

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE.
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

UNIVERSITE MOULOU D MAMMERI, TIZI-OUZOU
Faculté des Sciences
Département de Mathématiques

MEMOIRE DE MASTER

en

Mathématiques appliquées
Option Processus aléatoires et statistique de la décision

THEME

Cointégration Dans Les Séries Chronologiques

Présenté par

Refsi Malika et Touati Sonia

Sous la direction de

M^{elle} **Atil Lynda**

Devant le jury d'examen composé de :

Fellag Hocine	Professeur	UMMTO	Président
Belkacem Cherifa	Maître de conférences B	UMMTO	Examinatrice

Soutenu le : 29/10/2014

Remerciements

Nous remercions "Dieu" le tout puissant d'avoir guidé nos pas vers les portes du savoir, et nous avoir donné suffisamment de santé et de courage pour mener nos travail à terme.

Nous tenons à remercier vivement notre promotrice Melle Atil Lynda qui a accepté de dériquer ce travail, et qui nous a apporté ses précieux conseils, nous le remercions aussi de nous avoir aidé à affronter toutes les difficultés rencontrées dans ce projet.

Nous voudrons également remercier du fond du coeur M.Fellag Hocine, pour son soutient moral.

Nous exprimons nos remerciements à tous ce qui ont contribué de près ou de loin à la réalisation de ce modeste travail.

Enfin, nous exprimons nos meilleurs gratitudes à tous mes enseignants qui nous ont formé, du primaire jusqu'à nos jours.

Dédicaces

C'est avec une pensée pleine de reconnaissance inspirée par la générosité et la gentillesse que je dédie ce modeste travail :

A ma chère maman pour son grand et généreux amour, son sacrifice, sa compréhension et son soutien.

A ma grande sœur Yamina et son époux Djamal sans oublier leurs enfants Syphax et Daréce qui j'adore beaucoup.

A mes deux frères Meziene et Abdelkader et mes deux petites sœurs Fatima et Taous les chattes ! .

A mon oncle Arezki et mes tantes sans oublier les cousins et les cousines.

A mon cher fiancé Hocine et toute ma belle-famille .

A mes meilleures copines qui m'ont soutenu moralement :la jeune maman Hammama et sa fille Chahinaz, Fadila, Nora ,Saliha ...

A mon binôme Sonia et toute sa famille .

A tous ceux qui m'ont aidé, conseillé, et à tous ceux que j'aime et que je porte dans mon cœur.

Malika

Table des matières

Introduction générale	2
1 Les séries chronologiques.	5
1.1 Introduction.	5
1.2 Définition.	5
1.3 Représentations graphiques	6
1.4 Les composantes d'une série chronologique	9
1.5 Les opérateurs de Box-Jenkins.	12
1.5.1 Les effets des opérateurs de Box-Jenkins .	13
1.6 Premiers indices descriptifs.	13
1.7 Inversion , factorisation des polynômes et la présence d'une racine unitaire .	14
1.7.1 Inversion des polynômes.	14
1.7.2 Factorisation des polynômes.	16
1.8 Modélisation <i>ARIMA</i> des séries chronologiques.	17
1.8.1 Les processus autorégressifs, <i>AR</i> (Autoregressive).	18
1.8.2 Les processus moyennes mobiles, <i>MA</i> (Moving Average).	18
1.8.3 Les processus <i>ARMA</i> (p, q) (Autorégressive Moving Average).	18
1.8.4 Les processus intégrés <i>I</i> (d).	19
1.8.5 Les processus <i>ARIMA</i> (p, d, q).	19
1.9 Stationnarité dans un cadre univarié.	19
1.9.1 Exemples des processus stationnaires.	20
1.9.2 Exemples des processus non-stationnaires.	21
1.10 Stationnarité dans un cadre multivarié.	24
1.10.1 Définition d'un processus multivarié.	24
1.10.2 Processus vectoriels stationnaires du second ordre.	24
1.10.3 Processus bruit blanc multivarié.	25
1.10.4 La Fonction d'autocovariance et d'autocorrélation sous l'hypothèse de stationnarité.	25
1.10.5 Classe des modèles VARMA - Notions de causalité et d'inversibilité .	26
1.10.6 Le modèle VAR sous l'hypothèse de stationnarité.	28
1.10.7 Représentation <i>VMA</i> du processus <i>VAR</i> .	29
1.11 Modélisation <i>VAR</i> en différences.	31

2	La Cointégration et ses concepts.	33
2.1	Introduction.	33
2.2	Cointégration dans un cadre bivarié.	34
2.3	Cointégration dans un cadre multivarié.	34
2.4	Généralisation au cas trend stationnaire.	35
2.5	Modèles à correction d'erreur (ECM).	36
2.5.1	Représentation.	36
2.6	Représentation possibles des séries cointégrées.	37
2.6.1	Représentation en moyenne mobile.	38
2.6.2	Représentation en tendances communes.	39
2.6.3	Représentation autorégressive.	39
2.7	Théorème de représentation de Granger.	41
3	Tests de cointégration.	43
3.1	Introduction.	43
3.2	Tests de Racine Unitaire.	44
3.2.1	Un modèle général pour les tests.	45
3.2.2	Tests de Dickey-Fuller <i>DF</i>	46
3.2.3	Test de Dickey-Fuller augmenté <i>ADF</i>	47
3.2.4	Test de Phillips – Perron <i>PP</i>	48
3.3	L'approche de Engle et Granger (1987).	50
3.3.1	Estimation des modèles à correction d'erreur :Méthode d'estimation en deux étapes.	50
3.3.2	Tests univariés de cointégration (de Engle et Granger).	51
3.4	Approche multivariée de la cointégration : l'analyse de Johansen.	52
3.4.1	Test du nombre de relations de cointégration.	54
3.5	Quelques commentaires à propos de la puissance des tests de cointégration .	55
3.6	Exemple d'application sous R	56
3.6.1	Méthode de Engle et Granger (1987) sous R	59
3.6.2	Méthode de Johansen (1988) sous R :	62
3.7	Exemple du marché du merlu en Bretagne	63
3.7.1	Le choix du marché	63
3.7.2	le choix de l'échantillon et des données	63
3.7.3	La méthode économétrique de cointégration	64
3.7.4	Application de la méthode	64

Introduction générale

Au cours de la dernière décennie, un nombre important de travaux, tant théorique qu'empiriques, ont été publiés sur les problèmes de racines unitaires dans la représentation des séries chronologiques macroéconomiques. L'élément catalyseur de cette série de travaux de recherche a certes été l'étude de **Nelson et Plosser [1982]**, qui a démontré en utilisant les tests de **Dickey et Fuller [1979]**, que la plupart des séries macroéconomiques étaient probablement caractérisées par une telle racine unitaire dans leur représentation autorégressive. L'importance du problème vient du fait que la présence d'une racine unitaire autorégressive influe considérablement sur l'interprétation de certaines théories, et plus encore sur les méthodes statistiques et les distributions asymptotiques à utiliser. Citons comme exemple la théorie de la consommation et du revenu permanent avancée par **Hall [1978]** et certains problèmes causés par la présence d'une racine unitaire qui sont analysés par **Hankiw et Shapiro [1985]**.

Les études qui ont suivi les travaux de **Nelson et Plosser** ont analysé ces problèmes dans les modèles à une ou plusieurs variables. Une partie de cette littérature, se concentrant sur les propriétés des séries à une variable, a analysé la question de persistance des fluctuations macroéconomiques. Pour les modèles à plusieurs variables, les développements théoriques se sont axés sur la question de cointégration qui permet à certaines combinaisons linéaires d'un ensemble de séries, possédant chacune une racine unitaire, d'être à leur tour stationnaires.

L'objectif de ce travail est d'aborder le phénomène de cointégration qui ne peut pas être indépendant ni détacher de la notion des racines unitaires qui jouent un rôle très important au niveau de la modélisation, estimation et prévision dans les séries chronologiques, et dans ce but, le travail est réparti sur trois chapitres, organisés comme suit :

Le premier chapitre est un rappel général sur les séries chronologiques et ses concepts fondamentaux dans le cas univarié et multivarié notamment la notion de stationnarité qui est aussi inévitable dans notre travail, nous finirons par la modélisation *VAR* en différences qui nous amène et qui nous met face la notion de cointégration .

Par contre le deuxième chapitre est consacré à la notion de cointégration dans le cas bivarié, multivarié et au cas trend stationnaire, pour passer en suite au modèle à correction d'erreur **ECM**, et aux quelques représentations et caractérisations possibles des séries cointégrées pour leur intérêt particulier porté sur les problèmes de cointégration de leurs éléments .

Quant au troisième chapitre nous avons essayé de faire un large panorama des tests de racine unitaire et de cointégration les plus populaires qui sont souvent utilisés sous *R*. Nous terminons ce travail en citant deux exemples d'applications .

Chapitre 1

Les séries chronologiques.

1.1 Introduction.

Dans ce chapitre nous allons faire un rappel des séries chronologiques et ses concepts fondamentaux, nous allons commencer par une définition générale de ces séries chronologiques suivit de ses diverses représentations graphiques et de ses composantes que nous allons accompagner de simples exemples, comme nous allons rappeler de ses opérateurs, les estimateurs empiriques et le modèle *ARIMA* .Pour pouvoir en seconde étape de ce chapitre d'aborder la notion de stationnarité dans le cadre univarié et multivarié finissons par la modélisation *VAR* en différences .

1.2 Définition.

Définition 1.2.1. De manière générale, les séries chronologiques modélise un phénomène qui dépend du temps on les appelle aussi : Séries Temporelles, Chroniques ,(Time Series en anglais).

En clair, une série chronologique est une suite finie (x_1, \dots, x_n) de données indexées par le temps. L'indice temps peut être selon les cas la minute, l'heure, le jour, l'année etc.... Le nombre n est appelé la longueur de la série .Et lorsqu'on dispose de plusieurs séries (par exemple la série des consommations mensuelles et celle des prix mensuels) on parle de série chronologique multidimensionnelle .L'étude des séries temporelles à pour but de :

- Comprendre le passé : expliquer les variations observées
- Prédire les valeurs futures (proches)
- Etudier le lien avec d'autres séries.

Et les domaines d'applications des séries chronologiques sont très variés citons par exemple :

- Finance et économétrie : évolution des indices boursiers, des prix, des données économiques des entreprises, des ventes et achats de biens, des productions agricoles ou industrielles .

- Assurance : analyse des sinistres .
- médecine / biologie : suivi des évolutions des pathologies, analyse d'électroencéphalogrammes et d'électrocardiogrammes.
- Sciences de la terre et de l'espace : indices des marées, variations des phénomènes physiques (météorologie), évolution des taches solaires, phénomènes d'avalanches.
- Traitement du signal : signaux de communications, de radars, de sonars, analyse de la parole
- Traitement des données : mesures successives de position ou de direction d'un objet mobile (trajectographie) ... *etc*

Remarque 1.2.1. Les valeurs mesurées des série chronologiques peuvent être de deux types :

- Niveau : On mesure une grandeur à un instant donné (exemple : nombre mensuel de ventes d'un produit) ;
- Flux : On mesure la variation d'une grandeur durant une période (par exemple, variations mensuelles du nombre de demandeurs d'emploi).

1.3 Représentations graphiques

Plusieurs types de graphes peuvent être envisagés :

- a . Chronogramme :** L'étude d'une série temporelle commence par l'examen de son chronogramme. Il en donne une vue d'ensemble, montre certains aspects, comme des valeurs atypiques, d'éventuelles ruptures, un changement dans la dynamique de la série .

Exemple 1.3.1. (Populations) Les populations de la France et des Etats-Unis, en millions d'habitants,(fig (1.1)) constituent des séries où le temps explique bien le niveau de la série.

Leur graphique est obtenu par :

```
>> require(caschrono)
>> require(fSeries)
>> data(popfr)
>> plot.ts(popfr,xlab='année',ylab='population')
>> plot.ts(uspop,xlab='année',ylab='population')
```

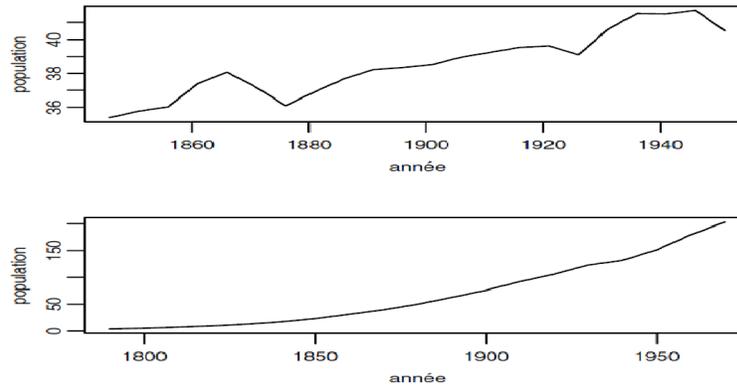


FIG. 1.1: Population française (haut), population des Etats-Unis (bas) en millions d'habitants.

Remarque 1.3.1. Les démographes peuvent être intéressés par la prévision de la population à 10 ans, à 20 ans. Remarquons que la population française varie de 35.4 à 41.7 sur 110 ans, tandis que celle des Etats-Unis varie de 3.93 à 203.00 sur 180 ans.

b . **Month plot :** Un month plot (qu'on pourrait traduire approximativement par « chronogramme par mois ») est une représentation simultanée des chronogrammes des séries associées à chaque saison, appelée conventionnellement mois. Il dessine un chronogramme de la sous-série correspondant à chaque niveau de la saison (mois de l'année, jour de la semaine ...).

Exemple 1.3.2. Les températures à Nottingham Castle durant 1920 à 1940

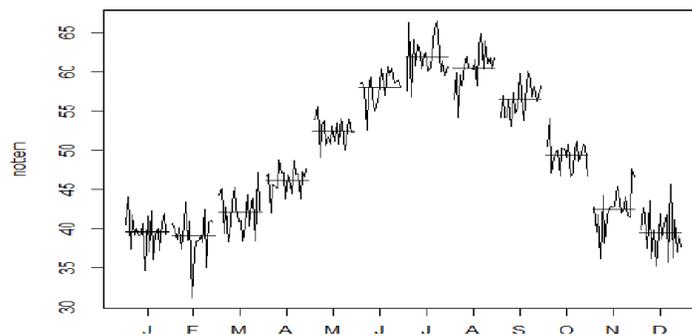


FIG. 1.2: Températures à Nottingham Castle.

Le graphique est obtenu par :

» data(nottem)

» monthplot(nottem)

c . Diagramme à échelle semi-logarithmique :

Le temps est porté sur l'axe des abscisses, gradué arithmétiquement et chaque valeur de la chronique est représentée par un point qui a pour ordonnée le logarithme décimal de la valeur observée. Les points successifs sont alors reliés par des segments de droite dans l'ordre croissant de leurs abscisses. On utilise un diagramme à échelle semi-logarithmique quand :

- les nombres à porter sur l'axe des ordonnées sont dans un intervalle trop grand pour utiliser une échelle arithmétique ;
- on veut mettre en évidence des propriétés du phénomène qui n'apparaîtraient pas aussi clairement en échelle arithmétique.

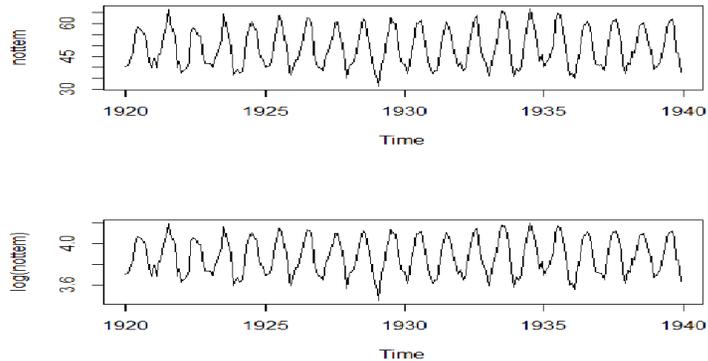


FIG. 1.3: Températures à Nottingham Castle (haut) , $\log(\text{Températures à Nottingham Castle})$ (bas).

Leur graphique est obtenu par :

```
>> data(nottem)
>> layout(1 :2)
>> plot(nottem)
>> plot(log(nottem))
```

1.4 Les composantes d'une série chronologique

A travers de nombreux exemples, les séries temporelles ont révélé un certain nombre de mouvements et de variations caractéristiques dont certains se manifestent à des degrés variés. L'analyse de ces mécanismes est importante, en particulier pour prévoir les variations futures. On peut généralement distinguer, dans l'évolution d'une chronique, cinq composantes fondamentales :

- La tendance à long terme ou trend :

On peut définir la tendance comme étant un comportement "moyen" de la série ou bien une représentation de mouvement profond de l'évolution à très long terme du phénomène. On note T_t sa valeur et sur un diagramme cartésien, on la représente par une courbe de tendance qui peut être :

- une tendance linéaire lorsque la série peut se décomposer en

$$x_t = at + b + \varepsilon_t \quad t = 1, 2, \dots \quad \text{où} \quad T_t = at + b \quad (1.1)$$

Plus généralement, on parle de tendance polynômiale lorsque la série peut se décomposer en

$$x_t = a_1 t^p + a_2 t^{p-1} + \dots + a_{p+1} + \varepsilon_t \quad t = 1, 2, \dots$$

$$\text{où} \quad T_t = a_1 t^p + a_2 t^{p-1} + \dots + a_{p+1} \quad (1.2)$$

expression dans laquelle ε_t est un résidu où ne figure plus la tendance.

- La tendance peut être multiplicative dans certaines séries :

$$x_t = T_t \varepsilon_t \quad t = 1, 2, \dots \quad (1.3)$$

où T_t prend l'une des formes (linéaire, polynômiale etc ...). C'est alors le logarithme des données (si elles sont positives!) qui présente une tendance additive.

Remarque 1.4.1. De même on pourra définir des tendances logarithmiques, exponentielles etc ...

- Les mouvements cycliques :

On identifie des mouvements cycliques qui s'équilibrent autour de la tendance à long terme et qui sont liés aux variations conjoncturelles (par exemple, à la succession des phases du cycle économique : prospérité, crise, dépression, reprise). On note C_t leur valeur. Sa périodicité étant assez grande, cette composante ne peut être mise en évidence que sur des chroniques assez longues, on l'ignore quand les données ne remontent pas suffisamment dans le temps.

Remarque 1.4.2. Les composantes tendances et cycle sont souvent regroupées et on parle, alors, de composante tendance – cycle, notée TC_t .

- Les variations saisonnières :

On appelle variations saisonnières ou saisonnalité, des fluctuations périodiques de même type que les mouvements cycliques mais de période plus courte. On note S_t leur valeur.

Les variations saisonnières peuvent avoir une période : journalière (trafic horaire du métro), hebdomadaire (nombre d'heures travaillées par jour), annuelle (indice mensuel de la production industrielle, chiffre d'affaires mensuel des grands magasins). Si p est la période, on a $S(t + p) = S(t)$.

Les variations saisonnières ont de multiples causes : cycle des saisons, dispositions réglementaires, dont les effets se produisent à date fixe. Voici quelques exemples :

- les congés : les congés annuels se traduisent chaque été par un ralentissement sensible de l'activité et une diminution des principales grandeurs économiques.
- l'inégalité des différents mois (nombres de jours, nombres de fêtes mobiles, nombre de "ponts").
- les facteurs climatiques qui influent sur les congés, sur l'activité de l'industrie du bâtiment, sur la consommation d'électricité.
- la périodicité de l'offre et de la demande de certains produits (rythme saisonnier de la production agricole, ventes de fin d'année, demandes d'automobile au printemps).

Remarque 1.4.3. 1. Il convient d'effectuer un test permettant de détecter l'existence d'une saisonnalité avant d'effectuer sur une série temporelle une correction de variations saisonnières (CVS). Le test le plus communément utilisé est celui de Fisher, le test se présente comme suit :

- Calcul de la somme U^* des carrés du modèle avec tendance simple
- Calcul de la somme U^{**} des carrés du modèle avec tendance et saisonnalité
- Calcul de la valeur du Fisher empirique

$$F_* = \frac{U^* - U^{**}}{K - P} / \frac{U^{**}}{T - K} \quad (1.4)$$

La valeur du F_* empirique est à comparer à la valeur du F théorique donné par la table de la loi de Fisher-Snedecor aux degrés de liberté $K - P$ et $T - K$, avec :

K , le nombre de paramètres indépendants à estimer dans le cadre du modèle avec tendance. et saisonnalité .

