}$ la k -ième dérivée de l_t par rapport à θ , \mathcal{N} un voisinage de θ_0 et $\|\cdot\|$ la norme de Frobenius

- Explication sur la notion de vraisemblance décomposable : le modèle de McCulloch et Tsay (1994) décrit par l'équation (36) donne un bon exemple car on voit que si $\mu_1 S_t = 0$ et $\sigma_1 S_t = 0$ on retrouve le modèle sans changement de régime ce qui permet l'écriture $\theta_t = \theta_0 + \eta_t$ avec $\eta_t = (\mu_1 S_t, \sigma_1 S_t, 0, 0, \mathbf{0})'$ et également de ne pas faire apparaître η_t dans la densité conditionnelle de y_t (p.s. pour que les vecteurs θ_0 et η_t possèdent les bonnes dimensions, on ajoute $\mathbf{0}$, un vecteur de 0 pour tenir compte des paramètres autorégressif et les deux autres 0 pour tenir compte de μ_0 et σ_0)
- Il existe d'autres hypothèses vraiment techniques qu'il vaut mieux ne pas présenter car ça dépasse déjà le niveau M2 je pense...
- Rappel norme de Frobenius : norme euclidienne dans un espace matricielle : $\|A\| := (\text{tr } A^* A)^{1/2} = (\text{tr } A A^*)^{1/2}$ avec A^* la matrice adjointe

Les tests optimaux : Carrasco et al. (2014)

- Le test de Carrasco et al. (2014) repose sur

$$TS_n(\gamma) = TS_n(\gamma, \hat{\theta}) = \Lambda_n - \frac{1}{2n} \hat{e}(\gamma)' \hat{e}(\gamma) \quad (37)$$

avec

$$\Lambda_n = n^{-1/2} \sum_t \lambda_t(\gamma, \hat{\theta}) \quad (38)$$

et

$$\begin{aligned} \lambda_t(\gamma, \hat{\theta}) = & \frac{1}{2} \left(\text{tr} \left((l_t^{(2)} + l_t^{(1)} l_t^{(1)'}) \mathbb{E}(\eta_t \eta_t') \right) \right. \\ & \left. + 2 \sum_{s < t} \text{tr} \left((l_t^{(2)} + l_t^{(1)} l_t^{(1)'}) \mathbb{E}(\eta_t \eta_s') \right) \right) \end{aligned} \quad (39)$$

- $\hat{e}(\gamma)$ sont les résidus de la régression OLS de $\lambda_t(\gamma, \hat{\theta})$ sur $l_t^{(1)}(\hat{\theta})$
- $\hat{\theta}$ sont les paramètres estimés du modèle contraint (sous H_0)

Les tests optimaux : Carrasco et al. (2014)

- Carrasco et al. (2014) analysent 2 statistiques de tests

1 Test de type sup

$$\sup TS = \sup_{\gamma \in \Gamma} TS_n(\gamma) \xrightarrow{d} \sup_{\gamma \in \Gamma} \left(\mathcal{N}(0, \Xi_{\gamma, \gamma}) - \frac{1}{2} \Xi_{\gamma, \gamma} \right) \quad (40)$$

2 Test de type exponentiel

$$\exp TS = \int_{\Gamma} \exp(TS_n(\gamma)) dJ_{\gamma} \xrightarrow{d} \int_{\Gamma} \exp \left(\mathcal{N}(0, \Xi_{\gamma, \gamma}) - \frac{1}{2} \Xi_{\gamma, \gamma} \right) dJ_{\gamma}$$

avec J une distribution à priori pour γ et à valeur sur Γ et

$$\Xi_{a,b} = \mathbb{E}_{\theta_0}((\lambda_t(a, \hat{\theta}) - d(a)'l_t^{(1)}(\theta_0))((\lambda_t(b, \hat{\theta}) - d(b)'l_t^{(1)}(\theta_0)))$$

avec $d(a) = d(a, \theta_0) = I(\theta_0)^{-1} \text{cov}(\lambda_t(a, \hat{\theta}), l_t^{(1)}(\theta_0))$ et $I(\theta_0)$ la matrice d'information de Fisher

