

CHAPITRE I

Modèles Économétriques Dynamiques à une équation

Michel LUBRANO

Janvier 2008

Contents

1	Introduction	3
2	Rappels	3
2.1	Stationnarité	3
2.2	Fonction d'auto-covariance	4
2.3	Opérateur retard	4
3	Deux modèles économétriques dynamiques traditionnels	5
3.1	Modèles à retards échelonnés	5
3.2	Anticipations adaptatives	6
3.3	Modèles auto-régressifs	7
3.4	Ajustement partiel à une cible	8
4	Analyse des propriétés dynamiques	8
4.1	Calcul des coefficients de retard individuels	9
4.2	Multiplicateurs	10
4.3	Retard moyen	10
5	Inférence dans les modèles dynamiques	11
5.1	Variables instrumentales	12
5.2	La méthode du maximum de vraisemblance	13
5.3	Première estimation par MLE	15
5.4	Estimation du modèle général	16
5.5	Polynômes d'Almon	16

<i>CONTENTS</i>	2
6 Application: la relation rendement risque en finance	17
7 Modèles ARE et modèles à corrections d'erreurs	19
7.1 Généralité du modèle	20
7.2 Modèles à correction d'erreur et solution de long terme	22
8 Stratégies de modélisation	23
8.1 Fondements théoriques	23
8.2 La méthode du général au particulier	25
9 Tests de mauvaise spécification	27
9.1 La théorie usuelle des tests d'hypothèses	27
9.1.1 Le test du rapport de vraisemblance	28
9.1.2 Le test de Wald	28
9.1.3 Le test de Lagrange	29
9.1.4 Relation entre les trois tests	29
9.2 Tests de mauvaise spécification	30
9.3 Tests selon une direction de régression	31
9.3.1 Auto-corrélation	31
9.3.2 Le test du RESET ou test de Ramsey	32
9.3.3 Test sur la forme fonctionnelle	33
9.3.4 Test de stabilité	33
9.4 Postsample prediction failure	34
9.4.1 Test d'exogénéité	35
9.5 Augmentation de la variance conditionnelle	35
9.5.1 Variances inégales	36
9.5.2 Arch	36
9.6 Augmentation des moments d'ordre supérieur	36
10 La demande de monnaie en Belgique	37
11 Conclusion	39
12 Lectures additionnelles	40
13 Exercices	40
13.1 Davidson, Hendry, Srba, Yeo (1978): Examen 2006	40
13.2 Exercice 2: Examen 2007	41
13.3 Exercice 3	41
13.4 Exercice 4	42

1 Introduction

Dans ce chapitre, on va s'intéresser à la modélisation des séries temporelles. Une série temporelle est une variable aléatoire dont les observations sont indexées par le temps. L'ordre des observations a donc une importance cruciale. En économie beaucoup de séries statistiques sont des séries temporelles, les séries macro-économiques par exemple ou les séries financières en général. Mais il existe d'autres types de données comme par exemple les données micro-économiques qui concernent l'observation du comportement des individus ou des ménages en un point du temps. L'unité d'observation est alors l'individu ou le ménage et non plus un point du temps.

On va s'attacher dans ce chapitre à étudier des modèles de régression à une équation qui vont expliquer le comportement dynamique d'une variable en fonction d'autres variables. On étudiera en particulier la relation risque-rendement qui est essentielle en finance comme illustration.

2 Rappels

Il est utile de commencer par rappeler quelques notions de base sur les séries temporelles.

2.1 Stationnarité

On supposera pratiquement tout le temps que les séries sont stationnaires, c'est à dire

- $E(x_t) = \mu < \infty \forall t$
- $\text{Var}(x_t) = \sigma^2 < \infty \forall t$
- $\text{Cov}(x_t, x_{t-k}) = \gamma_k \forall t$

Ces moments sont finis et indépendants de du temps. Cela définit la stationnarité au second ordre. Une définition plus stricte de la stationnarité requiert que la distribution entière soit invariante par translation. Dans le cas de la normalité, les deux définitions sont confondues car la normale est parfaitement caractérisée par ses deux premiers moments.