P , le nombre de paramètres à estimer dans le cadre du modèle à tendance seule , T est le nombre d'observations.

Hypothèse 1 ; $F_* > F$ alors la série est saisonnière.

Hypothèse 2 ; $F_* \leq F$ alors la série n'est pas saisonnière.

2. La dessaisonnalisation d'une série temporelle consiste à estimer sa composante saisonnière S_t et à l'extraire de la série brute (X_t) . On obtient ainsi une série corrigée des variations saisonnières (CVS), (X_t^{CVS}) .

3. La série (CVS) contient la composante résiduelle et ne doit pas être confondue avec la tendance.

- Les composantes irrégulières ou aléatoires :

Ces composantes prennent en compte les aspects aléatoires de la chronique ; c'est la partie imprévisible que l'on espère faible ; on considère qu'elle comprend, en plus de l'aléa, tout ce qui n'a pas été pris en compte par les autres composantes du modèle. On note ε_t leur valeur.

Elles font intervenir des composantes conjoncturelles ou accidentelles pour tenir compte des phénomènes particuliers, limités dans le temps (grèves, conditions météorologiques exceptionnelles, crash financier. Souvent ε_t est un bruit blanc.

Définition 1.4.1. Un bruit blanc est une suite de variables aléatoires non corrélées, d'espérance et de variance constante. Autrement dit, pour tout t

$$\mathbb{E}(X_t) = \mu, \quad \text{Var}(X_t) = \sigma^2 \quad \text{et} \quad \text{Cov}(X_t, X_{t+h}) = 0 \quad \text{pour} \quad h \neq 0$$

Si l'espérance μ est nulle, on dit que le bruit blanc est centré.

Le cas particulier le plus courant est la suite de variables aléatoires gaussiennes standard (espérance nulle et variance égale à 1) et indépendantes.

Exemple 1.4.1. La Moyenne mensuelle de nombre relative de tache solaire (**sunspot**) compté de 1749 à 1983. Rassemblé par l'observatoire fédéral Suisse, Zurich jusqu'à 1960, ensuite par l'observatoire astronomique Tokyo.

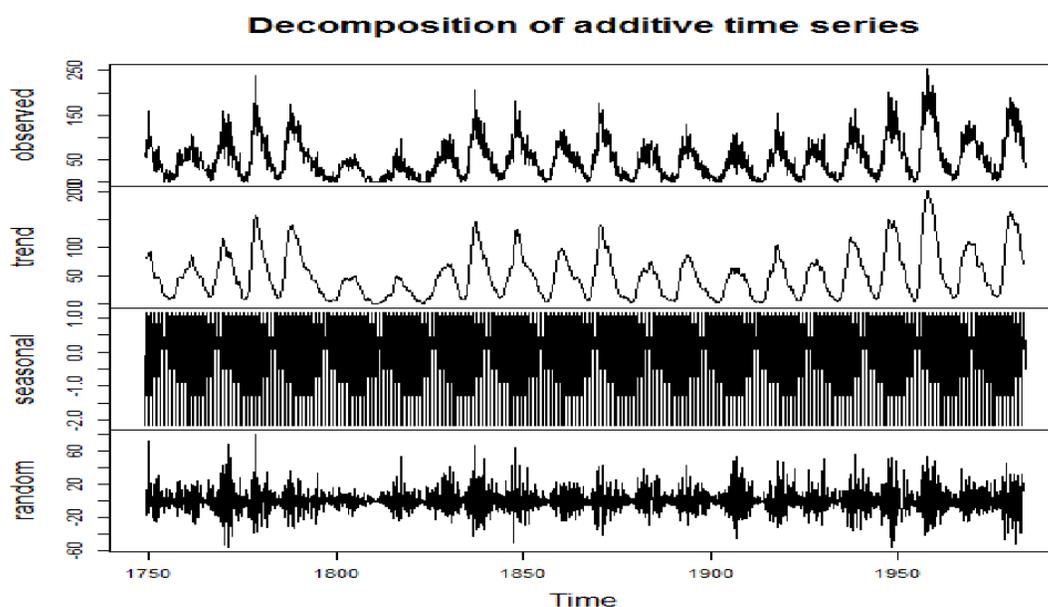


FIG. 1.4: La décomposition additive de la série sunspot.

Le programme relatif à ce graphe sous R est :

```

>> require(graphics)
>> plot(decompose(sunspots))

```

1.5 Les opérateurs de Box-Jenkins.

1- L'opérateur d'avance F (Farward).

L'opérateur est défini par :

$$FX_t = X_{t+1} \quad (1.5)$$

En d'autres termes :

$$F^p X_t = X_{t+p} \quad (1.6)$$

2- L'opérateur de recul B (Backward).

L'opérateur de recul B est défini comme agissant sur la série. A un instant t on fait correspondre la valeur de la série à l'instant $t - 1$. On définit ainsi une nouvelle série BX comme :

$$BX_t = X_{t-1} \quad (1.7)$$

On peut appliquer plusieurs fois cet opérateur, on définit ainsi des nouvelles séries :

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}$$

et

$$B^m X_t = X_{t-m} \quad (1.8)$$

3- L'opérateur de différenciation ∇ .

L'opérateur ∇ (prononcer "nabla") est défini par :

$$\nabla X_t = X_t - X_{t-1} \quad (1.9)$$

Nous entrons maintenant dans des considérations de notations. En effet, par écriture purement formelle on peut écrire :

$$\nabla X_t = X_t - BX_t = (1 - B)X_t \quad (1.10)$$

On peut écrire ∇ sous la forme d'un polynôme en B avec :

$$\nabla = (1 - B) \quad (1.11)$$

Ce mode d'écriture sous forme de polynôme en B est en fait très pratique mais totalement formel. Il ne faut pas oublier que, quand on écrit $(1 - B)X$, on définit à partir d'une série X , une nouvelle série qui à t , fait correspondre la différence entre la valeur de la série observée à l'instant t et celle observée à l'instant $t - 1$.

4- L'opérateur de désaisonnalisation ∇_s .

L'opérateur est défini par :

$$\nabla_s = X_t - X_{t-s} \quad (1.12)$$

En d'autres termes :

$$\nabla_s = (1 - B^s) \quad (1.13)$$

1.5.1 Les effets des opérateurs de Box-Jenkins .

1- L'opérateur ∇ .

1. permet d'éliminer la tendance de la série.
2. peut être répété plusieurs fois si la tendance n'est pas linéaire. Par exemple :

$$\nabla^2 X_t = (1 - B)^2 X_t = (1 - 2B + B^2) X_t$$

permet d'éliminer une tendance quadratique. Le nombre de fois où on applique ∇ est appelé ordre de différenciation.

2- L'opérateur ∇_s .

1. permet d'éliminer la saisonnalité de période s .
2. On peut également l'appliquer plusieurs fois.

$$\nabla_s^2 X_t = \nabla_s (X_t - X_{t-1})$$

Le nombre de fois où on applique ∇_s est appelé ordre de désaisonnalisation.

1.6 Premiers indices descriptifs.

Il est bien utile de disposer de quelques indices numériques qui "résumant" une série (x_1, \dots, x_n) .

1 Indices de tendance centrale.

Ce sont, classiquement, la moyenne et la médiane. la moyenne empirique définit par :

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j \quad (1.14)$$

2 Indices de dispersion.

Les plus courants sont la variance empirique (ou plus exactement l'écart type empirique qui en est la racine carrée) et l'étendue (différence entre la plus grande valeur et la plus petite). Elles indiquent la dispersion des observations autour de leur indice de tendance centrale ,définit par :

$$\hat{\sigma}_n(0) = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_n)^2 \quad (1.15)$$

3 Indices de dépendance.

3.1 Auto-covariances empiriques.

Ces indices $\hat{\sigma}_n(1), \hat{\sigma}_n(2), \dots$ donnent une idée de la dépendance entre les données. Plus exactement

$$\hat{\sigma}_n(1) = \frac{1}{n} \sum_{j=1}^{n-1} (x_j - \bar{x}_n)(x_{j+1} - \bar{x}_n) \quad (1.16)$$

indique la dépendance entre deux données successives,

$$\hat{\sigma}_n(2) = \frac{1}{n} \sum_{j=1}^{n-2} (x_j - \bar{x}_n)(x_{j+2} - \bar{x}_n) \quad (1.17)$$

indique la dépendance entre deux données écartées de deux unités de temps et ainsi de suite jusqu'à

$$\hat{\sigma}_{n-1}(2) = \frac{1}{n} (x_1 - \bar{x}_n)(x_n - \bar{x}_n) \quad (1.18)$$

Il est recommandé de ne prendre en considération que les auto-covariances empiriques $\hat{\sigma}_n(1), \dots, \hat{\sigma}_n(h)$ pour un h "pas trop grand" par rapport à la longueur de la série.

3.2 Auto-corrélations empiriques.

Ce sont les quotients des covariances sur la variance :

$$\hat{\rho}_n(j) = \frac{\hat{\sigma}_n(j)}{\hat{\sigma}_n(0)}, \quad j = 0, \dots, h. \quad (1.19)$$

Évidemment $\hat{\rho}_n(0) = 1$.

1.7 Inversion , factorisation des polynômes et la présence d'une racine unitaire .

L'intérêt de cette section est de montrer comment effectuer l'inversion et la factorisation d'un polynôme ,qui sont des outils essentiels pour aborder la notion de racines unitaires et le rôle que joue cette dernière sur les tests de stationnarité.

1.7.1 Inversion des polynômes.

Pour parler de l'inverse d'un polynôme de retard, il est commode dans un premier temps de considérer un polynôme particulier qui est le polynôme de retard de degré un défini par :

$$A(B) = 1 - \alpha B \quad (1.20)$$

Pour $|\alpha| < 1$ ce polynôme possède un inverse, c'est à dire que l'on peut définir :

$$A^{-1}(B) = \frac{1}{1 - \alpha B} = \sum_{i=0}^{\infty} \alpha^i B^i \quad (1.21)$$

en utilisant l'expression de la somme d'une progression géométrique. Considérons maintenant le polynôme $A(B)$ de degré p que l'on note :

$$A(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p \quad (1.22)$$

On définit l'équation caractéristique associée à ce polynôme comme l'expression en z

$$A(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p \quad (1.23)$$

On a le théorème suivant :

Théorème 1.7.1. *Le polynôme $A(B)$ est inversible si les p racines l_j de son équation caractéristique associée sont toutes extérieures au cercle unité. Son inverse est donné par :*

$$A^{-1}(B) = \prod_{j=1}^P \left[\sum_{k=0}^{\infty} \left(\frac{1}{l_j}\right)^k B^k \right] \quad (1.24)$$

La preuve de ce théorème utilise le résultat précédent en se ramenant à une série d'opérations élémentaires où l'on aurait à inverser que des polynômes de degré un.

Preuve 1.7.1. *Factoriser le polynôme $A(z)$ en utilisant les p racines de l'équation caractéristique :*

$$A(z) = \prod_{j=1}^P (z - l_j) \alpha_r \quad (1.25)$$

On peut remarquer que le produit des racines est égal à $1/\alpha_r$ car :

$$A(0) = 1 = \prod_{j=1}^P (-l_j) \alpha_r \quad (1.26)$$

D'autre part on a la factorisation :

$$(z - l_j) = -l_j \left(1 - \frac{z}{l_j}\right) \quad (1.27)$$

ce qui permet d'exprimer le polynôme en z sous la forme :

$$A(z) = \prod_{j=1}^P \left(1 - \frac{z}{l_j}\right) \quad (1.28)$$

On peut alors ramener le calcul de l'inverse de $A(z)$ au calcul simple suivant que l'on sait effectuer :

$$A^{-1}(z) = \prod_{j=1}^P \left(1 - \frac{z}{l_j}\right)^{-1} \quad (1.29)$$

car :

$$\left(1 - \frac{z}{l_j}\right)^{-1} = \sum_{k=0}^{\infty} \left(\frac{1}{l_j}\right)^k B^k \quad (1.30)$$

Cet inverse existe si les racines l_j de l'équation caractéristique sont toutes en dehors du cercle unité.

Définition 1.7.1. Alors si on veut définir les racines unitaires, elles représentent les l_j à l'intérieures du cercle unité.

i.e : $|l_j| < 1$

Remarque 1.7.1. Dans le cadre multivarié ,on ne parle pas de l'inverse d'un polynôme mais de l'inverse d'une matrice .

1.7.2 Factorisation des polynômes.

Une opération très pratique que l'on peut opérer sur les polynômes en général et les polynômes de retard en particulier est ce que l'on appelle la division des polynômes. Si l'on range un polynôme par puissances décroissantes, on peut effectuer une division par un autre polynôme aussi facilement (ou à peu près) que ce l'on effectue la division Euclidienne de deux nombres. Nous allons résoudre un exemple qui s'avérera très utile par la suite, car il permet de factoriser le polynôme $(1 - B)$:

$$\begin{array}{r} b_2 B^2 + b_1 B + b_0 \\ \hline -(b_2 B^2 + b_2 B) \\ \hline 0 + (b_1 + b_2)B + b_0 \\ \hline -((b_1 + b_2)B - (b_1 + b_2)) \\ \hline 0 + b_0 + b_1 + b_2 \end{array} \quad \begin{array}{l} | -B + 1 \\ - b_2 B - (b_1 + b_2) \end{array} \quad (1.31)$$

Cet exemple permet de montrer que :

$$\theta_2 B + \theta_1 B + \theta_0 = \left(-\theta_2 B - (\theta_1 + \theta_2) \right) (1 - B) + \Theta(1) \quad (1.32)$$

ou d'une manière générale :

$$\Theta(B) = \left(-\theta_0 - (\Theta(1) + \Theta^*(B)) \right) (1 - B) + \Theta(1) \quad (1.33)$$

Si q est le degré du polynôme $\Theta(B)$, le polynôme $\Theta^*(B)$ sera de degré $q - 1$ sans terme constant avec :

$$\Theta^*(B) = \theta_1^* B + \theta_2^* B^2 + \dots + \theta_{q-1}^* B^{q-1}$$

et ses coefficients sont définis par :

$$\theta_j^* = \sum_{i=j}^{q-1} \theta_{i+1} \quad j = 1, \dots, q-1 \quad (1.34)$$

Considérons maintenant le polynôme $A(B)$ de degré p :

$$A(B) = 1 - a_1B - a_2B^2 - \dots - a_pB^p \quad (1.35)$$

On peut montrer par une technique similaire de division de polynôme les relations suivantes :

$$A(B) = A(1) + (1 - B)[1 - A(1) - A^*(B)] \quad (1.36)$$

où les coefficients du polynôme $A^*(B)$ sont définis par :

$$a_j^* = -\left(\sum_{i=j}^{p-1} a_{i+1}\right) \quad j = 1, \dots, p-1 \quad (1.37)$$

Par de simples manipulations algébriques, on arrive à :

$$A(B) = A(1)B + (1 - B)[1 - A^*(B)] \quad (1.38)$$

et :

$$A(B) = (1 - [1 - A(1)]B) - (1 - B)A^*(B) \quad (1.39)$$

que l'on réécrira parfois comme :

$$A(B) = (1 - \rho B) - (1 - B)A^*(B) \quad (1.40)$$

Cette écriture sera utilisée dans les chapitres qui suivront notamment sur les tests de racine unitaire en posant $\rho = 1 - A(1)$. Cette dernière écriture permet de montrer que si un polynôme de retards comporte une racine unitaire $(1 - B)$, alors la somme de ses coefficients est égale à zéro. Il suffit pour cela d'y poser $B = 1$.

Remarque 1.7.2. On peut effectuer le même type de factorisation dans le cas des polynômes matriciels

1.8 Modélisation ARIMA des séries chronologiques.

Les modèles ARIMA permettent de combiner trois types de processus temporels : les processus autorégressifs (AR), les processus moyenne mobile (MA) et les processus intégrés (I). Dans le cas général, un modèle ARIMA (p, d, q) est une combinaison de ces trois types de processus, p, d et q désignant respectivement l'ordre du processus autorégressif, l'ordre d'intégration et l'ordre de la moyenne mobile. Il s'agit par la méthode de Box-Jenkins de construire un modèle restituant le mieux possible le comportement d'une série temporelle suivant trois étapes : identification, estimation et diagnostic.

1.8.1 Les processus autorégressifs, AR (Autoregressive).

Un processus (X_t) est dit autorégressif d'ordre p , $AR(p)$, si l'observation présente X_t est générée par une moyenne pondérée des observations passées jusqu'à la p -ième période sous la forme suivante :

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t \quad (1.41)$$

où α_i , $i = 1, \dots, p$ sont des paramètres positifs ou négatifs à estimer (ε_t) est un bruit blanc *i.e* les ε_t sont *i.i.d* suivant une loi $N(0, \sigma^2)$.

Les processus autorégressifs supposent donc que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents, plus un terme aléatoire d'erreur.

1.8.2 Les processus moyennes mobiles, MA (Moving Average).

Dans un processus X_t de moyenne mobile d'ordre q , chaque observation X_t est générée par une moyenne pondérée d'aléas jusqu'à la q -ième période dans le passé.

$$X_t = \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1.42)$$

Les moyennes mobiles suggèrent que la série présente des fluctuations autour d'une valeur moyenne. On considère que la meilleure estimation est représentée par la moyenne pondérée d'un certain nombre de valeurs antérieures (ce qui est le principe des procédures de moyennes mobiles utilisées pour le lissage de données). Ceci revient en fait à considérer que l'estimation est égale à la moyenne réelle, à laquelle on ajoute une somme pondérée des erreurs ayant entachées les valeurs précédentes.

Remarque 1.8.1. La somme d'un $AR(p_1)$ et d'un $AR(p_2)$ indépendants donnera un $MA(max(q_1; q_2))$. Voir ([17])

1.8.3 Les processus $ARMA(p, q)$ (Autorégressive Moving Average).

Les modèles $ARMA$ sont représentatifs de processus générés par une combinaison des valeurs passées et des erreurs passées.

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \cdots - \alpha_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1.43)$$

On peut aussi écrire le modèle $ARMA(p, q)$ sous la forme :

$$\alpha(B)X_t = \Theta(B)\varepsilon_t \quad (1.44)$$

où B est l'opérateur de retard et (ε_t) est le processus de bruit blanc.

Un processus $ARMA(p, q)$ est stationnaire si $\alpha(B) = 0$ à toutes ses racines en module strictement supérieur à 1.

Un processus $ARMA(p, q)$ est inversible si $\Theta(B) = 0$ à toutes ses racines en module strictement supérieur à 1.

Remarque 1.8.2. La somme d'un $AR(p_1)$ et d'un $AR(p_2)$ indépendants donnera un $ARMA(p_1 + p_2; max(p_1; p_2))$. Voir (Lubrano. M [Septembre 2008])

1.8.4 Les processus intégrés I(d).

Un processus X_t est dit intégré d'ordre d et on note $X_t \sim I(d)$, si sa différence d -ième est une représentation *ARMA* stationnaire et inversible .(et la notion de stationnarité sera définie dans la section prochaine)

Soit B l'opérateur tel que :

$$\begin{aligned} BX_t &= X_{t-1} \\ (1 - B)X_t &= X_t - X_{t-1} \end{aligned}$$

L'opérateur $1 - B$ est appelé opérateur différence première.

X_t est intégrée d'ordre d si $(1 - B)^d X_t$ est stationnaire.

Remarque 1.8.3. De façon générale voici quelques propriétés des séries intégrées

- Si $X_t \sim I(d)$ alors $a + bX_t \sim I(d)$ où a et b sont des constantes, avec $b \neq 0$.
- Si $X_t \sim I(0)$ et $Y_t \sim I(0)$ alors $aX_t + bY_t \sim I(0)$ où a et b sont des constantes.
- Si $X_t \sim I(0)$ et $Y_t \sim I(1)$ alors $aX_t + bY_t \sim I(1)$ où a et b sont des constantes, avec $b \neq 0$.