- Carrasco et al. (2014) : tests asymptotiquement optimaux

- Ces tests sont optimaux au sens de Neyman-Pearson dans la mesure où ils maximisent un critère d'optimalité. Pour faire simple, pour un seuil de rejet donné, la puissance du test est maximale en comparaison d'un autre test ayant un niveau d'optimalité moindre
- Ces tests ne dépendent que du score et des dérivées du score sous H_0 et ne nécessitent pas de spécifier et estimer le modèle sous H_1

Les tests optimaux : Carrasco et al. (2014)

- Implémentation du test $\sup TS$ avec le modèle MS-moyenne (8)

$$Y_t = \mu_0 + \mu_1 S_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad S_t \in \mathbb{S} = \{0, 1\}$$

- On cherche à tester l'hypothèse linéaire contre l'alternative MS

$$H_0 : \mu_1 = 0$$

- La log-vraisemblance sous H_0 est donnée par

$$l_t = -\log \sqrt{2\pi} - \frac{1}{2} \log \sigma^2 - \frac{(y_t - \mu_0)^2}{2\sigma^2}$$

car sous H_0 , $y_t = \mu_0 + \varepsilon_t$

Les tests optimaux : Carrasco et al. (2014)

- Afin d'appliquer le test, il nous faut calculer les dérivées première et seconde de l_t

$$l_{1,t}^{(1)} = \frac{\partial l_t}{\partial \mu_0} = \frac{\varepsilon_t}{\sigma^2}$$

$$l_{2,t}^{(1)} = \frac{\partial l_t}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{\varepsilon_t^2}{2\sigma^4}$$

$$l_{11,t}^{(2)} = \frac{\partial^2 l_t}{\partial \mu_0^2} = -\frac{1}{\sigma^2}$$

$$l_{22,t}^{(2)} = \frac{\partial^2 l_t}{\partial \sigma^4} = \frac{1}{2\sigma^4} - \frac{\varepsilon_t^2}{\sigma^6}$$

$$l_{12,t}^{(2)} = \frac{\partial^2 l_t}{\partial \mu_0 \partial \sigma^2} = -\frac{\varepsilon_t}{\sigma^4}$$

$$l_t^{(1)} = (l_{1,t}^{(1)}, l_{2,t}^{(1)})'$$

$$l_t^{(2)} = \begin{pmatrix} l_{11,t}^{(2)} & l_{12,t}^{(2)} \\ l_{12,t}^{(2)} & l_{22,t}^{(2)} \end{pmatrix}$$

Les tests optimaux : Carrasco et al. (2014)

- Calculons $\lambda_t(\gamma, \hat{\theta})$ pour $\eta_t = (\mu_1 S_t, 0, 0)'$ en deux parties :

$$\text{tr} \left((l_t^{(2)} + l_t^{(1)} l_t^{(1)'}) \mathbb{E}(\eta_t \eta_t') \right) = \frac{\varepsilon_t^2 - \sigma^2}{\sigma^4}$$

et

$$\sum_{s < t} \text{tr} \left(l_t^{(1)} l_t^{(1)' } \mathbb{E}(\eta_t \eta_s') \right) = \frac{1}{\sigma^4} \sum_{s < t} \rho_{S_t}^{t-s} \varepsilon_t \varepsilon_s$$

avec $\rho_{S_t} = p_{11} + p_{22} - 1$, ce qui d'après (39) donne

$$\lambda_t(\gamma, \hat{\theta}) = \frac{1}{2\sigma^4} \left((\hat{\varepsilon}_t^2 - \hat{\sigma}^2) + 2 \sum_{s < t} \rho_{S_t}^{t-s} \hat{\varepsilon}_t \hat{\varepsilon}_s \right)$$

- Passons au calcul de (38) qui donne

$$\Lambda_n = n^{-1/2} \sum_t \lambda_t(\gamma, \hat{\theta}) = n^{-1/2} \frac{1}{2\sigma^4} \sum_{s < t} \rho_{S_t}^{t-s} \hat{\varepsilon}_t \hat{\varepsilon}_s$$

car $\hat{\varepsilon}_t$ est supposé de moyenne nulle et donc $\sum_t (\hat{\varepsilon}_t^2 - \hat{\sigma}^2) = 0$