Le modèle suivant décrit un processus qui est toujours stationnaire

$$x_t = \epsilon_t + \beta\epsilon_{t-1}.$$

En effet la moyenne est toujours nulle et la variance est constante et égale à $\sigma^2(1 + \beta^2)$ quelle que soit la valeur des paramètres. Il s'agit d'un modèle moyenne mobile ou MA(1).

Il est par contre facile d'avoir un modèle non stationnaire en considérant par exemple

$$x_t = \mu + \delta t + \epsilon_t.$$

C'est un modèle avec tendance déterministe. La moyenne est égale à $\mu + \delta t$. Elle n'est pas constante dans le temps. La variance par contre est égale à σ^2 .

2.2 Fonction d'auto-covariance

La fonction d'auto-covariance va servir à décrire la mémoire du processus. On la définit comme

$$\gamma_k = E(y_t, t_{t-k}).$$

. Il est utile de la normaliser par σ^2 et d'en donner le graphique. Elle caractérise bien certains processus. Une transformation de cette fonction s'appelle la densité spectrale. Permet d'étudier les cycles dans le domaine des fréquences.

2.3 Opérateur retard

Comme les observations sont indexées par le temps, il est commode de définir un opérateur qui permette de faire des opérations sur cet index temporel. C'est l'opérateur retard que l'on note L pour lag en économétrie. Les statisticiens préfèrent en général utiliser B pour backshift. Cet opérateur a les propriétés suivantes

$$L^r y_t = y_{t-r}.$$

Une opération usuelle dans les séries temporelles consiste à prendre la différence première:

$$\Delta y_t = y_t - y_{t-1}.$$

On peut insérer cet opérateur dans des polynômes. On obtiendra alors des polynômes de retard comme par exemple

$$A(L) = 1 - \alpha L$$

qui va permettre d'écrire un modèle auto-régressif:

$$(1 - \alpha L)y_t = \epsilon_t.$$

L'inverse de ce polynôme se calcule facilement:

$$y_t = \frac{\epsilon_t}{A(L)} = \sum_{j=0}^{\infty} \alpha^j \epsilon_{t-j},$$

en utilisant la somme d'une progression géométrique. Cette écriture permet de voir deux choses. Premièrement, l'inversion nécessite que $|\alpha| < 1$. Deuxièmement cette condition est une condition de stationnarité pour le modèle auto-régressif. En effet, il faut utiliser cette forme pour calculer les moments et ceux-ci n'existent donc que si $|\alpha| < 1$.

Un dernier petit modèle va permettre d'illustrer encore une fois ces notions. C'est le modèle suivant

$$\Delta y_t = \epsilon_t.$$

Son espérance est nulle et sa variance est égale à σ^2 . On peut retrouver la valeur de y_t et ses moments en considérant

$$y_t = \Delta y_t + \Delta y_{t-1} + \dots + \Delta y_1 = \sum_{i=1}^t \Delta y_i$$

en supposant $y_0 = 0$. Comme les Δy_i sont indépendants entre eux et de même moments, on voit que l'espérance de y_t est nulle, mais que sa variance est égale à $t\sigma^2$. Le modèle en y_t est non stationnaire. Il s'agit d'une marche aléatoire.

3 Deux modèles économétriques dynamiques traditionnels

Nous allons maintenant examiner comment on a traditionnellement introduit la dynamique en économétrie dans le cadre d'un modèle conditionnel à une seule équation. Il s'agit de deux modèles très simples qui décrivent des comportements particuliers qu'il faut connaître et qui sont le *modèle d'ajustement partiel* et le *modèle d'anticipation adaptative*. Mais avant cela il faut définir deux types de modèles économétriques standards (modèles à retards échelonnés et modèle auto-régressif) qui en fait font un peu le pendant aux modèles AR et MA des séries temporelles. On consultera avec profit l'ouvrage de Harvey (1981), chapitre 7, mais aussi l'ouvrage plus récent de Hendry (1995).