Si $X_t \sim I(d_1)$ et $Y_t \sim I(d_2)$ alors $aX_t + bY_t \sim I(\max(d_1, d_2))$ où a et b sont des constantes non nulles.

1.8.5 Les processus *ARIMA*(p, d, q).

Un processus (X_t) est dit *ARIMA*(p, d, q), tel que p, d, q sont des nombres positifs ou nuls si le processus $(1 - B)^d X_t$ est un processus *ARMA*(p, q) stationnaire.

Les processus *ARIMA* sont utiles pour des processus qui ont des corrélations positives et lentement décroissantes car cette propriété des autocorrélations peut être le signe d'une tendance dans la série.

$$\alpha(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t \quad (1.45)$$

Le processus *ARIMA*(0, 1, 0) porte le nom de marche aléatoire (Random Walk Model) . Il est souvent utilisé pour analyser l'efficience des marchés financiers.

1.9 Stationnarité dans un cadre univarié.

Lorsque chacune des variables X_t vérifie $\mathbb{E}(X_t^2) < \infty$, la loi du processus est partiellement résumée par l'espérance des différentes variables et par leurs covariances. Ces moments dépendent en général du temps, ce qui est gênant quand on veut tirer de l'information sur la loi sous jacente de ce processus. Pour pouvoir obtenir une accumulation d'information on est amené à considérer des processus dits stationnaires.

Définition 1.9.1. Un processus X_t est stationnaire au premier ordre si :

★ $\forall t \in \mathbb{Z}, \mathbb{E}(X_t) = \mu$ indépendant de t .

Définition 1.9.2. Un processus X_t est stationnaire au second ordre (stationnarité faible) si :

★ $\forall t \in \mathbb{Z}, \mathbb{E}(X_t) = \mu$ indépendant de t .

★ $\forall t \in \mathbb{Z}, \mathbb{E}(X_t^2) < \infty$.

★ $\forall t \in \mathbb{Z}, \text{Cov}(X_t, X_{t+h}) = \sigma(h)$ indépendant de t

(c'est à dire les covariances sont stables par translation du temps).

Ce type de stationnarité est aussi une propriété d'invariance des deux premiers moments par translation dans le temps. Mais on ne dit rien sur les moments d'ordre supérieur, ce qui fait que cette définition, très commode par ailleurs, est sans doute trop floue.

Au lieu de considérer simplement les deux premiers moments d'un processus, on peut décider de s'intéresser à la distribution complète des observations. On peut alors accéder à une définition plus stricte de la stationnarité.

Définition 1.9.3. La notion de stationnarité peut se définir de façon forte par une stabilité en loi du processus : quelque soit n, t_1, \dots, t_n et h , on a l'égalité entre les lois jointes

$$\mathcal{L}(X_{t_1}, \dots, X_{t_n}) = \mathcal{L}(X_{t_1+h}, \dots, X_{t_n+h}) \quad (1.46)$$

Dans ce cas , c'est la distribution jointe qui est invariante par translation. Cette propriété est plus forte que la précédente car un processus stationnaire au second ordre peut posséder des moments d'ordre supérieur qui ne sont pas invariants par translation.

Remarque 1.9.1. • Cette propriété est la plus demandée mais elle aussi la plus difficile à obtenir en pratique

• On peut l'obtenir dans le cas gaussien avec des conditions précises

1.9.1 Exemples des processus stationnaires.

1 Bruits blancs.

Ce sont les processus stationnaires les plus élémentaires. Ils ont une fonction d'auto-covariance nulle (en dehors de $\sigma(0)$). Peu intéressants en eux-mêmes pour la prévision, ils servent de base à la construction de tous les modèles présentés dans la section précédente .

2 Moyennes mobiles.

On dit que ce sont des processus obtenus à partir d'un bruit blanc, car comme nous l'avons vu précédemment ils s'écrivent sous la forme :

$$X_t = \varepsilon_t + \theta_0 - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} \quad (1.47)$$

tel que ε_t est un bruit blanc centré de variance σ^2 et q est l'ordre de la moyenne mobile.

Un processus moyenne mobile est évidemment stationnaire, et on a :

$$\begin{aligned}
 \sigma(0) &= \sigma^2(1 - \theta_1^2 - \dots - \theta_q^2) \\
 \sigma(1) &= \sigma^2(\theta_1 - \theta_1\theta_2 - \dots - \theta_{q-1}\theta_q) \\
 &\vdots \\
 \sigma(q) &= \sigma^2\theta_1\theta_q \\
 \sigma(q+1) &= \sigma^2(q+2) = \dots = 0
 \end{aligned}
 \tag{1.48}$$

3 Processus autorégressif d'ordre 1.

C'est le plus simple des processus autorégressif tel que $p = 1$, $|\alpha| < 1$ et ε_t c'est un bruit blanc centré de variance σ^2

$$X_t = \alpha X_{t-1} + \varepsilon_t \tag{1.49}$$

est stationnaire et que ε_t est non-corrélé à X_s dès que $s < t$.

D'une part, $\mathbb{E}(X_t) = \alpha\mathbb{E}(X_{t-1})$ et donc $\mu = 0$ (X est centré); d'autre part, pour $h \geq 1$:

$$\begin{aligned}
 \sigma(h) &= \text{Cov}(X_t, X_{t-h}) = \text{Cov}(\alpha X_{t-1} + \varepsilon_t, X_{t-h}) \\
 &= \alpha\sigma(h-1) = \sigma(0)\alpha^{|h|}
 \end{aligned}
 \tag{1.50}$$

1.9.2 Exemples des processus non-stationnaires.

La notion de stationnarité (**faible, ou au second ordre**) se définit par une invariance des moments d'ordre 1 et 2 au cours du temps. On peut noter que la classe des processus non-stationnaire est alors relativement vaste, et surtout hétérogène : il existe différentes sources de non-stationnarité, et à chaque origine de non-stationnarité est associée une méthode propre de stationnarisation. **Nelson** et **Plosser** ont retenu, en 1982, deux classes de processus non-stationnaires : les processus **TS (trend stationary)** et les processus **DS (difference stationary)**. Les premiers correspondent à une non-stationnarité de type déterministe, alors que les seconds correspondent à une non-stationnarité de type stochastique.

1 Les processus TS.

Définition 1.9.4. (X_t) est un processus non-stationnaire **TS** s'il peut s'écrire sous la forme

$$X_t = f(t) + Z_t \tag{1.51}$$

où $f(t)$ est une fonction (déterministe) du temps, et (Z_t) est un processus stationnaire.

Dans ce cas, le processus X_t s'écrit comme la somme d'une fonction déterministe du temps et d'une composante stochastique stationnaire, éventuellement de type *ARMA*. Dès lors, il est évident que le processus ne satisfait plus la définition de la stationnarité du second ordre. En effet, on montre immédiatement que $\mathbb{E}(X_t) = f(t) + z$ où $z = \mathbb{E}(Z_t)$, dépend du temps, ce qui viole la seconde condition de la définition d'un processus stationnaire.

Exemple 1.9.1. L'exemple le plus simple d'un processus *TS* est celui d'une tendance linéaire perturbée par un bruit blanc. On pose $f(t) = a_0 + a_1t$ et $Z_t = \varepsilon_t$ et :

$$X_t = a_0 + a_1t + \varepsilon_t$$

avec $(a_0, a_1) \in \mathbb{R}^2$, et $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ est i.i.d.. Dans ce cas, on vérifie que le processus X_t est non stationnaire puisque l'espérance, $\mathbb{E}(X_t) = a_0 + a_1t$, dépend de t . En revanche, le processus Y_t défini par l'écart entre X_t et la composante déterministe $f(t) = a_0 + a_1t$, est quand à lui stationnaire : $Y_t = X_t - a_0 - a_1t = \varepsilon_t$ et est un bruit blanc, par définition stationnaire.

Une des propriétés importantes de ce type de processus réside dans l'influence des innovations stochastiques ε_t . En effet, nous allons montrer que lorsque un processus *TS* est affecté par un choc stochastique, l'effet de ce choc tend à disparaître au fur et à mesure que le temps passe : c'est la propriété de non persistance des chocs. De façon plus formelle, cette propriété est la suivante :

Proposition 1.9.1. *L'influence d'un choc ε_t à une date T sur un processus X_t défini par :*

$$X_t = f(t) + Z_t \tag{1.52}$$

avec Z_t stationnaire et $\mathbb{E}(Z_t) = 0$, est transitoire. La tendance du modèle étant déterministe, après le choc ε_T , la séquence des x_t converge ainsi vers sa valeur de long terme définie par $f(t)$. Il n'y a pas de persistance des chocs.

*(Voir **Hurlin. C**)*

2 Les processus *DS*.

Définition 1.9.5. Un processus non stationnaire X_t est un processus *DS* (Differency Stationary) d'ordre d , où d désigne l'ordre d'intégration, si le processus filtré défini par $(1 - B)^d X_t$ est stationnaire. On dit aussi que X_t est un processus intégré d'ordre d , noté $I(d)$.

Ou bien un processus sera dit intégré d'ordre d s'il devient $I(0)$ après d différenciations. Ainsi, on peut définir une classe de processus stochastiques qui ne satisfont pas les conditions de la stationnarité, mais dont la différence à l'ordre d satisfait les propriétés de la stationnarité. Par exemple, si un processus X_t n'est pas stationnaire, on dit que ce processus est *DS*, intégré d'ordre un, noté $I(1)$, si le processus définis par la différence première $\nabla X_t = X_t - X_{t-1}$ est quant à lui stationnaire. De la même façon, le processus X_t est $I(2)$ si le processus défini par la différence seconde

$(1 - B)^2 X_t = (1 - B) \nabla X_t = X_t - 2X_{t-1} + X_{t-2}$ est stationnaire.

On comprend alors que la définition des processus DS repose sur la présence de racines unitaires dans le polynôme associé à la dynamique autorégressive du processus.

Proposition 1.9.2. *Un processus non stationnaire X_t est un processus DS intégré d'ordre d , noté $I(d)$, si le polynôme $\Phi(B)$ défini en l'opérateur retard B , associé à sa composante autorégressive admet d racines unitaires :*

$$\Phi(B)X_t = \varepsilon_t \quad \text{avec} \quad \Phi(B) = (1 - B)^b \tilde{\Phi}(B) \quad (1.53)$$

où ε_t est un processus stationnaire, et si les racines du polynôme $\tilde{\Phi}(B)$ sont toutes supérieures strictement à l'unité en module.

En effet, on pose $\Phi(B)X_t = \varepsilon_t$ avec $\Phi(B) = (1 - B)^b \tilde{\Phi}(B)$. Si l'on admet que les racines du polynôme $\tilde{\Phi}(B)$ sont inférieures à l'unité en module, ce polynôme est inversible. On peut alors écrire la différence $d^{\text{ème}}$ de X_t sous la forme d'une somme de valeurs retardées de ε_t , :

$$(1 - B)^b \tilde{\Phi}(B)X_t = \varepsilon_t \quad \Leftrightarrow \quad (1 - B)^b X_t = \Psi(B)\varepsilon_t \quad (1.54)$$

où $\Psi(B) = \tilde{\Phi}(B)$. Si ε_t est un processus stationnaire, la somme pondérée de ses valeurs passées $\Psi(B)\varepsilon_t$ est elle aussi stationnaire. Donc, en résumé, dès lors que le polynôme autorégressif $\Phi(B)$ de X_t admet d racines unitaires, la quantité $(1 - B)^b X_t$ est stationnaire, donc le processus X_t est $I(d)$.

Exemple 1.9.2. Considérons le processus $ARMA(2, 1)$ suivant $\Phi(B)X_t = \Theta(B)$ et avec $\Phi(B) = (1 - 2.5B + 1.5B^2)$ et $\Theta(B) = (1 - 0.5B)$, et ε_t i.i.d. $(0, \sigma_\varepsilon^2)$. On admet que X_t est non stationnaire et l'on cherche à déterminer si X_t est un processus $I(d)$ et quel est alors son degré d'intégration. Pour cela il suffit de déterminer le nombre de racines unitaires de $\Phi(B)$. Soient λ_1 et λ_2 les racines de $\Phi(B) = 0$, on a $\lambda_1 = 1$ et $\lambda_2 = 2/3$. Dès lors, le processus X_t est $I(1)$, en effet :

$$\Phi(B) = \left(1 - \frac{1}{\lambda_1}B\right) \left(1 - \frac{1}{\lambda_2}B\right) = (1 - B) \left(1 - \frac{3}{2}B\right) = (1 - B)\tilde{\Phi}(B)$$

où $\tilde{\Phi}(B) = \left(1 - \frac{3}{2}B\right)$ admet une racine $\lambda_2 = 2/3$ inférieure à un en module.

Remarque 1.9.2. Dans la classe générale des processus DS , un type de processus apparaît de façon régulière, si bien que l'on lui a attribué un nom particulier : la marche aléatoire.

Remarque 1.9.3. On note que la notion d'un processus $I(0)$ est plus forte que la notion de stationnarité puisqu'elle implique l'existence d'une représentation $ARMA$ inversible.

1.10 Stationnarité dans un cadre multivarié.

La notion de stationnarité (au sens strict ou stationnarité forte) et au sens faible (ou stationnarité au second ordre)) peut se généraliser au cas multivarié tout comme le cas univarié . Dans ce qui suivra , nous nous intéresserons sur la stationnarité faible.

1.10.1 Définition d'un processus multivarié.

Considérons un vecteur X_t à n composantes :

$$X_t = \begin{pmatrix} X_{1,t} \\ \vdots \\ X_{n,t} \end{pmatrix} \quad (1.55)$$

Chacune des composantes $X_{j,t}$, $j = 1, \dots, n$, $t = 1, \dots, T$, est une variable aléatoire réelle. La suite $\{X_t, t \in \mathbb{Z}\}$ constitue un processus vectoriel à n dimension.

On va noter μ le vecteur des espérances de X_t .

$$\mu = \begin{pmatrix} \mathbb{E}[X_{1,t}] \\ \vdots \\ \mathbb{E}[X_{n,t}] \end{pmatrix} = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \quad (1.56)$$

L'autocovariance de la série $\Gamma_{X_t}(h)$, est une matrice

$$\Gamma_{X_t}(h) = \mathbb{E}((X_t - \mu)(X_{t-h} - \mu)') = Cov(X_t, X_{t-h}) = (Cov(X_{i,t}, X_{j,t-h}))_{1 \leq i, j \leq n, \forall h}. \quad (1.57)$$

Pour $h = 0$, on a la matrice de variance-covariance de la série $\Gamma_{X_t}(0) = \Sigma$. Sur la diagonale de cette matrice, se trouve la variance de chaque série, et sur les éléments hors diagonaux, les covariances entre deux séries.

1.10.2 Processus vectoriels stationnaires du second ordre.

Définition 1.10.1. (Rappels sur les espaces L^2) L'espace $L^2_{\mathbb{R}^2}(\Omega, A, \mathbb{P})$ est l'espace des variables de carré intégrable (variances-covariances finies). Comme on désigne par L^2 l'espace de Hilbert, muni du produit scalaire $\langle \cdot, \cdot \rangle$ et la norme associée $\| \cdot \|$. tel que

$$\langle X; Y \rangle = \mathbb{E}[XY]$$

$$\| X \|^2 = \langle X; X \rangle = \mathbb{E}[X^2] = \mathbb{V}[X] + \mathbb{E}^2[X]$$

Définition 1.10.2. (Processus vectoriel du second ordre) (X_t) est un processus du second ordre si et seulement si :

$$\forall t, \quad X_t \in L^2_{\mathbb{R}^2}(\Omega, A, \mathbb{P}) \quad \Leftrightarrow \forall i, \forall t, \quad X_{i,t} \in L^2_{\mathbb{R}}(\Omega, A, \mathbb{P})$$

$$\Leftrightarrow X_t' X_t = \|X_t\|_2^2 = \sum_{i=1}^n X_{i,t}^2 \in L^1_{\mathbb{R}}(\Omega, A, \mathbb{P})$$

Définition 1.10.3. (Processus vectoriel stationnaire au second ordre) Soit (X_t) un processus vectoriel de second ordre.

On dit que (X_t) est stationnaire au second ordre si et seulement si :

- $\forall t \in \mathbb{Z}, \quad \mathbb{E}[X_t] = \mu$ (indépendant de t).
- $\forall (t, h) \in \mathbb{Z}^2, \quad \mathbb{E}[(X_t - \mu)(X_{t-h} - \mu)'] = Cov(X_t, X_{t-h}) = \Gamma(h)$ (indépendant de t)

Remarque 1.10.1. Comme pour le cas univarié, la stationnarité stricte implique la stationnarité faible (la réciproque n'est pas vraie sauf pour les processus gaussiens).

1.10.3 Processus bruit blanc multivarié.

Un cas particulier de processus stationnaire est le processus bruit blanc. Ce processus, est très utilisé en analyse des séries chronologiques et fait l'objet de la définition suivante :

Définition 1.10.4. Un processus bruit blanc multivarié $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un processus stochastique non corrélé de moyenne nulle et de matrice de covariance Σ finie c'est à dire :

$$\mathbb{E}[\varepsilon_t] = 0$$

$$\Gamma(h) = \mathbb{E}[\varepsilon_t \varepsilon_{t+h}'] = \begin{cases} \Sigma, & h = 0 \\ 0, & h \neq 0 \end{cases}, \quad \forall t \in \mathbb{Z}$$

1.10.4 La Fonction d'autocovariance et d'autocorrélation sous l'hypothèse de stationnarité.

Fonction d'autocovariance.

Définition 1.10.5. Soit $\{X_t, t \in \mathbb{Z}\}$ un processus multivarié stationnaire, sa fonction d'autocovariance notée $\Gamma(h)$ est définie par :

$$\Gamma(h) = \mathbb{E}[(X_t - \mu)(X_{t-h} - \mu)'] = Cov(X_t, X_{t-h}) = (Cov(X_{i,t}, X_{j,t-h}))_{1 \leq i, j \leq n}, \forall h \quad (1.58)$$

Cette quantité possède les propriétés suivantes :

- $\Gamma(h)' = \Gamma(-h)$: la matrice $\Gamma(h)$ est symétrique ,en effet :

$$\begin{aligned} \Gamma(-h) &= \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)'] \\ &= \mathbb{E}[X_t X_{t+h}'] - \mu \mu' \\ &= \mathbb{E}[X_{t-h} X_{(t-h)+h}'] - \mu \mu' \quad \text{par stationnarité} \\ &= \mathbb{E}[X_{t-h} X_t'] - \mu \mu' \\ &= \mathbb{E}[(X_t X_{(t-h)}')'] - (\mu \mu')' \\ &= [\mathbb{E}[(X_t X_{(t-h)}')] - \mu \mu']' \\ &= \Gamma(h)' \end{aligned} \quad (1.59)$$

- si on note $Cov(x_{i,t}, x_{j,t-h}) = \gamma_{i,j}(h)$, $\forall i, j, h$ alors :

$$|\gamma_{i,j}(h)| \leq \sqrt{\gamma_{i,i}(0)\gamma_{j,j}(0)}$$

Fonction d'autocorrélation.

Définition 1.10.6. La fonction d'autocorrélation de retard h , $h \in \mathbb{Z}$, d'un processus multivarié du second ordre, de moyenne $\mathbb{E}(X_t) = \mu$ et de matrice de covariance $\Gamma(h)$, notée $R(h) = (\rho_{i,j}(h))_{1 \leq i, j \leq n}$, $\forall h$ est définie par :

$$\rho_{i,j}(h) = \frac{\gamma_{i,j}(h)}{\sqrt{\gamma_{i,i}(0)\gamma_{j,j}(0)}} \quad (1.60)$$

Pour un processus stationnaire, la fonction d'autocorrélation vérifie les mêmes propriétés que la matrice d'autocovariance $\Gamma(h)$, (i.e $\rho_{i,j}(h) = \rho_{i,j}(-h)$) :

$$\rho_{i,j}(h) = \frac{\gamma_{i,j}(h)}{\sqrt{\gamma_{i,i}(0)\gamma_{j,j}(0)}} = \frac{\gamma_{i,j}(-h)}{\sqrt{\gamma_{i,i}(0)\gamma_{j,j}(0)}} = \rho_{i,j}(-h) \quad (1.61)$$

car sous l'hypothèse de stationnarité la $\Gamma(h)$ est symétrique (i.e $\gamma_{i,j}(h) = \gamma_{i,j}(-h)$).