- $\eta_t = (\mu_1 S_t, 0, 0)'$ avec les deux 0 qui tiennent compte de μ_0 et σ afin que η_t et θ_0 aient la bonne dimension
- $\rho_{S_t} = p_{11} + p_{22} - 1$ correspond à λ_2 (seconde valeur propre) qui intervient dans la fonction d'autocovariance de la chaîne (voir S32)

Les tests optimaux : Carrasco et al. (2014)

■ L'implémentation du test se déroule ainsi

- 1 Estimer le modèle sous H_0 par MLE
 - 2 A l'aide des OLS, régresser $\lambda_t(\gamma, \hat{\theta})$ sur $l_t^{(1)}$ et collecter \hat{e}_t
 - 3 Calculer la statistique de test via (37) puis (40)
 - 4 Calculer la distribution asymptotique du test pour obtenir les valeurs critiques limites
- Ce dernier point est complexe et s'obtient en suivant Carrasco et al. (2014), ce qui dans le cas présent donne

$$\sup TS \xrightarrow{d} \sup_{\rho_{S_t} : \rho_l < \rho_{S_t} < \rho_u} \frac{1}{2} (\max(0, K))^2$$

avec ρ_l et ρ_u les bornes sup et inf pour ρ_{S_t} et

$$K = \text{sign}(\rho_{S_t}) \sqrt{1 - \rho_{S_t}^2} \sum_{i=0}^{\infty} \rho_{S_t}^i Z_i, \quad Z_i \sim \mathcal{N}(0, 1)$$

avec $\text{sign}(\rho_{S_t}) = 1$ si $\rho_{S_t} > 0$, 0 si $\rho_{S_t} = 0$ et -1 si $\rho_{S_t} < 0$.

Les tests optimaux : Carrasco et al. (2014)

EMPIRICAL AND ASYMPTOTIC CRITICAL VALUES OF SUPTS AND EXPST

Percentile/c.v.	supTS		expTS	
	Asymptotic	Empirical	Asymptotic	Empirical
$\rho \in [-0.7, 0.7]$				
99%	3.96	3.99	4.17	4.24
95%	2.45	2.51	1.82	1.87
90%	1.82	1.81	1.35	1.33
80%	1.21	1.19	1.04	1.02
70%	0.86	0.85	0.90	0.89
50%	0.45	0.43	0.76	0.75
10%	0.04	0.04	0.59	0.58
5%	0.01	0.01	0.55	0.55
1%	0.00	0.00	0.50	0.50
$\rho \in [-0.98, 0.98]$				
99%	4.52	4.19	3.83	3.81
95%	2.99	2.78	1.82	1.76
90%	2.32	2.09	1.38	1.31
80%	1.65	1.44	1.07	1.02
70%	1.25	1.07	0.93	0.89
50%	0.74	0.60	0.78	0.74
10%	0.11	0.07	0.57	0.56
5%	0.05	0.03	0.54	0.52
1%	0.00	0.00	0.48	0.47

Il s'agit de simulation pour le modèle de simple que l'on vient de détailler :

$$Y_t = \mu_0 + \mu_1 S_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

Les tests de type Lagrange Multiplier : Hamilton (1996)

- Les tests LM de Hamilton (1996) sont utiles pour tester :

- ⇒ de l'autocorrélation omise
- ⇒ une variable explicative omise
- ⇒ de l'hétéroscédasticité omise
- ⇒ ...

- Les tests LM sont intéressants car spécifiés uniquement sous H_0

- ⇒ repose sur le score $S_n(\theta; Y_t)$ du modèle estimé sous la nulle

$$S_n(\theta; Y_t) = \frac{\partial \log p(y_t | \mathbf{y}^{t-1}, \theta)}{\partial \theta}$$

- ⇒ pour chaque modèle il faut donc dériver le score

- Analysons le cas d'un modèle avec variables possiblement exogènes

$$Y_t = \mathbf{x}'_t \beta_{S_t} + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \quad S_t = \{1, 2\} \quad (41)$$

avec $\mathbf{x}' = (x_{1t}, \dots, x_{mt}, y_{t-1}, \dots, y_{t-p})$ et $\beta_i = (\beta_{1,i}, \dots, \beta_{m+p,i})$

Test LM et autocorrélation

- Exemple avec de l'autocorrélation à l'ordre 1 :

- Sous H_1 on a $Y_t = \mathbf{x}'_t \beta_{S_t} + \varphi(Y_{t-1} - \mathbf{x}'_{t-1} \beta_{S_{t-1}}) + \sigma_{S_t} \varepsilon_t$