3.1 Modèles à retards échelonnés

Un modèle à retards échelonnés se note:

$$y_t = \mu + \sum_{i=1}^s \beta_i x_{t-i} + u_t = \mu + B(L) x_t + u_t \quad (1)$$

Les coefficients β_i sont les coefficients de retards. Ils déterminent la façon dont y_t va répondre à un changement dans x_t . Dès lors que l'on suppose que les u_t sont des bruits blancs Gaussiens, il n'a pas de problème statistique particulier pour estimer les coefficients de ce modèle car les hypothèses usuelles des moindres carrés sont satisfaites et en particulier l'indépendance entre les régresseurs et le terme d'erreur. Toutefois une série temporelle évolue lentement à cause des effets mémoire si bien que les différents retards de la variable x_t auront tendance à être corrélés entre eux. On va donc se heurter à un problème de multi-colinéarité qui va gêner la précision dans l'estimation des coefficients de régression. On va imposer une structure particulière à la forme des coefficients de retard pour diminuer le nombre de paramètres à estimer. On résout le problème de multi-colinéarité en introduisant une information supplémentaire. Plusieurs structures sont possibles:

- Almon (1965): les coefficients sont contraints par un polynôme de degré n inférieur au nombre de retards, en général 2 ou 3. On aura

$$\beta_i = \sum_{j=1}^n \gamma_j i^j.$$

- retards échelonnés rationnels: la structure des retards est déterminée par le ratio de deux polynômes de retards:

$$y_t = \mu + \frac{B(L)}{A(L)} x_t + u_t$$

On consultera avec profit l'article de Griliches (1967). Ce type de modélisation est extrêmement souple dans la mesure où le cas le plus simple avec $B(L) = \beta_0 + \beta_1 L$ et $A(L) = 1 - \alpha L$ permet déjà une grande variété de configuration pour la structure des retards.

- retards géométriques ou de Koyck. C'est un cas particulier du précédent où l'on a posé $B(L) = 1$ et $A(L) = 1 - \alpha L$. Les coefficients de retards décroissent de manière exponentielle avec la longueur du retard.

Examinons en détail ce dernier cas. Dans le modèle initial à retards échelonnés, on impose une structure particulière sur les coefficients avec

$$\beta_i = \delta \alpha^i \text{ avec } |\alpha| < 1 \quad (2)$$

Les valeurs des β_i décroissent très vite avec le temps. Aussi il n'est pas très restrictif de supposer un nombre de retards infini. Cela peut même être très commode pour les calculs. Écrivons le modèle:

$$\begin{aligned} y_t &= \mu + \sum_{i=1}^{\infty} \beta_i x_{t-i} + u_t \\ &= \mu + \delta \sum_{i=1}^{\infty} \alpha^i x_{t-i} + u_t \\ &= \mu + \delta \sum_{i=1}^{\infty} (\alpha L)^i x_{t-i} + u_t \end{aligned}$$

Il peut alors se mettre sous la forme:

$$y_t = \mu + \frac{\delta}{1 - \alpha L} x_t + u_t \quad (3)$$

en utilisant la définition de la somme d'une progression géométrique. En multipliant par $(1 - \alpha L)$ les deux membres cette expression, on tombe sur un modèle auto-régressif:

$$y_t = (\mu - \alpha \mu) + \delta x_t + \alpha y_{t-1} + u_t - \alpha u_{t-1} \quad (4)$$

Ce modèle semble plus simple que le précédent et semble se prêter à une estimation commode, mais il comporte un terme d'erreurs auto-corrélées. L'estimation par moindres carrés est impossible à cause de la présence d'une variable endogène retardée, ce qui rend les variables dépendantes corrélées avec le terme d'erreur. Il faut avoir recours à une procédure par maximum de vraisemblance ou de variables instrumentales.