On peut facilement montrer que : $|\rho_{i,j}(h)| \leq \sqrt{\rho_{i,i}(0)\rho_{j,j}(0)}$.

On a aussi $\rho_{i,j}(0) = 1$.

1.10.5 Classe des modèles VARMA - Notions de causalité et d'inversibilité .

Cette classe de modèles est fort importante dans la modélisation et la prévision des séries chronologiques réelles vectorielles.

Définition 1.10.7. Un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ autorégressif moyenne mobile de dimension n et d'ordres p et q , noté $VARMA(p, q)$, est défini par : X_t

$$X_t - \Phi_1 X_{t-1} - \dots - \Phi_p X_{t-p} = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \dots + \Psi_q \varepsilon_{t-q}, \quad (1.62)$$

où $\Phi_1, \dots, \Phi_p, \Psi_1, \dots, \Psi_q$ sont des matrices réelles de dimension $n \times n$ et $\varepsilon_t \sim BB(0, \Sigma)$. (BB veut dire bruit blanc)

L'équation (1.62) peut être réécrite sous la forme suivante :

$$\Phi(L)X_t = \Psi(L)\varepsilon_t, \quad t \in \mathbb{Z} \quad (1.63)$$

où L est l'opérateur de retard temporel, et

$$\Phi(L) = I_n - \sum_{i=1}^p \Phi_i B^i$$

$$\Psi(L) = I_n + \sum_{i=1}^q \Psi_i B^i$$

sont des opérateurs matriciels de dimension $n \times n$, qui sont respectivement les opérateurs autorégressifs et moyenne mobile multivariés tels que $\Phi_p \neq 0$ et $\Psi_q \neq 0$, et ε_t est un bruit blanc tel que sa matrice de covariance est Σ .

Remarque 1.10.2. * Lorsque $p = 0$, le processus $VARMA(p, q)$ devient un processus moyenne mobile multivarié, d'ordre q , noté $VMA(q)$. Dans ce cas, l'opérateur autorégressif multivarié est la matrice identité.

* lorsque $q = 0$, le processus $VARMA(p, q)$ est le processus autorégressif multivarié d'ordre p noté $VAR(p)$, dans un tel cas c'est l'opérateur moyenne mobile qui est la matrice identité.

Notions de causalité et d'inversibilité.

Définition 1.10.8. Un processus stochastique multivarié X_t est dit causal si et seulement si, il peut s'exprimer sous forme d'une combinaison (finie ou non) convergente en moyenne quadratique du passé et du présent d'un bruit blanc.

Définition 1.10.9. Un processus stochastique multivarié X_t est dit inversible si et seulement si, on peut exprimer ε_t comme combinaison linéaire (finie ou infinie) convergente en moyenne quadratique du passé et du présent du processus stochastique c'est-à-dire s'il admet une représentation autorégressive.

La causalité dans les modèle $VARMA(p, q)$.

La notion de causalité d'un processus s'applique également aux processus $VARMA(p, q)$. Un moyen pour vérifier si un processus est causal est d'étudier la nature des racines du déterminant de l'opérateur autorégressif. Ceci fait l'objet du théorème suivant :

Théorème 1.10.1. (Brockwell et Davis, 1991). Soit $\{X_t, t \in \mathbb{Z}\}$ un processus $VARMA(p, q)$ pour lequel les polynômes matriciels $\Phi(\cdot)$ et $\Psi(\cdot)$, d'ordre respectifs p et q , n'ont pas des racines communes.

Alors, le processus $\{X_t, t \in \mathbb{Z}\}$ est causal si et seulement si le polynôme matriciel $\Phi(z)$ issu de la partie autorégressive a toutes ses racines qui sont strictement supérieures à 1 en valeur absolue. Autrement dit, il y a causalité du processus $\{X_t, t \in \mathbb{Z}\}$ si et seulement si

$$\det[\Phi(z)] \neq 0 \quad \forall z \in \mathbb{C} \quad \text{tel que} \quad |z| \leq 1 \quad (1.64)$$

De plus on a :

$$X_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, t \in \mathbb{Z}, \quad \text{avec} \quad \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} < \infty \quad (1.65)$$

où les matrices Θ_j sont déterminées uniquement par :

$$\Theta(z) = \sum_{j=0}^{\infty} \Theta_j z^j = \Phi^{-1}(z)\Psi(z), \quad |z| \leq 1 \quad (1.66)$$

De plus, un processus $VARMA(p, q)$ pouvant s'exprimer comme une combinaison linéaire infinie de valeurs passées de X_t fait intervenir le concept d'inversibilité. Le théorème suivant traite de l'inversibilité.

Théorème 1.10.2. (Brockwell et Davis, 1991). Le processus $VARMA(p, q)$ représenté par l'équation précédente est inversible si :

$$\det[\Phi(z)] \neq 0 \quad \forall z \in \mathbb{C} \quad \text{tel que } |z| \leq 1 \quad (1.67)$$

Dans ce cas, le processus $VARMA(p, q)$ admet la représentation suivante :

$$\varepsilon_t = \sum_{i=0}^{\infty} \Pi_i X_{t-i} = \Pi(B)X_t \quad (1.68)$$

et les matrices Π_i sont uniquement déterminées par :

$$\Pi(z) = \sum_{i=0}^{\infty} \Pi_i z^i = \Psi^{-1}(z)\Phi(z) \quad \text{tel que } |z| \leq 1 \quad (1.69)$$

1.10.6 Le modèle VAR sous l'hypothèse de stationnarité.

Définition 1.10.10. Un processus vectoriel $\{X_t, t \in \mathbb{Z}\}$, de dimension n , admet une représentation VAR d'ordre p , notée $VAR(p)$ si :

$$X_t = c + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + \varepsilon_t \quad (1.70)$$

où c désigne un vecteur de constantes de dimension n , et $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc vectoriel de matrice de covariance Σ . Nous pouvons réécrire l'équation précédente avec l'opérateur de retard B comme suit :

$$(I_n - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p)X_t = c + \varepsilon_t \quad (1.71)$$

ou bien

$$\Phi(B)X_t = c + \varepsilon_t \quad (1.72)$$

où $\Phi(B)$ est le polynôme matriciel de retard de degré p , donné par :

$$\Phi(B) = I_n - \sum_{i=1}^p \Phi_i B^i \quad (1.73)$$

et Φ_i sont des matrices réelles satisfaisant $\Phi_0 = I_n$ et $\Phi_p \neq 0$

Conditions de stationnarité d'un modèle VAR.

Le théorème suivant donne les conditions de stationnarité d'un processus $VAR(p)$

Théorème 1.10.3. *Le processus $VAR(p)$ représenté par l'équation (1.70) est stationnaire si :*

$$\det[\Phi(z)] \neq 0 \quad \forall z \in \mathbb{C} \quad \text{tel que} \quad |z| \leq 1 \quad (1.74)$$

Dans ce cas, le processus $VAR(p)$ admet la représentation linéaire suivante :

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-1} = \Psi(B) \varepsilon_t \quad (1.75)$$

où les matrices ψ_i sont déterminées par

$$\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i = \Phi^{-1}(B) \quad (1.76)$$

et les coefficients $\{\psi_i\}$ sont absolument sommables, c'est-à-dire que :

$$\sum_{i=0}^{\infty} \|\psi_i\| < \infty$$

avec $\|\cdot\|$ correspondant à la norme euclidienne d'une matrice.

1.10.7 Représentation VMA du processus VAR .

Tout comme dans le cas univarié, sous la condition de stationnarité, il est possible d'appliquer le théorème de **Wold** et de représenter X_t sous la forme d'un processus vectoriel moyenne mobile infini $VMA(\infty)$.

Proposition 1.10.1. *Tout processus $\{X_t, t \in \mathbb{Z}\}$, de dimension n , stationnaire, satisfaisant une représentation $VAR(p)$ admet une représentation moyenne mobile convergente définie par :*

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-1} = \mu + \Psi(B) \varepsilon_t \quad (1.77)$$

avec $\mu = \mathbb{E}(X_t) = \Phi(B)^{-1}c$, où ε_t est un bruit blanc vectoriel et où la séquence des matrices de dimension $(n \times n)$, $\{\psi_i\}_{i=0}^{\infty}$ satisfait $\psi_0 = I_n$ et est absolument sommable au sens où les éléments $\psi_{j,k}^i$ de ψ_i satisfont la condition :

$$\sum_{s=0}^{\infty} |(\psi_{j,k}^i)^s| < \infty, \quad \forall i \geq 1, \quad \forall (j, k) \in \mathbb{N}^2 \quad (1.78)$$

Il est possible de déterminer de façon générale la forme des matrices de l'opérateur polynômial associée à la représentation $VMA(\infty)$.

Proposition 1.10.2. *Le polynôme matriciel $\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$ associé à la représentation $VMA(\infty)$ d'un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ telle que*

$$X_t = c + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t$$

satisfait la récurrence suivante :

$$\begin{cases} \psi_0 = I_n \\ \psi_1 - \psi_0 \Phi_1 = 0 \\ \psi_2 - \psi_1 \Phi_1 - \psi_0 \Phi_2 = 0 \\ \vdots \\ \psi_i - \sum_{j=1}^i \psi_{i-j} \Phi_j = 0 \end{cases}$$

où $\Phi_j = 0$ pour $j > p$.

Preuve 1.10.1. *Une façon simple d'obtenir la représentation $VMA(\infty)$ d'un processus $VAR(p)$ consiste à identifier les polynômes matriciels $\Phi(B)^{-1}$ et $\Psi(B)$.*

En effet, supposons que le processus soit centré (i.e $c = 0$), on a :

$$X_t = \Phi(B)^{-1} \varepsilon_t = \Psi(B) \varepsilon_t \Leftrightarrow \Phi(B) \Psi(B) = I_n \quad (1.79)$$

Cette égalité peut se réécrire sous la forme :

$$\lim_{k \rightarrow \infty} [(I_n - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p)(I_n - \Psi_1 B - \Psi_2 B^2 - \cdots - \Psi_k B^k)] = I_n \quad (1.80)$$

Ainsi par identification des termes de même ordre, on montre que les matrices de l'opérateur polynômial associées à la forme $VMA(\infty)$ satisfont bien à une équation de récurrence correspondant à celle de la proposition précédente.

Exemple 1.10.1. Soit un processus bivarié stationnaire $\{Y_t, t \in \mathbb{Z}\}$ admettant une représentation $VAR(1)$ telle que :

$$\Phi(B)Y_t = c + \varepsilon_t$$

avec $\varepsilon_t \sim N(0, \Sigma)$, $c = (3, 1)'$ et

$$\Phi(B) = \Phi_0 + \Phi_1 B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -0,2 & -0,7 \\ -0,3 & -0,4 \end{pmatrix} B = \begin{pmatrix} 1 - 0,2B & -0,7B \\ -0,3B & 1 - 0,4B \end{pmatrix}$$

Par application du théorème de Wold, on sait que ce processus peut être représenté sous une forme $VMA(\infty)$ telle que :

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-1} = \mu + \Psi(B) \varepsilon_t \quad (1.81)$$

Immédiatement, on montre que :

$$\mu = \mathbb{E}[Y_t] = \begin{pmatrix} 0,8 & -0,7 \\ -0,3 & 0,6 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 9,25 \\ 6,29 \end{pmatrix}$$

Par définition, on a $\Phi(B)\Psi(B) = I_2$, ce qui peut se réécrire sous la forme :

$$\lim_{k \rightarrow \infty} [(I_2 - \Phi_1 B)(\Psi_0 - \Psi_1 B - \Psi_2 B^2 - \dots - \Psi_k B^k)] = I_2 \quad (1.82)$$

Par identification des membres de même terme, on montre que :

$$\begin{aligned} \Psi_0 &= I_2 \\ \Psi_1 &= \Phi_1 = \begin{pmatrix} -0,2 & -0,7 \\ -0,3 & -0,4 \end{pmatrix} \\ \Psi_2 &= \Phi_1 \Psi_1 = \Phi_1^2 = \begin{pmatrix} -0,2 & -0,7 \\ -0,3 & -0,4 \end{pmatrix}^2 \end{aligned}$$

Et de façon générale, on a :

$$\Psi_n = \Phi_1 \Psi_{n-1} = \Phi_1^n = \begin{pmatrix} -0,2 & -0,7 \\ -0,3 & -0,4 \end{pmatrix}^n \quad \forall n \geq 1$$

On retrouve ainsi la formule générale que l'on avait établit précédemment :

$$Y_t = \mu + \Psi(B)\varepsilon_t = \mu + \sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}$$

Anisi on a :

$$Y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 9,25 \\ 6,29 \end{pmatrix} + \sum_{i=0}^{\infty} \begin{pmatrix} -0,2 & -0,7 \\ -0,3 & -0,4 \end{pmatrix}^i \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix}$$

1.11 Modélisation VAR en différences.

Nous allons maintenant montrer sous quelles conditions on peut modéliser un vecteur non stationnaire au moyen d'un modèle VAR en différences. On part du polynôme matriciel $A(B)$ et l'on effectue le même type de factorisation que ce que l'on a fait dans le cadre univarié, ce qui donne :

$$A(B) = (I - \Pi B) - (A_1^* B + A_2^* B^2 + \dots + A_{p-1}^* B^{p-1})(I - B)$$

avec

$$-\Pi = A_1 + \dots + A_p \quad \text{et} \quad A_s^* = \sum_{i=s+1}^p A_i$$

et comme le modèle VAR s'écrit sous cette forme :

$$A(B)X_t = \varepsilon_t \quad (1.83)$$

c-à-d

$$[(I - \Pi B) - (A_1^* B + A_2^* B^2 + \dots + A_{p-1}^* B^{p-1})(I - B)]X_t = \varepsilon_t$$

donc

$$X_t = \Pi X_{t-1} + A^*(I - B)X_t + \varepsilon_t X_t - X_{t-1} = (\Pi - 1)X_{t-1} + A^*(B)(I - B)X_t + \varepsilon_t$$

Ceci permet de considérer la reparamétrisation suivante d'un modèle *VAR*

$$\nabla X_t = (\Pi - 1)X_{t-1} + A^*(B)\nabla X_t + \varepsilon_t \quad (1.84)$$

Si $\Pi = I$, c'est à dire si $\sum_i A_i = I$, alors on pourra écrire le modèle *VAR* en différences :

$$\nabla X_t = A^*(B)\nabla X_t + \varepsilon_t \quad (1.85)$$

Comment va-t-on interpréter la condition $\sum_i A_i = I$? Elle signifie que la série X_t contient n racines unitaires, c'est à dire en fait que toutes les composantes de X_t sont non-stationnaires. Un modèle *VAR* en différences est donc un bon moyen pour modéliser des séries non stationnaires. Mais la condition $\sum_i A_i = I$ signifie plus. Elle indique qu'aucune information n'est perdue dans cette approche. On peut avoir toutes les composantes de X_t non-stationnaires et $I(1)$ sans que cette condition soit satisfaite. On rentre alors dans la problématique de la cointégration que nous allons maintenant définir dans le chapitre suivant (Chapitre 2).

Chapitre 2

La Cointégration et ses concepts.

2.1 Introduction.

La modélisation VAR précédente (chapitre 1) repose sur une interprétation du théorème de représentation de **Wold** et donc sur une hypothèse de stationnarité. Si les séries que l'on doit modéliser au moyen d'un VAR ne sont pas stationnaires, la stationnarité sera en général obtenue par différenciation des séries. On aura donc un modèle VAR en différences comme il a été montré au chapitre 1 . Mais une différenciation systématique des séries dans un modèle VAR peut conduire à perdre de l'information car il repose sur l'imposition d'une contrainte paramétrique. Par exemple si le niveau des séries est utile pour décrire la mémoire du processus, un VAR en différence requerrait un nombre infini de retards. Dans ce cas, il est donc nécessaire de conserver le niveau des séries, quitte à mélanger niveau et différences. Les modèles $CVAR$ ou VAR cointégrés appartiennent à cette classe de modèles particuliers .La notion de cointégration a été introduite dès 1974 par Engle et Newbold, sous le nom de ”**spurious regressions**”, ou régressions fallacieuses, puis formalisée par Engle et Granger en 1987, et enfin par Johansen en 1991 et 1995. Et nous allons distinguer le cas bivarié du cas général pour mettre en lumière certains problèmes spécifiques.

Notons bien que pour la simplicité de ce travail nous ne considérerons que des séries $I(0)$ ou $I(1)$. Et avant de donner la définition de la cointégration, nous allons brièvement citer cette définition,

Définition 2.1.1. Soit un processus linéaire défini par sa représentation MA infinie $X_t = C(B)\varepsilon_t$.

Ce processus est $I(0)$ si la matrice $C(1)$ est finie et strictement différente de zéro (c-à-d inversible \Rightarrow stationnarité). Un processus sera dit intégré d'ordre d s'il devient $I(0)$ après d différenciations.

2.2 Cointégration dans un cadre bivarié.

Exemple 2.2.1. Soient Y_t et X_t deux séries chronologiques vectorielles et $I(1)$, une combinaison linéaire quelconque de ces deux variables définie par :

$$Z_t = X_t - \alpha Y_t \quad (2.1)$$

sera encore $I(1)$ (d'après les propriétés des séries intégrées qu'on a vu dans le chapitre 1). Mais on peut très bien trouver pour certaines séries une valeur de α telle que Z_t soit $I(0)$. Et l'idée sous-jacente de cette exemple est la suivante :

A court terme, X_t et Y_t peuvent avoir une évolution divergente (elles sont toutes les deux $I(1)$ donc non stationnaires), mais elles vont évoluer à long terme. Il existe donc une relation stable à long terme entre X_t et Y_t . Cette relation est appelée relation de cointégration ou encore relation de long terme. Z_t mesure donc l'ampleur du déséquilibre entre X_t et Y_t et est appelée "erreur d'équilibre". Et les exemples correspondant à une telle situation sont nombreux en économie, on peut penser à la relation entre consommation et revenu, à celle liant les taux d'intérêt à court terme et long terme, ... etc.

Définition 2.2.1. On dit que le vecteur X_t de deux composantes est cointégré si

1. toutes les composantes de X_t sont $I(d)$ (même ordre d'intégration)
2. il existe un vecteur β non nul tel que $\beta' X_t \sim I(d - b)$.
avec b est un entier et $0 < b \leq d$

Le vecteur β est dit vecteur cointégrant.

2.3 Cointégration dans un cadre multivarié.

Dans le cas où $X_t \in \mathbb{R}^n$ la définition de cointégration reste la même que celle d'au dessus, sauf qu'il est très possible qu'il y ait plus d'un seul vecteur cointégrant β .

On peut supposer en fait qu'il y a r vecteurs cointégrants linéairement indépendants avec $r \leq n - 1$ que l'on rangera dans une matrice β de taille $n \times r$.

Par construction le rang de cette matrice est r qui sera appelé le rang de cointégration de X_t .

Dans le cas bivarié, pour peu qu'on le normalise, le vecteur de cointégration est défini de manière unique et est directement interprétable.

Dans le cas multivarié, cette propriété n'est plus vraie. A partir du moment où il peut y avoir plusieurs vecteurs de cointégration, ceux-ci ne sont plus uniques.

La matrice β qui est de dimension $n \times r$ est alors une base de l'espace de cointégration à r dimensions. Il est toujours possible de définir une autre matrice de cointégration par simple changement de base. Il y a un problème d'identification et d'interprétation.