- Sous H_0 on a $\varphi = 0$ et le score “empirique” est donné par

$$s_t(\hat{\theta}) = \frac{\partial \log p(y_t | \mathbf{y}^{t-1}, \hat{\theta})}{\partial \hat{\theta}} \bigg|_{\varphi=0} = \sum_{i=1}^2 \sum_{j=1}^2 \psi_{t,j,i} \mathbb{P}(S_t = j, S_{t-1} = i | \mathbf{y}^t, \hat{\theta}) \\ + \sum_{\tau=1}^{t-1} \sum_{i=1}^2 \sum_{j=1}^2 \psi_{\tau,j,i} \left(\mathbb{P}(S_\tau = j, S_{\tau-1} = i | \mathbf{y}^\tau, \hat{\theta}) \right. \\ \left. - \mathbb{P}(S_\tau = j, S_{\tau-1} = i | \mathbf{y}^{t-1}, \hat{\theta}) \right)$$

$$\text{avec } \psi_{t,j,i} = (y_t - \mathbf{x}'_t \beta_j)(y_{t-1} - \mathbf{x}'_{t-1} \beta_i) / \sigma_j^2 = \frac{\partial \log p(y_t | \mathbf{y}^{t-1}, \mathbf{x}^{t-1}, \mathbf{S}^{t-1}, \theta, \varphi)}{\partial \varphi}$$

- La statistique de test est également donnée par

$$LM_n = \left(n^{-1/2} \sum_{t=1}^n s_t(\hat{\theta}) \right)' \left(\frac{1}{n} \sum_{t=1}^n s_t(\hat{\theta}) s_t(\hat{\theta})' \right)^{-1} \left(n^{-1/2} \sum_{t=1}^n s_t(\hat{\theta}) \right) \xrightarrow{d} \chi_1^2$$

Ici, l'autocorrélation des résidus à l'ordre 1 est captée par $\varphi(Y_{t-1} - \mathbf{x}'_{t-1}\beta_{S_{t-1}})$

Test LM et variables omises

- Sous H_1 on a $Y_t = \mathbf{x}'_t \beta_{S_t} + \mathbf{z}'_t \delta + \sigma_{S_t} \varepsilon_t$
- Sous H_0 on a $\delta = 0$ et le score “empirique” est donné par

$$s_t(\hat{\theta}) = \frac{\partial \log p(y_t | \mathbf{y}^{t-1}, \hat{\theta})}{\partial \hat{\theta}} \bigg|_{\delta=0} = \sum_{i=1}^2 \sum_{j=1}^2 \psi_{t,j,i} \mathbb{P}(S_t = j, S_{t-1} = i | \mathbf{y}^t, \hat{\theta})$$

$$\text{avec } \psi_{t,j,i} = (y_t - \mathbf{x}'_t \beta_j) \mathbf{z}'_t / \sigma_j^2 = \frac{\partial \log p(y_t | \mathbf{y}^{t-1}, \mathbf{x}^{t-1}, \mathbf{S}^{t-1}, \theta, \delta)}{\partial \delta}$$

- La statistique de test est donnée par

$$LM_n = \left(n^{-1/2} \sum_{t=1}^n s_t(\hat{\theta}) \right)' \left(\frac{1}{n} \sum_{t=1}^n s_t(\hat{\theta}) s_t(\hat{\theta})' \right)^{-1} \left(n^{-1/2} \sum_{t=1}^n s_t(\hat{\theta}) \right) \xrightarrow{d} \chi_l^2$$

avec l la taille du vecteur $\mathbf{z}'_t = (z_{1t}, \dots, z_{lt})$

Avec un raisonnement similaire on pourrait tester l'a présence d'un effet ARCH (test sur la nullité du paramètre devant ε^2)

Plan

- 1 Quelques faits stylisés
- 2 Chaîne de Markov
- 3 Modèle MS simple
- 4 Modèle MS-AR

- 5 Extensions
- 6 Estimation
- 7 Tests
- 8 Applications

Hamilton (1989)

Who Hamilton (1989)

Model modèle MS(2)-AR(4) décrit en (18)