3.2 Anticipations adaptatives

Le modèle d'*anticipations adaptatives* a été introduit par Nerlove (1958) pour des exemples agricoles et par Cagan (1956) pour des problèmes monétaires. On suppose que la décision de production y_t dépend du prix futur x_{t+1} . Au moment de prendre sa décision de planter l'agriculteur doit anticiper le prix futur. On aura donc:

$$y_t = \beta x_{t+1}^* + u_t \quad (5)$$

Comment anticiper le prix? Un schéma d'anticipations adaptatives relie la révision dans les anticipations aux erreurs commises par le passé selon la formule:

$$x_{t+1}^* - x_t^* = \gamma(x_t - x_t^*) \quad (6)$$

avec $0 < \gamma < 1$ qui mesure la vitesse de réaction. Résolvons cette équation pour trouver l'anticipation en fonction des prix passés:

$$(1 - (1 - \gamma)L)x_{t+1}^* = \gamma x_t \quad (7)$$

et en posant $\alpha = 1 - \gamma$

$$x_{t+1}^* = \frac{\gamma}{1 - \alpha L} x_t \quad (8)$$

ce qui montre que l'anticipation est formée comme un retard échelonné infini sur les prix passés. On va maintenant substituer le modèle d'anticipation dans l'équation qui définit la production:

$$y_t = \beta \frac{\gamma}{1 - \alpha L} x_t + u_t \quad (9)$$

que l'on peut résoudre en multipliant tout par $(1 - \alpha L)$:

$$(1 - \alpha L)y_t = \beta(1 - \alpha)x_t + (1 - \alpha L)u_t \quad (10)$$

qui est un modèle auto-régressif à erreurs auto-corrélées. Il existe maintenant d'autres modèles d'anticipation et en particulier les anticipations rationnelles. Avec les anticipations adaptatives, la variable anticipée est une simple extrapolation de son passé, sans tenir compte aucunement des autres variables. Les anticipations rationnelles de Muth (1961).

3.3 Modèles auto-régressifs

On peut décider alternativement de considérer dès le départ une forme particulière de dynamique en incluant l'endogène retardée dans un modèle statique. On a donc simplement:

$$y_t = \mu + \alpha y_{t-1} + \beta x_t + u_t \quad (11)$$

où u_t est cette fois-ci un bruit blanc. Ce modèle est donc très simple. Une condition de stationnarité est bien sûr que $|\alpha| < 1$. On peut exprimer ce modèle en utilisant l'opérateur retard:

$$(1 - \alpha L)y_t = \mu + \beta x_t + u_t \quad (12)$$

et en divisant les deux membres par $(1 - \alpha L)$ on arrive à un modèle à retards échelonnés infinis mais avec cette fois-ci une dynamique sur les termes d'erreur à la différence du modèle simple à retards échelonnés. On a:

$$y_t = \frac{\mu}{1 - \alpha} + \frac{\beta}{1 - \alpha L} x_t + \frac{1}{1 - \alpha L} u_t \quad (13)$$

ou bien en terme de sommes infinies:

$$y_t = \frac{\mu}{1 - \alpha} + \beta \sum_{i=0}^{\infty} \alpha^i x_{t-i} + \sum_{i=0}^{\infty} \alpha^i u_{t-i} \quad (14)$$

Un modèle auto-régressif peut être généralisé en un modèle auto-régressif à retard échelonnés (ARE) que l'on détaillera par la suite:

$$A(L) y_t = \mu + B(L) x_t + u_t \quad (15)$$

Le modèle ARE et le modèle à fonction de transfert ont la même dynamique systématique, mais une dynamique différente sur les résidus.