2.4 Généralisation au cas trend stationnaire.

La notion de cointégration de **Engle** et **Granger** peut se généraliser au cas où $Z_t = \beta X_t$ est stationnaire autour d'une tendance.

Une série multivariée non stationnaire peut se décomposer en la somme d'une composante déterministe et d'une composante stochastique admettant une représentation $AR(p)$, soit par exemple :

$$X_t = C + \delta t + u_t \quad \text{avec} \quad A(B)u_t = \varepsilon_t \quad (2.2)$$

tel que $A(B)$ est un polynôme de retard matriciel .

On peut alors généraliser la définition de la cointégration avec **Campbell and Perron** (1991) :

Définition 2.4.1. Le vecteur aléatoire $X_t \in \mathbb{R}^n$ est cointégré s'il existe au moins un vecteur cointégrant $\beta_i \in \mathbb{R}^n$ tel que $\beta_i' X_t$ soit trend stationnaire. S'il existe r vecteurs cointégrants linéairement indépendants β_i que l'on regroupe dans la matrice cointégrante β de dimension $n \times r$, le vecteur aléatoire X_t est cointégrée de rang r .

Lorsque Engle et Granger ont étudié la notion de cointégration ,ils se sont placés dans le cas sans composante déterministe pour aboutir à un concept particulier de cointégration,et cette fois-ci on procède de la même manière qu'eux on ne définit la cointégration que sur la partie stochastique des séries (et dans cet exemple on parle de u_t).

On pourrait aussi imposer une forme plus restrictive de cointégration en demandant à ce même vecteur de cointégration β_i qui annule les racines unitaires, qu'il rende aussi constante la tendance, ce qui dans notre cas signifierait $\beta \delta t = 0$ (i.e $T_t = \mu$).

La définition de **Engle et Granger** (1987) implique que toutes les variables soient intégrées et intégrées de même ordre. La définition ci dessus est plus large car elle permet de définir une relation de cointégration entre des variables plus ou moins hétérogènes :

- premièrement, des variables qui sont toutes stationnaires autour d'une tendance seront trivialement cointégrées selon cette définition.
- deuxièmement, toute combinaison linéaire entre une variable $I(1)$ et une variable $I(0)$ est encore $I(1)$, il faut au moins deux variables $I(1)$ dans une relation pour avoir cointégration. Mais à partir du moment où l'on a cointégration, l'ajout d'une ou de plusieurs variables stationnaires autour d'une tendance ne modifie pas la propriété de cointégration.

Remarque 2.4.1. Ce que nous voulons dire dans ce dernier paragraphe est que l'on a plus de chance d'obtenir une cointégration d'après la définition précédente.

2.5 Modèles à correction d'erreur (ECM).

Les modèles dits à correction d'erreur ont été introduits pour permettre de modéliser conjointement les dynamiques de court terme (représentées par les variables en différence première) et de long terme (représentées par les variables en niveau). Engle and Granger (1987) définissent de la manière suivante un modèle à correction d'erreurs :

Définition 2.5.1. Une série temporelle vectorielle X_t a une représentation en mécanisme correcteur d'erreurs si elle peut se modéliser sous la forme :

$$A(B)\nabla X_t = \alpha\beta'X_{t-1} + \varepsilon_t \quad (2.3)$$

avec $A(0) = I$ et $A(1)$ ayant tous ses éléments finis (ou I est la matrice identité).

Ce modèle correspond en fait à un modèle VAR . C'est donc une forme réduite, et cela à cause de l'hypothèse $A(0) = I$. On appellera cette forme un modèle $VAR-ECM$. Dans le cas le plus simple, le polynôme $A(B)$ peut se réduire à la matrice identité.

β est bien sûr le vecteur de cointégration. Il peut s'interpréter naturellement comme la solution de long terme du modèle quand $r = 1$. Son interprétation est plus délicate quand $r > 1$.

La matrice α de dimension $n \times r$ représente la matrice des poids de la solution de long terme dans chaque équation du VAR .

2.5.1 Représentation.

considérons un vecteur de deux composantes X_t et Y_t .

La dynamique de court terme s'écrit :

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + v_t \quad (2.4)$$

La dynamique de long terme s'exprime de la manière suivante :

$$Y_t = aX_t + b + \varepsilon_t \quad (2.5)$$

Car à long terme, on a $Y_{t-1} = Y_t$, $X_{t-1} = X_t$ et la dynamique de court terme devient à long terme

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + v_t \\ (1 - \alpha_1)Y_t &= (\alpha_2 + \alpha_3)X_t + \alpha_0 + v_t \\ Y_t &= \frac{\alpha_2 + \alpha_3}{1 - \alpha_1} X_t + \frac{\alpha_0}{1 - \alpha_1} + \frac{v_t}{1 - \alpha_1} \end{aligned}$$

D'où

$$Y_t = aX_t + b + \varepsilon_t$$

$$\text{avec } a = \frac{\alpha_2 + \alpha_3}{1 - \alpha_1}, \quad b = \frac{\alpha_0}{1 - \alpha_1} \quad \text{et} \quad \varepsilon_t = \frac{v_t}{1 - \alpha_1} \quad (2.6)$$

Le ECM s'obtient à partir de la dynamique de court terme :

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + v_t \\ Y_t - Y_{t-1} &= \alpha_0 + \alpha_1 Y_{t-1} - Y_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + v_t \\ Y_t - Y_{t-1} &= \alpha_0 + \alpha_1 Y_{t-1} - Y_{t-1} + \alpha_2 X_t - \alpha_2 X_t + \alpha_2 X_t + \alpha_3 X_{t-1} + v_t \\ \nabla Y_t &= -(1 - \alpha_1) Y_{t-1} + \alpha_2 (X_t - X_{t-1}) + \alpha_0 + (\alpha_2 + \alpha_3) X_{t-1} + v_t \\ \nabla Y_t &= -(1 - \alpha_1) \left(Y_{t-1} - \frac{\alpha_2 + \alpha_3}{Y_{t-1}} + \frac{\alpha_0}{Y_{t-1}} \right) + \alpha_2 \nabla X_t + v_t \\ \nabla Y_t &= \gamma \nabla X_t + \mu (Y_{t-1} - a X_{t-1} - b) + v_t \end{aligned}$$

D'où

$$\alpha_2 = \gamma, \quad \mu = -(1 - \alpha_1), \quad a = \frac{\alpha_2 + \alpha_3}{1 - \alpha_1} \quad \text{et} \quad b = \frac{\alpha_0}{1 - \alpha_1} \quad (2.7)$$

Remarque 2.5.1. Si une série non stationnaire X_t est cointégrée de rang r , elle peut s'exprimer sous la forme d'un processus autorégressif à correction d'erreur .

2.6 Représentation possibles des séries cointégrées.

L'hypothèse de cointégration a des conséquences importantes pour la représentation paramétrique du processus de génération des données. Si l'on suppose que le vecteur aléatoire X_t est $I(1)$ et cointégré, alors ∇X_t est stationnaire et admet une représentation en moyenne mobile infinie. Cette représentation est cependant très particulière du fait que la matrice des coefficients d'impact totale est singulière. Cette singularité se traduira aussi au niveau de la forme AR du modèle par une autre singularité de matrice (perte de rang). Enfin, comme r relations de cointégration annulent r tendances stochastiques sur les n initialement présentes, il restera dans le modèle $n - r$ tendances stochastiques dont les propriétés vont dépendre de la forme du terme constant. C'est tous ces résultats que l'on va détailler maintenant et qui sont rassemblés par Engle and Granger (1987) dans un théorème de représentation appelé théorème de représentation de Granger. Ce théorème a été repris dans la littérature sous différentes formes, en particulier par **Johansen** (1991).

Remarque 2.6.1. Ces représentations seront ensuite utilisées dans des tests et méthode d'estimation qui seront présentés plus tard (Chapitre 3) .

2.6.1 Représentation en moyenne mobile.

Un processus multivarié stationnaire accepte une représentation de **Wold** multivariée. Considérons le vecteur aléatoire X_t de \mathbb{R}^n dont toutes les composantes sont $I(1)$. On commence par différencier la série dans toutes ses composantes et l'on peut alors écrire :

$$(1 - B)X_t = C(B)\varepsilon_t \quad (2.8)$$

où $C(B)$ est un polynôme de retards matriciel infini. Que se passe-t-il au niveau de cette représentation quand en plus X_t est cointégré, c'est à dire que $\beta'X_t \sim I(0)$? Le théorème de représentation de Granger montre que $C(1)$ est nécessairement singulier et de rang $n - r$, c'est à dire que l'on va au moins avoir $|C(1)| = 0$.

Théorème 2.6.1. *Soit un vecteur aléatoire X_t de \mathbb{R}^n intégré d'ordre un. Ce vecteur est cointégré de rang r si et seulement si la matrice $C(1)$ de sa représentation $MA(\infty)$ est de rang $n - r$.*

Preuve 2.6.1. *Pour montrer cette propriété, il suffit de se baser sur une décomposition particulière de la série. Le polynôme matriciel $C(B)$ peut toujours se réécrire, par analogie avec le cas univarié :*

$$C(B) = C(1) + (1 - B)(I_n - C(1) - C^*(B)) \quad (2.9)$$

En posant $\psi(B) = I_n - C(1) - C^*(B)$ cette décomposition permet d'écrire :

$$(1 - B)X_t = (C(1) + (1 - B)\psi(B))\varepsilon_t \quad (2.10)$$

En divisant les deux membres par $(1 - B)$ et en appelant X_0 les conditions initiales et en les posant égales à μ , on a :

$$X_t = \mu + C(1) \sum_{j=0}^{t-1} \varepsilon_{t-1} + \psi(B)\varepsilon_t \quad (2.11)$$

Si X_t est cointégré et que β est la matrice de cointégration, il vient que dans :

$$\beta'X_t = \beta'\mu + \beta'C(1) \sum_{j=0}^{t-1} \varepsilon_{t-1} + \beta'\psi(B)\varepsilon_t \quad (2.12)$$

le membre de gauche est par définition stationnaire et $I(0)$. Le membre de droite doit également l'être. Le dernier terme de droite est stationnaire par construction. Par contre le terme précédent de droite comporte une marche aléatoire vectorielle, non stationnaire par définition. Il faut donc dans cette écriture que $\beta'C(1) = 0$. Comme β est une matrice $n \times r$ de rang plein, $\beta'C(1) = 0$ implique que le rang de $C(1)$ est égal à $n - r$.

Remarque 2.6.2. Il n'est pas nécessaire de supposer que toutes les composantes de X_t sont $I(1)$. Il en suffit de deux. Les autres peuvent être $I(0)$. (d'après la généralisation au cas trend stationnaire.)

Corollaire 2.6.1. *S'il y a r relations indépendantes de cointégration entre les composantes de X_t , la matrice $C(1)$ associée à la représentation de **Wold** a les deux propriétés :*

$$\beta' C(1) = 0$$

$$C(1)\alpha = 0$$

où β est la matrice $n \times r$ contenant les r vecteurs de cointégration et α la matrice $n \times r$ de poids.

Remarque 2.6.3. Le fait que la matrice $C(1)$ soit singulière implique que la représentation en moyenne mobile (2.8) n'est pas inversible. Il ne sera donc pas possible de trouver une représentation AR sur les différences qui ait un nombre fini de retards.

2.6.2 Représentation en tendances communes.

Une série qui est $I(1)$ présente une tendance stochastique, c'est à dire qu'elle peut s'écrire sous la forme d'une accumulation de chocs passés. Dans le cas multivarié, une série $I(1)$ de dimension n comportera n tendances stochastiques qui pourront être éventuellement corrélées. La propriété de cointégration introduit une contrainte forte entre les tendances stochastiques au point de dire que les composantes de X_t vont avoir des tendances communes. Il y aura donc moins que n tendances stochastiques distinctes. Plus précisément, si le rang de cointégration est r , r tendances stochastiques auront disparu et il ne restera que $n - r$ tendances distinctes qui deviendront alors les tendances communes aux n composantes de X_t .

L'écriture qui met en avant ces $n - r$ tendances communes est due à **Stock and Watson (1988)** et on peut la formaliser dans le théorème suivant :

Théorème 2.6.2. *Toute série $X_t \in \mathbb{R}^n$ cointégrée de rang r peut se mettre sous la forme d'une combinaison linéaire de n trends (tendance) déterministes δt , de $n - r$ marches aléatoires ou tendances stochastiques F_{τ_t} et n termes aléatoires stationnaires $\psi(B)\varepsilon_t$:*

$$X_t = \mu + \delta t + F_{\tau_t} + \psi(B)\varepsilon_t \quad (2.13)$$

Preuve 2.6.2. (Voir *Lubrano. M [Octobre 2008]*)

2.6.3 Représentation autorégressive.

Le fait que $C(1)$ soit singulière implique qu'un modèle VAR en différences n'est pas compatible avec l'hypothèse de cointégration. Mais un VAR sur les niveaux est parfaitement possible. Considérons une série temporelle multivariée X_t de \mathbb{R}^n que l'on décompose en :

$$X_t = \mu + \delta t + u_t \quad \text{tel que} \quad A(B)u_t = \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma) \quad (2.14)$$

On peut combiner les deux équations de ce système en

$$A(B)(X_t - \mu - \delta t) = \varepsilon_t \quad (2.15)$$

$A(B)$ est un polynôme matriciel de retards avec $A(B) = I - A_1B - A_2B^2 - \dots - A_pB^p$. On cherche à obtenir une expression similaire à celle utilisée pour les tests de Dickey-Fuller Augmenté **ADF** (nous allons traiter au section prochaine) univariés qui nous permettra une discussion sur les propriétés du système en cas de cointégration des composantes de X_t . On a besoin pour cela de la généralisation multivariée d'une des formules de factorisation de $A(B)$ donnée dans le chapitre 1 et qui précise que :

$$A(B) = A(1)B + (1 - B)(I_n - A^*(B)) \quad (2.16)$$

En posant $\Pi = -A(1)$ on obtient :

$$\nabla X_t = \delta - \Pi(X_{t-1} - \mu - \delta(t-1)) + A^*(B)(\nabla X_t - \delta) + \varepsilon_t \quad (2.17)$$

qu'il est plus commode de réécrire en :

$$\nabla X_t = m + \Pi X_{t-1} + A^*(B)(\nabla X_t - \delta) + \varepsilon_t \quad (2.18)$$

avec

$$m = -\Pi\mu + \delta$$

$$X_{t-1}^* = X_{t-1} - \delta(t-1)$$

Comme on ne considère que le cas où X_t comporte au plus une seule racine unitaire dans chacune de ses composantes, ∇X_t sera toujours stationnaire. Dans quels cas alors le membre de droite du modèle sera-t-il stationnaire ? Cela ne dépend que de ΠX_{t-1}^* et donc du rang de Π .

- ◇ Si Π est de rang plein, on a n combinaisons linéaires indépendantes de ΠX_{t-1}^* qui doivent être stationnaires. Les composantes de X_t ne peuvent pas être cointégrées, car on ne peut avoir au plus que $n-1$ relations de cointégration indépendantes. Pour que ΠX_{t-1}^* soit stationnaire, il faut que chaque composante de X_t soit stationnaire autour d'une tendance. On a donc un modèle *VAR* qui doit s'écrire en niveau avec une tendance.
- ◇ Si Π est de rang nul, on a $\Pi = 0$. Il n'existe aucune combinaison linéaire de X_t qui soit trend stationnaire. On doit donc écrire le modèle *VAR* en différence.
- ◇ Le dernier cas est le plus intéressant. Le rang de Π est égal à r compris entre 1 et $n-1$. A ce moment là on peut décomposer la matrice Π en produit de deux matrices de dimension $n \times r$ α et β :

$$\Pi = \alpha\beta' \quad (2.19)$$

Pour que ΠX_{t-1}^* soit stationnaire, il suffit que $\beta' X_{t-1}^*$ le soit. Dans ce cas, la matrice β de rang plein et égal à r , est la matrice de cointégration. C'est elle qui rend X_t stationnaire autour d'une tendance. On peut imposer un concept plus fort de cointégration déterministe en demandant que les tendances disparaissent avec $\beta'\delta = 0$.

On peut alors donner le théorème suivant :

Théorème 2.6.3. *Soit un vecteur aléatoire X_t de \mathbb{R}^n dont toutes les composantes sont $I(1)$ et admettant la représentation autorégressive :*

$$A(B)(X_t - \mu - \delta t) = \varepsilon_t$$

Ce vecteur est cointégré de rang r si et seulement si dans la factorisation :

$$\nabla X_t = m - A(1)X_{t-1} + A^*(B)(\nabla X_t - \delta) + \varepsilon_t \quad (2.20)$$

la matrice $A(1)$ est singulière, de rang r et peut se mettre sous la forme :

$$-A(1) = \alpha\beta' \quad (2.21)$$

où α et β sont deux matrices de dimension $n \times r$.

2.7 Théorème de représentation de Granger.

Quand on est dans un modèle *VAR* simple, le passage de la forme autorégressive $A(B)\nabla X_t = \varepsilon_t$ à la forme moyenne mobile $\nabla X_t = C(B)\varepsilon_t$ se fait par inversion du polynôme $A(B)$. Comme la cointégration implique que les matrices $A(1)$ et $C(1)$ ont subi une perte de rang, cette inversion directe n'est plus possible.

Le théorème de représentation de **Granger** donné dans **Engle and Granger (1987)** permet de montrer comment on peut quand même opérer un passage entre la représentation *AR* et la représentation *MA* du modèle. D'autres auteurs ont donné une formulation alternative de ce théorème et en particulier **Johansen (1991)**.

Nous reprendrons ici la présentation de **Boswijk (1992)**. Partons d'un modèle *VAR* sur les niveaux (les dynamiques de long terme) :

$$A(B)X_t = m + \varepsilon_t \quad (2.22)$$

Remplaçons dans (2.22) le polynôme $A(B)$ par sa factorisation (2.16) :

$$\nabla X_t = m + A(1)X_{t-1} + A^*(L)\nabla X_t + \varepsilon_t. \quad (2.23)$$

Cela nous conduit directement à représentation en modèle correcteur d'erreur si $A(1)$ est de rang r . Le passage de cette forme *AR* à la forme *MA* se fait au moyen du théorème suivant :

Théorème 2.7.1. *Soit X_t un processus stochastique de dimension n admettant la représentation *VAR* de (2.22) avec l'équation caractéristique associée à $A(B)$ ayant $n - r$ racines égales à l'unité et toutes les autres en dehors du cercle unité. Supposons que (2.22) admette la représentation en terme correcteur d'erreurs (2.23) avec $A(1) = -\alpha\beta'$. Alors :*

- ∇X_t est stationnaire de représentation de Wold :

$$\nabla X_t = C(B)(m + \varepsilon_t) = \mu + C(B)\varepsilon_t \quad (2.24)$$

Le polynôme de retard $C(B)$ se décompose en :

$$C(B) = C(1) + C^*(B)(1 - B) \quad (2.25)$$

et est relié au polynôme $A(B)$ par :

$$C(1) = \beta_{\perp}(\alpha'_{\perp}[I - A(1) - A^*(1)]\beta_{\perp})^{-1}\alpha'_{\perp} \quad (2.26)$$

où α_{\perp} et β_{\perp} sont les compléments orthogonaux de α et β , chacun de dimension $n \times (n - r)$ vérifiant $\alpha'_{\perp}\alpha = 0$ et $\alpha'_{\perp}\alpha = 0$. Le rang de $A(1)$ est égal à r et le rang de $C(1)$ est égal à $n - r$.

- X_t est non-stationnaire et $I(1)$ de représentation :

$$X_t = X_0 + C(1)(mt + \sum_{j=1}^t \varepsilon_t) + C^*(B)\varepsilon_t \quad (2.27)$$

- $\beta'X_t$ est stationnaire. Les colonnes de β sont les vecteurs de cointégration et r est le rang de cointégration.
- Les tendances communes sont représentées par $\alpha_{\perp} \sum_{j=1}^t \varepsilon_t$

On trouvera une preuve d'une version plus élaborée de ce théorème dans **Johansen (1991)** et la preuve de cette version-ci dans **Boswijk (1992)**. Notons que cette version du théorème de représentation de **Granger** part de l'existence d'une représentation en terme correcteur d'erreurs et en déduit la propriété de cointégration. La version originale du théorème donnée dans **Engle et Granger (1987)** procède de manière strictement inverse et montre que l'hypothèse de cointégration implique l'existence d'une représentation en terme correcteur d'erreurs.