Data le PNB réel américain trimestriel entre 1951T2 et 1984T4

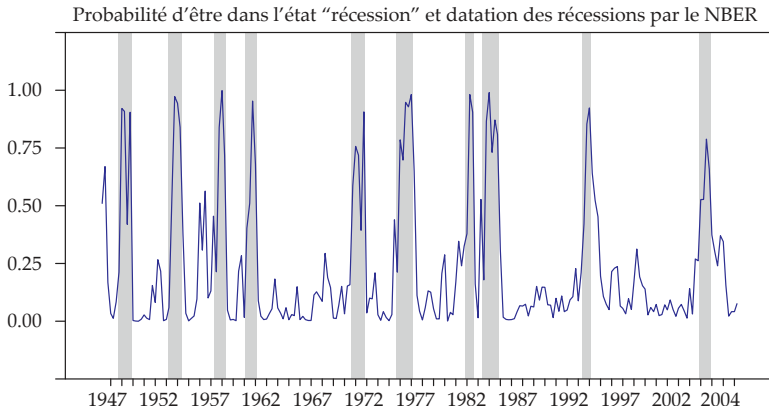
Results

MAXIMUM LIKELIHOOD ESTIMATES OF PARAMETERS AND ASYMPTOTIC STANDARD ERRORS
BASED ON DATA FOR U.S. REAL GNP, $t = 1952 : \text{II}$ TO $1984 : \text{IV}$

Parameter	Estimate	Standard error
α_1	1.522	0.2636
α_0	-0.3577	0.2651
p	0.9049	0.03740
q	0.7550	0.09656
σ	0.7690	0.06676
ϕ_1	0.014	0.120
σ_2	-0.058	0.137
ϕ_3	-0.247	0.107
ϕ_4	-0.213	0.110

Hamilton (1989)

- Chauvet et Hamilton (2005) étendent l'analyse de Hamilton (1989) pour dater les points de retournement de la croissance réelle américaine



Filardo (1994)

Who Filardo (1994)

Model modèle MS(2)-AR(0)-TVTP(k) décrit en (20)

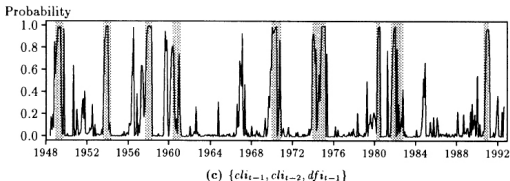
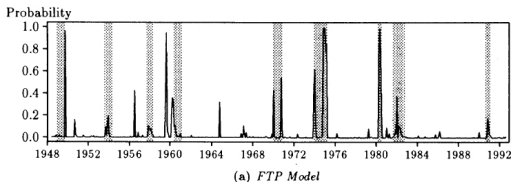
Data l'indice de prod industriel américain mensuel entre 1948 et 1992 ;
pour z_t , différent indicateurs du business cycle sont considérés

Results

Parameter	TVTP models using the composite leading index				
	FTP	(2)	(3)	(4)	(5)
	(1)	$\{cli_{t-1}\}$	$\{cli_{t-1}, cli_{t-2}\}$	$\{cli_{t-1}, dff_{t-1}\}$	$\{cli_{t-1}, cli_{t-2}, dff_{t-1}\}$
μ_0	-1.712 (.289)	-.896 (.172)	-1.288 (.228)	-.582 (.203)	-.778 (.164)
$\mu_0 + \mu_1$.330 (.074)	.461 (.079)	.371 (.063)	.494 (.068)	.455 (.056)
θ_{q0}	.458 (.329)	1.293 (.430)	-3.577 (.672)	-.493 (.596)	-2.524 (.540)
θ_{q1}		-1.033 (.583)	-2.493 (.848)	-1.208 (.628)	-2.133 (.547)
θ_{q2}			-6.699 (1.012)		-5.968 (.782)
θ_{q3}				-1.784 (.638)	
θ_{p0}	3.906 (.371)	4.205 (.507)	3.826 (.550)	3.371 (.770)	2.812 (.455)
θ_{p1}		1.654 (.359)	.726 (.559)	-1.803 (.811)	-.870 (.478)
θ_{p2}			.167 (.502)		
θ_{p3}				3.014 (1.323)	1.631 (.527)
p value*		.002	.000	.001	.000
AIC/SC	129.3/167.7	127.2/174.2	126.7/182.2	127.9/183.5	125.2/180.7