Remarque 2.7.1. Autrement dit la version originale du théorème de représentation de Granger présente l'intérêt de faire la synthèse entre les deux approches :

- L'approche **ECM** issue de l'idée de concilier des préoccupations de théorie économique avec une écriture rigoureuse des équations économétriques
- L'approche **VAR** issue d'une approche purement statistique, de type "boîte noire".

Chapitre 3

Tests de cointégration.

3.1 Introduction.

Comme nous l'avons vu dans le chapitre précédent ,la cointégration est la stationnarité d'une combinaison linéaire des composantes non stationnaires du vecteur aléatoire ,la cointégration est une propriété qui permet donc de réduire le nombre de trend stochastiques dans un vecteur aléatoire.

Par rapport au cas univarié, la représentation autorégressive de la série vectorielle devient nettement plus complexe, même si une écriture matricielle laisse entrevoir une certaine similitude. Les tests de cointégration dans les modèles autorégressifs avec cointégration deviennent eux aussi plus complexes.Et dans ce contexte deux approches ont été distinguées dans la littérature .

1. **L'approche *ECM* ou bien l'approche de Engle and Granger (1987)** : Elle permet de tester si une régression statique est cointégrante ou non. Mais elle ne permet pas d'examiner des restrictions sur les vecteurs de cointégration,et pour cette raison la méthode a connu des développements avec Phillips (1991) ,mais elles restent des méthodes qu'on peut qualifier d'univariées car elles n'ont clairement défini que le cas où il n'y a qu'un seul vecteur de cointégration (ce que nous appelons par la cointégration univariée)
2. **L'approche *VAR* ou bien l'approche de Johansen (1988) qui a été développé par la suite dans Johansen (1991), Johansen and Juselius (1990) et Johansen (1995)** : C'est elle qui est majoritairement utilisée dans les logiciels, et qu'on peut qualifier de multivariée .

Comme ces deux approches font appellent aux tests de racine unitaire univariés (tests de stationnarité) auxquels nous avons consacré la première partie, afin de mieux saisir l'approche d'Engle et Granger (1987) dans la seconde section .

3.2 Tests de Racine Unitaire.

La présence d'une racine unitaire dans les données a des conséquences très importantes sur le plan statistique. On peut les résumer en quelques lignes :

- Les propriétés asymptotiques générales des estimateurs (vitesse de convergence, normalité asymptotique) ne tiennent plus. c'est à dire il faut avoir recours à une théorie asymptotique spéciale.
- La présence de régresseurs comportant une racine unitaire dans une régression peut conduire à estimer des régressions apparemment très bonnes entre des variables qui sont totalement indépendantes entre elles.
- Une série trend stationnaire et une série stationnaire en différence se comportent de manière radicalement opposée dans le long terme, alors que :
 Une série trend stationnaire a tendance à se repositionner autour de son trend déterministe après un choc aléatoire , (C'est ce que l'on appelle par la propriété de "**mean reversion**").
 Et une série stationnaire en différence ne revient pas autour de sa tendance après un choc, puisque le choc affecte aussi la tendance stochastique de la série.

Il existe donc un ensemble de raisons statistiques variées pour s'intéresser à la présence (ou l'absence) d'une racine unitaire dans une série, pour cela on doit savoir si la racine existe, et si c'est le cas dans quelle partie du processus .

On peut décomposer toutes séries temporelles en deux composantes : une composante déterministe TD_t et une composante stochastique u_t tel que :

$$y_t = TD_t + u_t \quad (3.1)$$

En fait ce qui importe principalement dans la partie déterministe c'est son ordre .On distingue trois cas :

1. $TD_t = 0$ ou pas de partie déterministe
2. $TD_t = \mu$ ou seulement un terme constant.
3. $TD_t = \mu + \delta t$ ou cette fois-ci une constante et un trend.

Par contre L'hypothèse de racine unitaire concerne la partie stochastique u_t .

Et comme la racine unitaire se trouve dans la partie stochastique u_t , on va modéliser cette partie stochastique au moyen d'un processus *ARMA* :

$$\tilde{A}(B)u_t = C(B)\varepsilon_t \quad (3.2)$$

A cause de la présence du terme TD_t , on peut supposer que la moyenne de u_t est nulle, la moyenne du processus étant contenue dans TD_t .

On va supposer que la partie moyenne mobile est inversible, c'est à dire que toutes les racines de $C(B)$ sont en dehors du cercle unité. De plus on va supposer que le polynôme rationnel infini $C^{-1}(B)\tilde{A}(B)$ peut s'approximer par un polynôme $A(B)$ de degré fini et que

donc il suffit de considérer un processus autorégressif pour la partie stochastique donné par :

$$A(B)u_t = \varepsilon_t \quad (3.3)$$

On peut maintenant distinguer deux modèles alternatifs pour y_t .

- * Dans le premier cas y_t est trend stationnaire, si toutes les racines de l'équation caractéristique $A(z) = 0$ sont en dehors du cercle unité. Le processus de u_t est donc stationnaire et celui de y_t est stationnaire autour d'un trend.
- * Dans le deuxième cas où y_t est stationnaire en différence, si $A(z) = 0$ comporte une racine sur le cercle unité et toutes les autres racines sont en dehors du cercle unité. On a donc la factorisation :

$$A(B) = (1 - B)A^*(B) \quad (3.4)$$

et $(1 - L)u_t$ est stationnaire. Dans ce cas, $(1 - L)y_t$ est stationnaire autour d'une moyenne fixe.

3.2.1 Un modèle général pour les tests.

Le but des tests de racine unitaire est donc de tester la présence d'une racine unitaire dans la partie autorégressive de la partie stochastique du processus de y_t . Pour faire ce test il est nécessaire de trouver un modèle qui contienne à la fois les deux modèles représentatifs que l'on a dégagé, trend stationnaire et stationnaire en différence de manière à ce qu'une hypothèse puisse apparaître comme une restriction paramétrique dans le modèle général. Le modèle que l'on a adopté peut se réécrire après substitution :

$$A(B)(y_t - TD_t) = \varepsilon_t$$

Il est très intéressant de développer cette écriture dans un cas simple, celui où $A(B) = (1 - \rho B)$ et prenons le cas général $TD_t = \mu + \delta t$. Il devient :

$$(1 - \rho B)(y_t - \mu + \delta t) = \varepsilon_t \quad (3.5)$$

$$y_t = \rho y_{t-1} + (1 - \rho)(\mu + \delta t) + \rho\delta + \varepsilon_t$$

Enfin, il est habituel de considérer un modèle plus simple pour le test qui est :

$$y_t = \rho y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t \quad (3.6)$$

avec :

$$\beta_0 = (1 - \rho)\mu + \rho\delta \quad (3.7)$$

$$\beta_1 = (1 - \rho)\delta \quad (3.8)$$

Cette écriture est généralement employée pour les tests car l'estimateur de ρ dans les deux cas (les deux modèles représentatifs de y_t que l'on a dégagé) est numériquement le même. Seuls changent les estimateurs des coefficients de TD_t .

3.2.2 Tests de Dickey-Fuller DF .

Le modèle de régression que l'on a mis en avant :

$$y_t = \rho y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t \quad (3.9)$$

est appelée régression de Dickey-Fuller à la suite de leur papier de 1981. L'idée d'un test de racine unitaire est très simple. Il suffit d'estimer cette régression par la méthode des moindres carrés et de tester ensuite $\rho = 1$ au moyen de la **statistique de Student**.

On soustrait y_{t-1} des deux côtés de la régression (3.9) :

$$\nabla y_t = (\rho - 1)y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t \quad (3.10)$$

Deux statistiques de test.

En partant de la régression de test (3.10) , on va pouvoir employer deux statistiques de test. La première est la **statistique de Student** usuelle, que l'on va noter τ , pour bien souligner le fait qu'elle a une distribution asymptotique qui n'est pas normale :

$$\tau = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\rho_T}} \quad (3.11)$$

La seconde statistique est basée sur le fait que $z = T(\hat{\rho}_T - 1)$ converge en distribution vers une **fonctionnelle de Browniens** qui ne dépend pas du paramètre de nuisance σ^2 .

L'écriture de cette statistique sous $H_0 : \rho = 1$ est :

$$z = T(\hat{\rho}_T - 1) = \frac{T^{-1} \sum y_{t-1} \varepsilon_t}{T^{-1} \sum y_{t-1}^2} \xrightarrow{\mathcal{L}} \frac{0.5(W^2(1) - 1)}{\int_0^1 W^2(r) dr} \quad (3.12)$$

Avec :

$W(r)$: est un mouvement Brownien standard.

T : la taille de l'échantillon à simuler.

Et la distribution de τ sous $H_0 : \rho = 1$ est :

$$\tau \xrightarrow{\mathcal{L}} \frac{0.5(W^2(1) - 1)}{(\int_0^1 W^2(r) dr)^{1/2}} \quad (3.13)$$

La seule différence, c'est que le dénominateur de la distribution de τ est égale à la racine carrée du dénominateur de la distribution de z . Ces deux statistiques auront donc des comportements similaires, mais des valeurs critiques différentes. La statistique τ est la plus utilisée. Mais la statistique z va servir de base au test de **Philipps-Perron**.

La distribution des tests en fonction de la composante stochastique TD_t .

◇ **Cas sans terme déterministe** : si $TD_t = 0$ alors le modèle s'écrit :

$$\nabla y_t = (\rho - 1)y_{t-1} + \varepsilon_t \quad (3.14)$$

L'hypothèse nulle et l'hypothèse alternative sont :

$$\begin{cases} H_0 : \rho = 1, \\ H_1 : \rho < 1. \end{cases}$$

◇ **Cas avec constante** : si $TD_t = \mu$ le modèle se complique un peu. On a :

$$\nabla y_t = (\rho - 1)y_{t-1} + \beta_0 + \varepsilon_t \quad (3.15)$$

L'hypothèse nulle n'est plus la même que précédemment. On veut toujours tester $\rho = 1$ contre $\rho < 1$ mais l'autre paramètre doit aussi être considéré. On a :

$$\begin{cases} H_0 : \rho = 1 \quad \beta_0 = 0, \\ H_1 : \rho < 1. \end{cases}$$

◇ **Cas avec trend** : si enfin $TD_t = \mu + \delta t$ le modèle est le suivant :

$$\nabla y_t = (\rho - 1)y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t \quad (3.16)$$

L'hypothèse nulle a encore un peu changé. Cette fois-ci on a :

$$\begin{cases} H_0 : \rho = 1 \quad \beta_1 = 0, \\ H_1 : \rho < 1. \end{cases}$$

3.2.3 Test de Dickey-Fuller augmenté ADF .

Les tests développés jusqu'à présent ne reposent pas sur l'hypothèse de normalité des erreurs. Ils sont valides, du moins asymptotiquement, sous des hypothèses relativement générales concernant les erreurs. Mais ils ne sont plus valables dès que les erreurs sont autocorrélées. On a dans la littérature deux types de corrections. La première revient, dans une approche paramétrique, à modifier la régression de test. La seconde consiste à modifier la statistique de test proprement dite.

Cependant, nous allons adopter la première correction car il est relativement facile de modifier la régression de test pour tenir compte de l'autocorrélation.

Le modèle autorégressif de base sur lequel on s'est basé est le suivant :

$$A(B)(y_t - TD_t) = \varepsilon_t \quad (3.17)$$

Nous nous limitons pour l'instant au cas où $A(B)$ était un polynôme de degré un. On considère $A(B)$ de degré $p \geq 1$.

Factorisons ce polynôme $A(B)$ selon la formule explicitée dans le chapitre 1 :

$$A(B) = (1 - \rho B) - (1 - B)A^*(B)$$

où $A^*(B)$ est un polynôme de degré $p - 1$ sans terme constant :

$$A^*(B) = \alpha_1 B + \alpha_2 B^2 + \dots + \alpha_{p-1} B^{p-1}$$

avec $\rho = 1 - A(1)$.

On va supposer que le polynôme $A^*(B)$ a toutes ses racines en dehors du cercle unité et s'intéresser à tester l'hypothèse nulle d'une seule racine unitaire $\rho = 1$.

Effectuons la multiplication entre la factorisation de $A(B)$ et $(y_t - TD_t)$:

$$(1 - \rho B)(y_t - TD_t) = A^*(B)(\nabla y_t - \nabla TD_t) + \varepsilon_t \quad (3.18)$$

ce qui montre que l'on obtient le même type de modèle que précédemment, mais que cette fois-ci on a simplement rajouté les différences premières retardées de y_t .

Et pour bien voir cette simple différence prenons un exemple dans le cas où $p = 2$ et $TD_t = (\mu + \delta t)$ on a $A^*(B) = a_1^* B$ et on obtient :

$$y_t = \rho y_{t-1} + (1 - \rho)(\mu + \delta t) + (\rho + a_1^*)\delta + a_1^* \nabla y_{t-1} + \varepsilon_t \quad (3.19)$$

Sous l'hypothèse nulle $\rho = 1$, on retrouve les mêmes propriétés de la régression à savoir que le trend (*i.e.* δt) disparaît, que le terme constant initial (*i.e.* μ) disparaît et est remplacé par une fonction du paramètre du trend (*i.e.* $(\rho + a_1^*)\delta$).

Le test de l'hypothèse nulle s'effectue de la même manière que précédemment au moyen d'une statistique de Student que l'on va noter τ, τ_μ et τ_τ par analogie avec le cas précédent :

- $DT_t = 0$.
- $DT_t = \mu$
- $DT_t = \mu + \delta t$

Ce type de correction paramétrique a été proposée initialement par **Dickey and Fuller (1979)**, d'où l'appellation de test de Dickey-Fuller augmenté ou ADF (augmentation de la régression initiale des retards de ∇y_t).

3.2.4 Test de Phillips – Perron PP.

Si le test ADF apporte une correction paramétrique au test de Dickey – Fuller simple en prenant en compte l'éventuelle autocorrélation des erreurs, le test de **Phillips – Perron (1988)**, quant à lui, propose une correction non paramétrique au test de Dickey – Fuller simple afin de régler le problème de l'hétéroscédasticité des erreurs, sans avoir à ajouter des variables endogènes en différences retardées comme dans les tests ADF.

La procédure du test reste donc basée sur les trois modèles autoregressifs du test de Dickey et Fuller simple, repris ci-après, et consiste à tester l'existence de racine unitaire [$H_0 : \rho = 0$] contre celle de stationnarité [$H_1 : \rho < 0$].

1. $\nabla y_t = \rho y_{t-1} + \varepsilon_t$

2. $\nabla y_t = \rho y_{t-1} + \mu + \varepsilon_t$
3. $\nabla y_t = \rho y_{t-1} + \mu + \delta t + \varepsilon_t$

La statistique de test de Phillips-Perron est une statistique de Student corrigée de la présence d'autocorrélation par la prise en compte d'une estimation de la variance de long terme de ε_t , robuste à la présence d'autocorrélation et d'hétéroscédasticité.

Phillips et Perron (1988) ont donc suggéré d'ajouter à la statistique de Student du coefficient autorégressif ρ , un facteur de correction afin d'éliminer les paramètres de nuisance, associés à l'existence de corrélations dans la composante stochastique du processus générateur de données [PGD], qui perturbent les résultats des tests de Dickey et Fuller simple.

L'estimation de cette variance de long terme est donnée par :

$$\psi = \mu_\varepsilon(0) + 2 \sum_{j=1}^q \mu_\varepsilon(j) \quad (3.20)$$

ou $\mu_\varepsilon(j)$, $j = 0, 1, \dots, q$, est le coefficient d'autocovariance d'ordre j de ε_t et q est le paramètre de troncature [pour ne pas calculer les coefficients d'autocovariance jusqu'à un ordre infini]. On préfère en général utiliser la correction de **Newey-West** :

$$\psi = \mu_\varepsilon(0) + 2 \sum_{j=1}^q \frac{q+1-j}{q+1} \mu_\varepsilon(j) \quad (3.21)$$

On calcule le paramètre de troncature q par les formules suivantes :

- Formules de Schwert (1989), qui propose deux valeurs de q , notées q_4 et q_{12} ,

$$q_4 = \text{int} \left[4 \left(\frac{N}{100} \right)^{\frac{1}{4}} \right] \quad q_{12} = \text{int} \left[12 \left(\frac{N}{100} \right)^{\frac{1}{4}} \right] \quad (3.22)$$

où N est le nombre d'observations et $\text{int}[y]$ désigne la partie entière de y .

- Formule de Newey et West (1984), qui suggèrent de choisir q tel que :

$$q = \text{int} \left[4 \left(\frac{N}{100} \right)^{\frac{2}{9}} \right] \quad (3.23)$$

- Plus généralement, on retient :

$$q = N^{\frac{1}{4}} \quad (3.24)$$

Toutefois, en pratique, ces différentes formules conduisent à des résultats proches, ce qui pose donc moins d'exigence sur la formule à choisir de quatre susmentionnées. De plus, si les résultats sont quasiment identiques, on choisira, par parcimonie, la formule donnant lieu au paramètre q le moins élevé.

3.3 L'approche de Engle et Granger (1987).

L'approche de Engle et Granger (1987) comme nous l'avons mentionner à l'introduction ne fournit aucune garantie qu'un vecteur cointégrant unique à estimer, on l'appelle aussi par l'approche de **OLS** (Ordinary Least Square-Moindres Carrés Ordinaires)

3.3.1 Estimation des modèles à correction d'erreur :Méthode d'estimation en deux étapes.

La méthode d'estimation proposée par Engle et Granger (1987) procède en deux étapes. On estime tout d'abord le vecteur de cointégration par la méthode des moindres carrés ordinaires, puis en remplaçant ce vecteur par son estimation dans le modèle général, on estime dans une seconde étape les coefficients dynamiques du modèle.

- **Première étape** :La première étape consiste à régresser par simples moindres carrés ordinaires une des composantes de la variable multivariée sur les autres pour obtenir une estimation du vecteur de cointégration. Supposons que le modèle multivarié s'écrive :

$$A(B)(\nabla X_t - \nabla TD_t) = \alpha\beta'(X_{t-1} - \nabla TD_{t-1}) + \varepsilon_t \quad (3.25)$$

et que le rang de cointégration soit égal à un. β est alors un vecteur colonne.

On va choisir d'en normaliser le premier élément et de partitionner X_t de manière conforme en :

$$\begin{aligned} \beta' &= [1, \tilde{\beta}] \\ X_t &= [X_{1t}, X_{2t}] \end{aligned}$$

Tel que X_{1t} et X_{2t} dont à la séparation de X_t endeux groupes, où X_{1t} est une variable scalaire $I(1)$ et X_{2t} est un vecteur $(m \times 1)$ de variable $I(1)$.

Alors pour estimer $\tilde{\beta}$, il suffit d'utiliser la méthode des moindres carrés ordinaires dans la régression :

$$X_{1t} = \tilde{\beta}'X_{2t} + TD_t + \tilde{\varepsilon}_t \quad (3.26)$$

- **Deuxième étape** : La méthode d'estimation consiste à reporter dans le modèle **ECM-VAR** l'estimation du vecteur de cointégration et d'estimer par la méthode des moindres carrés ordinaires les paramètres du modèle :

$$A(B)(\nabla X_t - \nabla TD_t) = \alpha\tilde{\beta}'(X_{t-1} - \nabla TD_{t-1}) + \varepsilon_t \quad (3.27)$$

ou une des équations du modèle ECM structurel correspondant.

Le fait d'utiliser l'estimation de $\tilde{\beta}'$ au lieu de sa vraie valeur n'apporte aucune perturbation

Afin de simplifier cette approche :

Exemple 3.3.1. Nous considérons le cas d'une séries bivarié, telque ses composantes sont X_t et Y_t de même ordre $I(1)$.

Donc reprenons les étapes précédente :

1. en estimant par la méthode des moindres carrés ordinaires la relation de long-terme :

$$Y_t = \tilde{b} + \tilde{b}X_t + \tilde{\varepsilon}_t; \quad (3.28)$$

2. puis en estimant, toujours par la méthode des moindres carrés ordinaires la relation de court-terme :

$$\nabla Y_t = \lambda \nabla X_t + \mu \hat{\varepsilon}_{t-1} + \eta_t. \quad (3.29)$$

Le coefficient μ doit être significativement négatif, dans le cas contraire, on rejette l'hypothèse d'une modélisation de la forme *ECM*.

3.3.2 Tests univariés de cointégration (de Engle et Granger).

Ce test est construit en deux étapes :

Etape 1 : Tester l'ordre d'intégration des variables. Une des conditions nécessaires pour qu'il y ait cointégration étant que les composantes de la série doivent être intégrées de même ordre. Ces tests reposent sur l'utilisation des tests de *DF* et *ADF*.

Et si on reprend l'exemple précédent on doit tester si $Y_t \sim I(1)$ et $X_t \sim I(1)$

Etape 2 : Estimation de la relation de long-terme :

$$Y_t = b + aX_t + \varepsilon_t \quad (3.30)$$

En réécrivant l'équation (3.30) comme suit, on a :

$$\varepsilon_t = Y_t - b - aX_t \quad (3.31)$$

Ainsi, le test valide la présence de cointégration si le vecteur résiduel issu de cette estimation est stationnaire.

Remarque 3.3.1. (Tests de cointégration basés sur les résidus.)

En fait le test de cointégration sur la relation de long terme n'est qu'un test des résidus ε_t , donc il suffit de tester la présence d'une racine unitaire dans ces résidus pour tester la cointégration. Ce test se fait au moyen d'un test du type de celui de Dickey et Fuller souvent appelé test de Engle et Granger (**EG**)

Mais comme l'autocorrélation est souvent un problème, on emploiera un test **AEG** (Augmented Engle Granger) qui revient à rajouter à la régression de test des retards de $\nabla \varepsilon_t$.

Plusieurs tables pour les valeurs critiques des tests de cointégration ont été publiées à commencer dans l'article de Engle and Granger (1987). On peut citer aussi Engle and Yoo (1987) qui complète les tables précédentes. Mais on préférera se fier à celle de **MacKinnon (1991)** qui semble plus précise.

La méthode de Engle et Granger (1987) nous permet d'estimer aisément un modèle à correction d'erreur en deux étapes. Elle fournit également un certain nombre de tests de cointégration faciles à mettre en oeuvre. L'inconvénient de cette approche est qu'elle

	5%	10%
n=2		
sans trend	-3.34	-3.04
avec trend	-3.78	-3.50
n=3		
sans trend	-3.74	-3.45
avec trend	-4.12	-3.84
n=4		
sans trend	-4.10	-3.81
avec trend	-4.43	-4.15
n=5		
sans trend	-4.41	-4.13
avec trend	-4.72	-4.44
n=6		
sans trend	-4.71	-4.42
avec trend	-4.98	-4.70

Ces valeurs ont été calculées à partir de MacKinnon (1991).

FIG. 3.1: Valeurs critiques asymptotiques pour le test AEF de cointégration.

ne permet pas de distinguer plusieurs relations de cointégration. En effet si on étudie simultanément n variables avec $n > 2$, on peut avoir jusqu'à $(n - 1)$ relations de cointégration.

La méthode de Engle et Granger (1987) ne nous permet d'obtenir qu'une seule relation de cointégration.

Afin de pallier cette difficulté, Johansen a proposé une approche multivariée de la cointégration fondée sur la méthode du maximum de vraisemblance.

3.4 Approche multivariée de la cointégration : l'analyse de Johansen.

Considérons un vecteur X_t contenant n ($n \geq 2$) variables toutes $I(1)$. La représentation $VAR(p)$ de X_t est donnée par :

$$X_t = \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \varepsilon_t \quad (3.32)$$

avec Φ_i matrice de dimension $(n \times n)$ et $\varepsilon_t \sim N(0, \Sigma)$ On peut réécrire ce modèle sous forme d'un *VECM* (Vector Error Correction Model). Pour cela considérons l'équation suivante :

$$\begin{aligned} X_t - X_{t-1} &= (\Phi_1 - I)X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + \varepsilon_t \\ \Leftrightarrow \nabla X_t &= (\Phi_1 - I)(X_{t-1} - X_{t-2}) + (\Phi_1 + \Phi_2 - I)X_{t-2} + \dots + \Phi_p X_{t-p} + \varepsilon_t \\ \Leftrightarrow \nabla X_t &= (\Phi_1 - I)\nabla X_{t-1} + (\Phi_1 + \Phi_2 - I)(X_{t-2} - X_{t-3}) + \dots + \Phi_p X_{t-p} + \varepsilon_t \end{aligned}$$

Et ainsi de suite. On se ramène finalement à une représentation susceptible d'être un *VECM* :

$$\nabla X_t = B_1 \nabla X_{t-1} + B_2 \nabla X_{t-2} + \cdots + B_{p-1} \nabla X_{t-p+1} + \Pi X_{t-1} + \varepsilon_t$$

où les matrices B_i sont fonctions des matrices Φ_i et

$$\Pi = \Phi_1 + \Phi_2 + \cdots + \Phi_p + I = \sum_{k=1}^p \Phi_k - I \quad (3.33)$$

Pour illustrer cette généralisation, considérons l'exemple d'un *VAR(2)* :

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \varepsilon_t \quad (3.34)$$

On obtient de cette façon l'équation :

$$X_t - X_{t-1} = \Phi_1 X_{t-1} - X_{t-1} + \Phi_2 X_{t-2} + \Phi_2 X_{t-1} - \Phi_2 X_{t-1} + \varepsilon_t \quad (3.35)$$

\Leftrightarrow

$$\begin{aligned} \nabla X_t &= (\Phi_1 - I)X_{t-1} - \Phi_2(X_{t-1} - X_{t-2}) + \Phi_2 X_{t-1} + \varepsilon_t \\ \nabla X_t &= -\Phi_2 \nabla X_{t-1} + (\Phi_1 + \Phi_2 - I)X_{t-1} + \varepsilon_t \\ \nabla X_t &= B_1 \nabla X_{t-1} + \Pi X_{t-1} + \varepsilon_t \end{aligned} \quad (3.36)$$

où $B_1 = -\Phi_2$ et $\Pi = \Phi_1 + \Phi_2 - I$

De façon générale, la matrice Π et B_i peuvent s'écrire sous la forme :

$$\Pi = \left(\sum_{k=1}^P \Phi_k \right) - I_n \quad (3.37)$$

et

$$B_i = \left(\sum_{j=1}^i \Phi_j \right) - I_n \quad (3.38)$$

La matrice Π peut s'écrire en produits de deux matrices α et β'

$$\Pi = \alpha \beta' \quad (3.39)$$

où le vecteur α est la force de rappel vers l'équilibre et β le vecteur de paramètres d'ajustement (vecteur de coefficients des relations de long terme).

Le test de Johansen sera donc centré sur l'examen de Π la matrice de coefficients de long terme. À l'équilibre, tous les ∇X_{t-i} sont nuls et de ce fait $\Pi \nabla x_{t-k}$. Dans cette optique de croissance équilibrée, le test de cointégration à la Johansen est calculé en cherchant le rang de la matrice Π via ses valeurs propres.

Remarque 3.4.1. Le rang de la matrice Π détermine le nombre de relations de cointégration présentes entre les n variables du vecteur X_t . donc

- Si le rang de la matrice Π (c'est à dire le nombre de colonnes linéairement indépendantes) est égal à la dimension n du VAR , alors toutes les variables du VAR sont stationnaires ($I(0)$) et le problème de la cointégration ne se pose pas.
- Si en revanche, le rang de la matrice Π satisfait : $1 \leq r \leq n - 1$ alors il existerait r relations de cointégration et la représentation $VECM$ sera valide :

$$\nabla X_t = B_1 \nabla X_{t-1} + B_2 \nabla X_{t-2} + \dots + B_{p-1} \nabla X_{t-p+1} + \alpha \beta' X_{t-1} + \varepsilon_t$$

3.4.1 Test du nombre de relations de cointégration.

Afin de déterminer le nombre de vecteurs de cointégration r , à l'aide de la méthode de maximum de vraisemblance, Johansen parvient à proposer deux tests pour déterminer les valeurs propres non nulles correspondant aux r relations de cointégration :

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad (3.40)$$

et

$$\lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_i) \quad (3.41)$$

où T est le nombre d'observations, r est le nombre de vecteur de cointégration sous l'hypothèse nulle et $\hat{\lambda}_i$ est la valeur estimée de la i ème valeur propre de la matrice Π .

Nous ne développerons ici que le test de la trace car il est le plus utilisé.

Test de la trace.

La statistique $\lambda_{trace}(r)$ est associée au test où l'hypothèse nulle suppose que le nombre de vecteur de cointégration est inférieur ou égal à r contre l'hypothèse alternative qu'ils soient supérieurs à r (pour le test de la valeur propre maximale, $H_0 : r = r$ contre $H_1 : r = r + 1$). $\lambda_{trace}(r)$ suit une loi de probabilité tabulée par Johansen et Juselius. Et ce test fonctionne par exclusion d'hypothèses alternatives :

$$\text{Test 1} : = \begin{cases} H_0 : r = 0, & \text{l'hypothèse aucune relation de cointégration} \\ H_1 : r > 0, & \text{il existe au moins une relation de cointégration} \end{cases}$$

- Si $\lambda_{trace}(0)$ est supérieur à la valeur lue dans la table au seuil α , on rejette H_0 , il existe au moins une relation, on passe alors à l'étape suivante, sinon on arrête et $r = 0$.

$$\text{Test 2} : = \begin{cases} H_0 : r = 1, & \text{l'hypothèse d'une seule relation de cointégration} \\ H_1 : r > 1, & \text{il existe plus d'une relation de cointégration} \end{cases}$$

- Si $\lambda_{trace}(1)$ est supérieur à la valeur lue dans la table au seuil α , on rejette H_0 , il existe au moins deux relations, on passe alors à l'étape suivante, sinon on arrête et $r = 1$.

⋮
⋮

Et ainsi de suite jusqu'à la dernière étape (si elle est nécessaire) :

$$\text{Test } n : = \begin{cases} H_0 : r = (n - 1), & \text{l'hypothèse de (n-1) relations de cointégration} \\ H_1 : r > (n - 1), & \text{il existe au moins (n-1) relation de cointégration} \end{cases}$$

- Si $\lambda_{trace}(n - 1)$ est supérieur à la valeur lue dans la table au seuil α , on rejette H_0 , il existe au moins n relations, (en fait dans ce cas les n variables sont $I(0)$) sinon $r = n - 1$.

En conclusion, on soulignera que l'intérêt de la théorie de la cointégration est qu'elle fournit une méthode d'analyse des séries chronologiques non stationnaires. De plus, grâce aux modèles à correction d'erreur, elle permet de modéliser simultanément les dynamiques de long terme et de court terme des séries temporelles.

3.5 Quelques commentaires à propos de la puissance des tests de cointégration .

Souvent, le macroéconomiste a le choix entre différents échantillons de données pour certaines séries chronologiques. En particulier, il peut avoir recours à des séries dont la fréquence d'échantillonnage varie selon la période retenue .Ainsi, avec des données macroéconomiques, on dispose souvent d'observations trimestrielles pour la période commençant à la fin de la seconde guerre mondiale, tandis que la série mensuelle ne commence qu'au début des années soixante .Par contre, avec des données annuelles, les échantillons couvrent en générale des horizons plus longs. En règle générale, un échantillon de données annuelles contient donc environ 100 observations, un échantillon de données trimestrielles plus de 160 observations et un échantillon de données mensuelles, plus de 300 . **Il est intéressant de savoir lequel de ces échantillons donne les tests les plus puissants .Est-ce qu'un nombre plus élevé d'observations donne nécessairement des tests plus puissants ?**

Comme nous l'avons vu, certains tests de cointégration sont directement des tests de racines unitaire appliqués à des résidus estimés. On peut donc extrapoler certains resultats obtenus pour les tests de racines unitaires pour une variable directement aux tests de l'hypothèse nulle d'absence de cointégration .Pour les tests d'hypothèses d'une racine unitaire

contre les alternatives stationnaires, il s'avère que la puissance des tests dépend beaucoup plus de l'horizon temporel retenu que du nombre d'observations .

- Pour un nombre donné d'observations, la puissance s'accroît avec un accroissement de l'horizon temporel des données .
- Pour un horizon donné du temps, l'ajout d'observations supplémentaires obtenues un échantillon à une plus grande fréquence qui ne permet qu'un accroissement marginal de la puissance, cette accroissement est devenant négligeable lorsque l'intervalle entre les observation décroît .
- Dans le cas d'applications contenant petit nombre de données annuelles sur une longue période, on obtient des tests plus puissants que dans le cas d'échantillons contenant plus d'observations sur une durée plus courte .

Ces résultats démontrent que les tests de racine unitaire et les tests de cointégration devraient être construits en utilisant les données annuelles couvrant une longue période historique .

Par contre, l'utilisation de données dans un passé plus lointain peut créer des problèmes d'un autre ordre.

Premièrement : Il se peut que la qualité des données historiques soit douteuse et que les méthodes antérieures de construction d'indices introduisent un biais contre l'une ou l'autre des hypothèses.

Deuxièmement : L'utilisation d'un échantillon couvrant une longue période rend plus probable la présence dans la série étudiée d'un changement structurel majeur du processus caractérisant soit la composante tendancielle soit la composante stochastique . Un tel changement structurel crée un biais en faveur de l'hypothèse de racine unitaire .

Donc, même si l'utilisation d'un échantillon couvrant une longue période est souhaitable du point de vue de la puissance des tests, il faut interpréter les résultats avec prudence à cause de ces effets secondaires possibles .

3.6 Exemple d'application sous R .

Exemple 3.6.1. Dans cet exemple nous allons mettre en oeuvre les tests vus dans ce chapitre en traitant deux séries chronologiques, prix de la douzaine d'oeufs et d'un poulet, aux Etats Unis durant 1930 à 1983, et les données relatives sont :

```
year chic egg
1930 468491 3581
1931 449743 3532
1932 436815 3327
1933 444523 3255
1934 433937 3156
```

1935 389958 3081
1936 403446 3166
1937 423921 3443
1938 389624 3424
1939 418591 3561
1940 438288 3640
1941 422841 3840
1942 476935 4456
1943 542047 5000
1944 582197 5366
1945 516497 5154
1946 523227 5130
1947 467217 5077
1948 499644 5032
1949 430876 5148
1950 456549 5404
1951 430988 5322
1952 426555 5323
1953 398156 5307
1954 396776 5402
1955 390708 5407
1956 383690 5500
1957 391363 5442
1958 374281 5442
1959 387002 5542
1960 369484 5339
1961 366082 5358
1962 377392 5403
1963 375575 5345
1964 382262 5435
1965 394118 5474
1966 393019 5540
1967 428746 5836
1968 425158 5777
1969 422096 5629
1970 433280 5704
1971 421763 5806
1972 404191 5742
1973 408769 5502
1974 394101 5461
1975 379754 5382
1976 378361 5377
1977 386518 5408

```
1978 396933 5608
1979 400585 5777
1980 392110 5825
1981 384838 5625
1982 378609 5800
1983 364584 5656
```

Leur présentation graphique est obtenue par le programme suivant :

```
>> Egd<-read.table("eggs.txt", header=T)  "eggs.txt" c'est le fichier où les
données sont enregistrées .
>>attache(Egd)
>>library(stats)
>>chic<-ts(chic, start=1930 , f=1)
>> egg<-ts(egg, start=1930, f=1)
>> layout(1 :2)
>> plot.ts(egg)
>>plot.ts(chic)
```

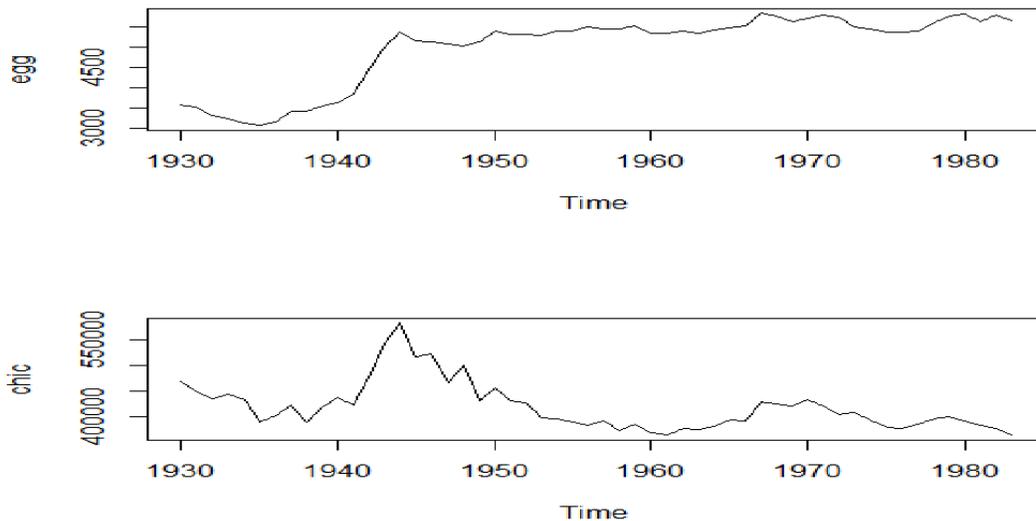


FIG. 3.2: prix de la douzaine d'oeufs et d'un poulet, durant 1930 à 1983

A partir de ces présentations graphiques on peut suspecter qu'il n'y a pas une relation de cointégration entre prix de la douzaine d'oeufs et d'un poulet, durant 1930 à 1983 car ces

deux séries ne peuvent pas être ajustées par une droite .Or cette proposition n'est qu'une hypothèse qu'il faut confirmer ou infirmer par les tests de cointégration.

3.6.1 Méthode de Engle et Granger (1987) sous R .

Cette méthode se fait en trois étapes .

(i) Tester l'ordre d'intégration des variables (egg, chic), et pour cela sous **R** il y a plusieurs méthodes, prenons par exemple :

1.

```
>> library(urca)
>>chic.df <- ur.df(y=chic, lags=1, type='trend')
>>summary(chic.df)
```

```
.
                        Augmented Dickey-Fuller Test Unit Root Test

Test regression trend

Call : lm(formula = z.diff    z.lag.1 + 1 + tt + z.diff.lag)

Coefficients :
.
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.329e+04 4.260e+04  1.955 0.0564 .
z.lag.1      -1.821e-01 9.112e-02 -1.998 0.0514 .
tt           -3.156e+02 2.670e+02 -1.182 0.2429
z.diff.lag   -8.620e-02 1.435e-01 -0.601 0.5510
---
Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error : 25030 on 48 degrees of freedom Multiple
R-squared : 0.1067, Adjusted R-squared : 0.05085 F-statistic :
1.911 on 3 and 48 DF, p-value : 0.1404

Value of test-statistic is : -1.998 1.4499 2.0188

Critical values for test statistics :
.
  1pct 5pct 10pct
tau3  -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```

L'observation statique est inférieure aux extrémités des valeurs critiques, alors on accepte l'hypothèse H_0 : la série **chic** est non stationnaire

2.