Filardo (1994)

- Les résultats concluent en faveur du modèle (5) et donc de la spécification TVTP plutôt que FTP



Maheu et al. (2012)

Who Maheu et al. (2012)

Data rendement du S&P500 hebdomadaire entre 1885 et 2010 (noté r_t)

Model modèle MS(4)-moyenne-variance décrit en (9) avec restrictions

$$P = \begin{pmatrix} p_{11} & p_{12} & 0 & p_{14} \\ p_{21} & p_{22} & 0 & p_{24} \\ p_{31} & 0 & p_{33} & p_{34} \\ p_{41} & 0 & p_{43} & p_{44} \end{pmatrix}$$

et les anticipations de long termes sur les rendements

$$\mathbb{E}(r_t | S_t = 1, 2) = \frac{\pi_1}{\pi_1 + \pi_2} \mu_1 + \frac{\pi_2}{\pi_1 + \pi_2} \mu_2 < 0$$

$$\mathbb{E}(r_t | S_t = 3, 4) = \frac{\pi_3}{\pi_3 + \pi_4} \mu_3 + \frac{\pi_4}{\pi_3 + \pi_4} \mu_4 > 0$$

■ Identification des régimes

- $\mu_1 < 0$: bear state
- $\mu_2 > 0$: bear rally state
- $\mu_3 < 0$: bull correction state
- $\mu_4 > 0$: bull state
- $\sigma_{S_t}^2$: pas de restriction

Maheu et al. (2012)

	mean	95% DI		mean	95% DI
π_1	0.070	(0.035, 0.117)	μ_1	-0.94	(-1.50, -0.45)
π_2	0.157	(0.073, 0.270)	μ_2	0.23	(0.04, 0.43)
π_3	0.304	(0.216, 0.397)	μ_3	-0.13	(-0.31, -0.01)
π_4	0.469	(0.346, 0.579)	μ_4	0.30	(0.22, 0.38)
			σ_1	6.01	(5.41, 6.77)
			σ_2	2.63	(2.36, 3.08)
			σ_3	2.18	(1.94, 2.39)
			σ_4	1.30	(1.20, 1.37)
$P = \begin{pmatrix} 0.921 & 0.076 & 0 & 0.003 \\ 0.015 & 0.966 & 0 & 0.019 \\ 0.010 & 0 & 0.939 & 0.051 \\ 0.001 & 0 & 0.039 & 0.960 \end{pmatrix}$			μ_1/σ_1	-0.16	(-0.25, -0.07)
			μ_2/σ_2	0.09	(0.02, 0.17)
			μ_3/σ_3	-0.06	(-0.14, -0.01)
			μ_4/σ_4	0.23	(0.17, 0.31)

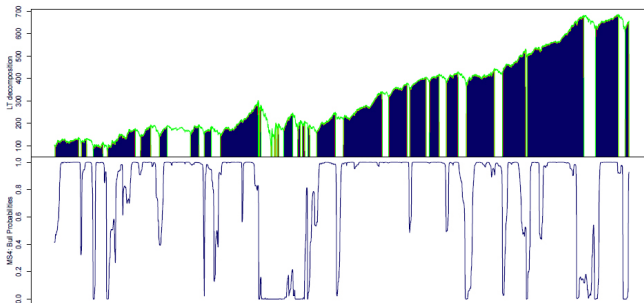
\Rightarrow Proba non-conditionnelle bearish market : $\pi_1 + \pi_2 = 0.227$

\Rightarrow Proba non-conditionnelle bullish market : $\pi_3 + \pi_4 = 0.773$

- Les signes des coefficients permettent d'identifier les états
- Donc la probabilité d'observer un marché haussier sur la période est de 77.3% quand la probabilité d'observer un marché baissier est de 22.7%

Maheu et al. (2012)