```
>> library(tseries)
>> adf.test(egg, k=1)
```

```
.                               Augmented Dickey-Fuller Test
```

```
data : egg
Dickey-Fuller = -1.6336, Lag order = 1, p-value = 0.7221
alternative hypothesis : stationary
```

le p-value est supérieure à 0.05 donc on accepte H_0 et on rejette l'alternative. On passe par conséquent à la différenciation des deux séries, et les résultats du test de **Dickey-Fuller** appliqué aux séries différenciées sont données ci-dessous :

```
>> library(tseries)
>> require(quadprog)
>> chic1 <- -diff(chic)
>> egg1 <- -diff(egg)
```

```
.                               Augmented Dickey-Fuller Test
```

```
data : egg1
Dickey-Fuller = -4.3367, Lag order = 1, p-value = 0.01
alternative hypothesis : stationary
Message d'avis :
In adf.test(egg1, k = 1) : p-value smaller than printed p-value
```

```
>> adf.test(chic1, k=1)
>> adf.test(egg1, k=1)
```

```
.                               Augmented Dickey-Fuller Test
```

```
data : chic1
Dickey-Fuller = -4.3796, Lag order = 1, p-value = 0.01
alternative hypothesis : stationary
```

```

Message d'avis :
In adf.test(chic1, k = 1) : p-value smaller than printed p-value

```

On conclue que les deux séries différenciées sont stationnaires car les p-values sont toutes inférieures à 0.05.

Ainsi on a deux séries qui sont $I(1)$ (même ordre), donc on peut envisager l'existence de relations de cointégration entre elles.

(ii) L'estimation de la relation de long-terme (**ECM**) :

```

>>Engle< -lm(chic egg)
>>summary(Engle)

```

```

Call :
lm(formula = chic    egg)

```

Coefficients :

```

.           Estimate Std. Error t value Pr(>|t|)
(Intercept) 470461.481 36111.963 13.028 <2e-16 ***
egg        -10.219   .      7.133 -1.433 0.158
---

```

```

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error : 45950 on 52 degrees of freedom
Multiple R-squared : 0.03798, Adjusted R-squared : 0.01948
F-statistic : 2.053 on 1 and 52 DF, p-value : 0.1579 .

```

(iii) On extrait le vecteur résiduel issu de cette estimation par :

```

>>residual< - resid(Engle)

```

Et enfin de cette étape on teste la stationnarité de vecteur résiduel (i.e : si le vecteur résiduel est $I(0)$ ou non) .

```

>> adf.test(residual, k=1)

```

```

.           Augmented Dickey-Fuller Test

```

```

data : residual
Dickey-Fuller = -2.0247, Lag order = 1, p-value = 0.5645
alternative hypothesis : stationary

```

le p-value est supérieur à 0.05, donc on rejette l'altirnative (i.e : résiduel n'est pas un $I(0)$) .

On conclue que les deux séries (chic, egg) sont pas cointégré.

3.6.2 Méthode de Johansen (1988) sous R :

(i) Les deux séries sont intégrées et de même ordre $I(1)$, ceci est vérifié dans la méthode précédente .

(ii) Test du nombre de relations de cointégration (vecteurs de cointégration) .

```
>> library(urca)
>> chegg=cbind(chic,egg)
>>chegg.vecm <- ca.jo(chegg,ecdet ="const", type="trace", K=2, spec="longrun",season=4)
>> summary(chegg.vecm)
```

```
.                Johansen-Procedure
```

```
Test type : trace statistic , without linear trend and constant in cointegration
Eigenvalues (lambda) : [1] 0.2675294 0.1013629 0.0000000
```

```
Values of teststatistic and critical values of test :
```

```
.          test 10pct 5pct 1pct
r <= 1 | 5.56 7.52 9.24 12.97
r = 0 | 21.75 17.85 19.96 24.60
```

```
Eigenvectors, normalised to first column : (These are the cointegration
relations) .
```

	chic.l2	egg.l2	constant
chic.l2	1.00000	1.00000	1.00000
egg.l2	4.85545	88.10556	-47.42672
constant	-393810.49290	-896077.46799	-226055.04523

```
Weights W : (This is the loading matrix)
```

	chic.l2	egg.l2	constant
chic.d	-1.638589e-01	-0.046540914	-1.142952e-15
egg.d	-2.771808e-05	-0.000542545	-4.821596e-18

On conclue que les deux séries (chic, egg) ne sont pas cointégrées aussi, d'après le test Johansen (la trace ou valeur propre maximale), donc l'hypothèse qu'on a mis est correcte.

Remarque 3.6.1. Si on veut utiliser le test de la valeur propre maximale il suffit de remplacer dans type **trace** par **eigen** .

3.7 Exemple du marché du merlu en Bretagne .

dans cette exemple nous allons traiter une application faite par l'économiste des pêches "José Antonio Pérez Agûndez" qu'a analysé l'étendue d'un marché en étudiant l'évolution des prix d'une seule espèce, le merlu, sur plusieurs marchés afin de déterminer s'il existe ou pas une évolution similaire des prix sur l'ensemble des marchés à long terme.

3.7.1 Le choix du marché .

Le merlu est une des espèces structurantes de la pêche en France (Bretagne). Elle est toujours classée parmi celles qui ont un poids important pour le revenu des pêcheurs, d'où l'importance de l'étude de ce marché. Une très forte chute de la production a été observée depuis le début de la décennie. Comme on peut voir dans la figure (3.3), la production de merlu des années quatre-vingt tournait autour des 20 000 tonnes par an alors qu'en 1998, la production a été de 6 539 tonnes.

Comme on doit savoir que le merlu est passé de l'espèce la plus importante en valeur de la pêcherie française en 1992 à la sixième position en 1998 .

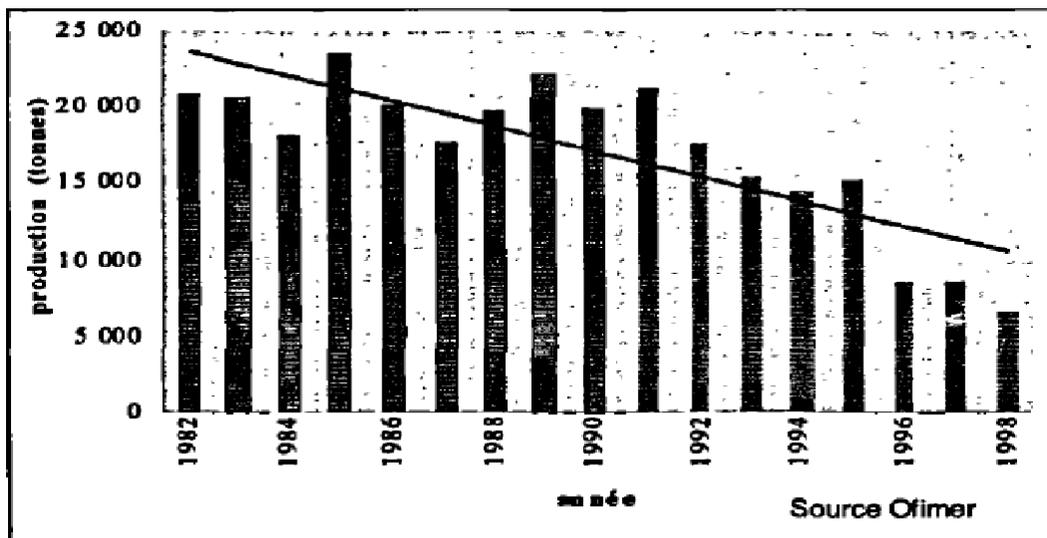


FIG. 3.3: Production française de merlu.

3.7.2 le choix de l'échantillon et des données .

Il a pris comme échantillon des criées distribuées spatialement dans le sud de la Bretagne, lesquelles représentent, grosso modo, la moitié de la production nationale .

Les données utilisées pour cette étude ont été obtenues auprès de l'Ofimer qui alimente et gère la base **RIC** (Réseau Inter Criée : Base de données enregistrant toutes les ventes de produits de mer des criées en France). On utilisera des séries de prix moyens mensuels ou bien des séries de prix hebdomadaires correspondant aux ventes de merlu par criée durant 1994 à 1999, toutes catégories confondues. On dispose d'information sur les 10 principaux ports de Bretagne produisant du merlu qui sont les suivants :

Classement	Port	Quantité (tonnes)
1	LORIENT	1433
2	LATURBALLE	834
4	CONCARNEAU	798
5	LE GUILVINEC	701
9	SAINT GUENOLE	515
11	LOCTUDY	497
18	LESCONIL	177
21	DOUARNENEZ	93
22	QUIBERON	93
36	AUDIERNE	5

Tableau. 3.1 : Classement et production moyenne annuelle de merlu dans les principales criées en Bretagne (période 1994-1997) .

3.7.3 La méthode économétrique de cointégration .

La méthode économétrique de cointégration est un outil courant pour déterminer l'étendue d'un marché, cette méthode a été traitée dans la littérature en utilisant des espèces différentes pour un même marché afin de tester des interdépendances entre elles .

Et dans cette application, il a effectué dans un premier temps des estimations en construisant des modèles bivariés afin d'avoir un premier élément sur les relations bilatérales entre les criées. Cependant, cette analyse étant restrictive, on a élaboré des modèles multivariés en utilisant différentes séries de prix pour déterminer d'une manière plus ample si l'ensemble du marché breton de merlu peut être considéré cointégré et donc évoluant sous une même tendance.

3.7.4 Application de la méthode .

Après l'application du test ADF aux variables étudiées, on a constaté le rejet de l'hypothèse alternative de la stationnarité .Seule la variable prix de « Le Guilvinec » a été trouvée stationnaire.

Des tests ADF ont été également appliqués aux variables en différences premières afin de tester l'existence de plus d'une racine unité. Les résultats montrent l'existence d'une seule racine unité car toutes les variables deviennent stationnaires en différences premières .

Ces tests nous amènent à supprimer la variable prix de « Le Guilvinec » de l'analyse car, pour que plusieurs variables soient cointégrées il est nécessaire qu'elles soient intégrables au même ordre. Or, les prix de « Le Guilvinec » sont $I(0)$ et ceux des autres ports sont

$I(1)$.

Maintenant on peut Comparer les variables par couples afin d'avoir une première vision générale de l'existence de relations de cointégration entre les variables retenues. Il s'agit de faire une première analyse pour déterminer les variables susceptibles de faire partie du groupe des variables cointégrées, et pour cela l'utilisation de l'approche de Johansen pour des tests bivariés a pour avantage (par rapport à l'approche de Engle Granger) de considérer toutes les variables comme endogènes .

La totalité des tests effectués est présentée dans le tableau suivant :

	Douarnenez	Audierne	St Guénole	Lesconil	Loctudy	Concarneau	Lorient	Quiberon
Douarnenez								
Audierne	-							
St Guénole	*	*						
Lesconil	C	C	C					
Loctudy	C	C	C	C				
Concarneau	C	*	*	C	C			
Lorient	-	*	C	C	C	C		
Quiberon	*	C	C	C	C	C	C	C
La Turballe	-	*	C	C	C	C	C	C

(C) : cointégration entre les deux variables

(*) : existence d'un vecteur de cointégration, mais une des variables n'est pas statistiquement significative

(-) : inexistence de cointégration entre les deux variables.

Tableau. 3.2 : Résultats des tests bivariés de cointégration .

D'où dans une première lecture de ce tableau, on peut voir que de manière générale les marchés sont assez bien cointégrés par couples. Ainsi, jï Agûndez l_i a trouvé des relations stationnaires de long terme dans 73% des combinaisons possibles. Dans 22% des cas a trouvé un vecteur de cointégration dans le modèle estimé, mais des tests d'exclusion lui amène au rejet d'une des variables. Pour seulement 1% des cas il n'a pas trouvé aucune relation de cointégration entre les deux variables du modèle.

On peut donc interpréter dans un premier temps ces résultats comme de forts indices de cointégration dans l'ensemble.

Remarque 3.1. (Les variables statistiquement significatives .)

Une fois la variable dépendante et les variables explicatives candidates sélectionnées, nous avons de bonnes raisons de les inclure dans notre modèle. Cependant, après avoir exécuté notre modèle, nous constatons que certaines de nos variables explicatives sont statistiquement significatives et que d'autres ne le sont pas (On dit que notre variable explicative est statistiquement significative si cette variable explicative peut améliorer notre modèle).

Et pour savoir la quelles de ces variables explicatives sont significatives, on utilise les Moindres carrés ordinaires (test d'exclusion) qui calcule un coefficient pour chaque une de ces variables explicatives et qui calcule la probabilité de ce dernier qu'il soit réellement égal à zéro (ou très proche de zéro) et que cette variable explicative n'est pas significative, ou bien le coefficient est statistiquement significatif avec un niveau de confiance de 95 pour cent et que cette variable explicative est importante pour notre modèle \square .

Cependant, cette analyse est insuffisante pour conclure des inter relations entre la totalité des variables retenues pour cette étude. il a effectué ainsi des combinaisons de groupes de ports pour lesquels il a testé l'existence de cointégration afin d'avoir un cadre d'analyse plus large. Les combinaisons effectuées sont les suivantes :

(Il y a sept combinaisons de quatre ports)

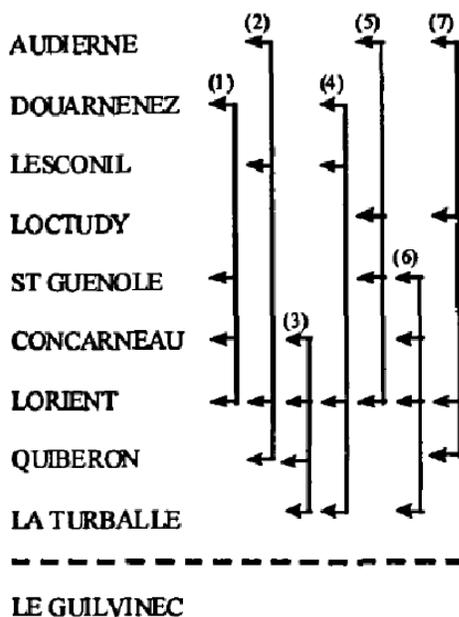


FIG. 3.4: Groupes formés pour tester la cointégration .

Les remarques découlent des résultats :

- Dans tous les exemples de groupes construits il a trouvé au moins un vecteur de cointégration reliant les variables de chaque modèle. Dans tous les cas, le test de trace et le test de valeur propre maximum convergent vers la significativité d'un même nombre de vecteurs de cointégration.
- Des tests d'exclusion ont été effectués et cela permet de déterminer si les variables intégrantes du vecteur de cointégration sont statistiquement significatives, autrement dit, si toutes les variables rentrent significativement dans la construction de la relation stationnaire de long terme. Ces tests assurent la robustesse des relations de cointégration.
Tous les tests d'exclusion appliqués aux variables de chaque groupe n'ont pas montré d'exclusion quelconque.
- Il a trouvé d'autres combinaisons de ports qui ne se trouve pas sur la figure précédente pour lesquels des tests d'exclusion il ont amené au rejet de certaines variables. Cependant, les différents groupes présentés (ainsi que d'autres obtenus) peuvent être interprétés comme significatifs d'un marché cointégré dans son ensemble. Autrement dit, les sous-marchés intégrant le marché breton varient sous une même tendance à long terme malgré des chocs ponctuels à court terme .Il s'agit donc d'un marché géographiquement unifié.
- Les ports étant plus faiblement cointégrés sont ceux ayant un moindre niveau de production de merlu. C'est le cas de Douamenez et de Audieme. Plus la production d'un port est forte et plus les débouchés de cette production pourront s'étendre dans une aire géographique plus importante.

On conclue, en mettant la cointégration comme outil économétrique elle nous a permis d'étudier la délimitation géographique d'un marché .D'où les résultats mettent en évidence l'existence d'un marché breton de merlu cointégré dans son ensemble malgré l'intégration plus faible de ports tels que Douamenez et Audieme à très faible production de merlu.

Conclusion générale .

Au début de ce travail, nous avons abordé la problématique de la cointégration des séries chronologiques dans le cadre univarié d'une manière générale et spécifiquement au phénomène de stationnarité dans le cadre multivarié .Mais nous avons constaté que la maîtrise des phénomènes temporels est beaucoup plus puissante et intéressante dans le cas multivarié car on peut mettre en évidence l'évolution d'un phénomène en liaison avec d'autres .Comme on a vu dans les exemples étudiés, la cointégration s'avère une méthode satisfaisante dans la mesure où des séries non-stationnaires sont étudiées.

L'utilisation de l'approche de Johansen dans les tests de cointégration est plus avantageux que l'approche de Engle et Granger qui considère toutes les variables comme endogènes, et qui nous permet d'examiner des restrictions sur les vecteurs de cointégration, et pour cette raison ; est qualifié de multivarié .

Bibliographie

- [1] Adouka L, Kchirid El M, Benzekoura A et Bouguelli Z, "Modélisation de la fonction de la demande des dépenses publiques Algérienne à l'aide de modèle ECM", European Scientific Journal August (2013) edition vol.9, No.22 ISSN : 1857 – 7881, (Print) e - ISSN 1857- 7431, PP.159 - 181 .
- [2] Aragon. Y, " Séries temporelles avec R , Méthodes et cas", © Springer-Verlag France, 2011.
- [3] Brockwell, P.J. et Davis, R.A[1991], "Time Series :Theory and Methodes", Springer verlag .
- [4] Cannault J,"Séries Temporelles linéaires ", polycopié de deuxième année Master, ENSAE .
- [5] Charpentier A, "Séries temporelles : théorie et applications", ENSAE,University Paris DAUPHINE .
- [6] Dickey, D. A. et Fuller, W. A. [1979], "Distribution of the Estimators for Autoregressive Times Series with a Unit Root",Journal of the American Statistical Association, 74, PP. 427-431 .
- [7] Engle, R. F. et Granger, C. W. J.[1987],"Co-Integration and Error Correction :Representation, Estimation and Testing", Econometrica, 55,PP.251-276 .
- [8] Engle, R. F. et Yoo, B. S.[1987], "Forecsting and Testing in Cointegrated Systems", Journal of Econometrics, 35, PP.143-159 .
- [9] Hall, R. E.[1978], "Stochastic Implication of the Life Cyclé-Permanent Income Hypothesis :Theory and Evidence", Journal of Polical Economy, 86, PP . 971-987 .
- [10] Hurlin. C, "Séries Temporelles", Polycopie de Cours de Tronc Commun, U.F.R. Economie Appliquée.
- [11] Johansen, S.[1991],"Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", Econometrica, 59, PP. 1551-1580 .
- [12] Johansen, S. et Juselius, K.[1990], "Maximum Kikelihood Estimation and Inference on Cointegration, with Application to the Demand for Money", Ox Ford Bulletin of Economics and Statics, 52, PP .169-210 .
- [13] Jonathan D. Cryer et Kung-Sik Chan,"Time Series Analysis With Applications in R", © 2008 Springer Science+Business Media, LLC .
- [14] Lubrano. M, "Introduction à la modélisation des séries temporelles univariées", Polycopié, Septembre 2008.
- [15] Lubrano. M, "Inférence et Tests dans les Modèles Cointégrés", Octobre 2007.
- [16] Lubrano. M, "Modélisation mutilvariée et cointégration", Polycopié, Octobre 2008.
- [17] Lubrano. M, "Tests de Racine Unitaire",Polycopie, Septembre 2008.

-
- [18] Maddala. G. S et Kim. In-Moo, "Unit Roots Cointegration and And Structural Change", © Cambridge University Press (1998).
- [19] Mbaye Baïdy Baro et Sene Serigne Moustapha, Document de Travail "Analyse et prévision des séries temporelles par la methode de BOX et JENKINS", DPEE/DSC/BPTCT@Décembre 2007.
- [20] Nelson, C. R. et Plosser, C. I. [1982], "Trends and Random Walks in Macroeconomic Time Series", Journal of Monetary Economics, 10, PP . 139-162 .
- [21] Perron P et Campbell .John Y, "Racines unitaires en macroéconomie :le cas multidimensionnel", Annales d'économie et de statistique - N°27 - 1992 .
- [22] Peter J. Brockwell et Richard A. Davis, "Introduction to Time Series and Forecasting", © 2002, 1996 Springer-Verlag New York, Inc.
- [23] Robert H. Shumway et David S. Stoffer, "Time Series Analysis and Its Applications With R Examples", © Springer Science+Business Media, LLC 2011 .
- [24] Shiller R.J et Perron P(1985), "Testing the Random Walk Hypothesis :Power Versus Frequency of Observation", Economie Letters, 18, PP.381-386 .
- [25] Viano. M.-C et Philippé. A, "Cours de Séries Temporelles", Polycopie de cours, années 1999 à 2004, U.F.R. de Mathématiques Pures et Appliquées