■ Graphique du S&P500 entre 1885 et 2010

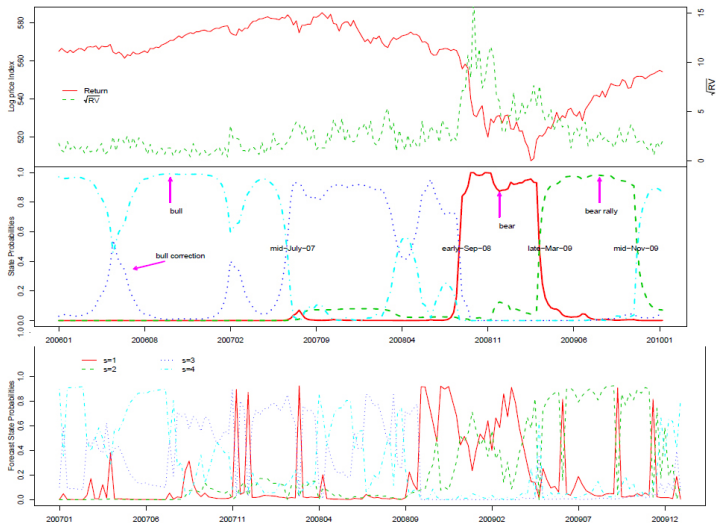


■ Le graphique des probabilités lissées pour les états haussiers

$$\mathbb{P}(S_t = 4|\mathbf{y}, \theta) + \mathbb{P}(S_t = 3|\mathbf{y}, \theta)$$

Maheu et al. (2012)

■ Estimation et prédiction :



- Ailliot, P., Pene, F., 2015. Consistency of the maximum likelihood estimate for non-homogeneous Markov-switching models. *ESAIM : Probability and Statistics*, 19, 268-292.
- Baum, L. E., T. Petrie, G. Soules, et 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.
- Bickel, P. J., Y. Ritov, et T. Rydén, 1998. Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *The Annals of Statistics* 26, 1614-1635.
- Carrasco, M., Hu, L., Ploberger, W., 2014. Optimal Test for Markov Switching Parameters. *Econometrica* 82, 765-784.
- Chauvet, M., Hamilton, J. D., 2006. Dating business cycle turning points. *Contributions to Economic Analysis*, 276, 1-54.
- Constant, K., Davin, M., de Truchis, G. and Keddad, B., 2015. How the European debt crisis has impacted the renewable energy sector? Work in progress.
- Douc, R., Moulines, E., Rydén, T., 2004. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *The Annals of statistics* 32, 2254-2304.
- Filardo, A., 1994. Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics* 12, 299-308.
- Garcia, R., 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* 39, 763-788.
- Goldfeld, S.M., Quandt, R.E., 1973. A Markov model for switching regressions. *Journal of Econometrics* 1, 3-16.

- Hamilton, J. D., 1988. Rational expectations econometric analysis of changes in regime : An investigation on the term structure of interest rates. *Journal of Economic Dynamics and Control* 12, 385-423.
- Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 357-384.
- Hamilton, J. D., 1994. *Time Series Analysis*. Princeton, NJ : Princeton University Press.
- Hamilton, J.D., 1996. Specification testing in Markov-switching time-series models. *Journal of Econometrics*.
- Hansen, B.E., 1992. The Likelihood Ratio Test under Non-standard Conditions : Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics* 7, 61-82.
- Kim, C.-J., Piger, J., Startz, R., 2008. Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics* 143, 263-273.
- Krolzig HM. 1997. *Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*. Springer : Berlin.
- Leroux, B. G., 1992. Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and Their Applications* 40, 127-143.
- Lindgren, G., 1978. Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics* 5, 81-91.
- Quandt, R.E., 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* 53, 873-880.

- Quandt, R. E., 1972. A new approach to estimating switching regressions. *Journal of the American Statistical Association* 67, 306-310.
- Maheu, J.M., McCurdy, T.H., Song, Y., 2012. Components of bull and bear markets : bull corrections and bear rallies. *Journal of Business & Economic Statistics* 37-41.
- McCulloch, R. E. and R. S. Tsay 1994. Statistical analysis of economic time series via Markov switching models. *Journal of Time Series Analysis* 15, 523-539.
- Otranto, E., 2005. The multi-chain Markov switching model. *Journal of Forecasting* 24, 523-537.
- Phillips, K. L., 1991. A two-country model of stochastic output with changes in regime. *Journal of International Economics* 31, 121-142.
- Poskitt, D.S., Chung, S.H., 1996. Markov chain models, time series analysis and extreme value theory. *Advances in Applied Probability* 28, 405-425.
- Timmermann, A. (2000). Moments of Markov switching models. *Journal of Econometrics*, 96(1), 75-111.