3.4 Ajustement partiel à une cible

Le modèle *d'ajustement partiel* a une très longue histoire en économétrie et on peut le faire remonter également à Nerlove (1958). On considère une variable économique y_t à laquelle on fixe un objectif y_t^* à atteindre. Ce peut être par exemple un niveau désiré d'investissement ou d'emploi pour une entreprise. Ce niveau désiré est déterminé par un modèle auxiliaire en fonction de variables exogènes:

$$y_t^* = x_t' \beta \quad (16)$$

Pour toutes une séries de raisons il est en général coûteux d'ajuster trop rapidement une variable à sa cible. On pose alors:

$$\Delta y_t = \gamma(y_t^* - y_{t-1}) + u_t \quad (17)$$

avec $0 < \gamma \leq 1$. Cette relation décrit un ajustement dont la vitesse dépend de γ . Pour $\gamma = 1$ l'ajustement est instantané. En remplaçant y_t^* par sa valeur dans la relation d'ajustement, on obtient:

$$y_t = (1 - \gamma)y_{t-1} + x_t' \beta \gamma + u_t \quad (18)$$

qui est exactement un modèle auto-régressif avec résidus bruits blancs. Par rapport au modèle ARE, on remarque l'absence de retard sur x_t .

Dans les domaines économétriques de demande de facteur (demande de travail par exemple ou bien fonction d'investissement), il est essentiel de modéliser les coût d'ajustement. Ce type de modélisation de la dynamique a été utilisé de manière extensive lors de la construction du modèle européen COMET par Barten, d'Alcantara, and Carrin (1976). Il faut toutefois noter que ce mécanisme d'ajustement est assez restrictif. On n'arrive jamais complètement à atteindre la cible. Alors qu'avec le modèle à correction d'erreurs que l'on détaillera plus loin, on arrive bien mieux à corriger et à atteindre la cible.

4 Analyse des propriétés dynamiques

Nous allons ici analyser la caractérisation de la dynamique de la partie systématique d'un modèle à retards rationnels du type

$$y_t = \frac{B(L)}{A(L)} x_t + u_t$$

On s'intéresse au cheminement moyen de y_t c'est à dire à l'espérance de y_t conditionnellement aux valeurs présentes et passées des exogènes. Comme les erreurs sont de moyenne nulle, on arrive à

$$E(y_t) = A^{-1}(L) B(L) x_t \quad (19)$$

On voit que cette espérance est une fonction des coefficients de retards d'un modèle à retards échelonnés infinis. Il s'agit donc de déterminer cette suite infinie de coefficients et de tenter de la résumer au moyen de diverses caractéristiques. Pour cela il faut bien sûr que cette suite soit convergente, ce qui implique entre autres que les racines du polynôme $A(L)$

soient toutes situées à l'extérieur du cercle unité, c'est à dire que le polynôme $A(L)$ soit inversible. Les racines de ce polynôme correspondent à l'équation caractéristique $A(z) = 0$, c'est à dire

$$A(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_r z^r = 0.$$

Une solution d'équilibre pour le modèle est alors possible étant donné que si x_t est constant, y_t sera aussi constant et l'on aura:

$$\bar{y} = \frac{\beta_0 + \beta_1 + \dots + \beta_s}{1 - \alpha_1 - \alpha_2 \dots - \alpha_r} \bar{x} = \frac{B(1)}{A(1)} \bar{x} \quad (20)$$

Les valeurs de $A(1)$ et $B(1)$ se trouvent en remplaçant l'opérateur L dans le polynôme par z et en posant $z = 1$. L'expression trouvée correspond à la somme des coefficients du polynôme.

4.1 Calcul des coefficients de retard individuels

L'effet total est donc obtenu comme le ratio entre la somme des coefficients de $B(L)$ et la somme des coefficients de $A(L)$, c'est à dire sans calculer la suite des coefficients de retards. Mais nous pouvons parfois avoir besoin de celle-ci. Définissons:

$$D(L) = \frac{B(L)}{A(L)} = \delta_0 + \delta_1 L + \delta_2 L^2 + \dots \quad (21)$$

On peut calculer cette suite simplement par une opération de division de polynômes classique. On peut aussi procéder par identification au moyen de la formule:

$$B(L) = A(L) D(L) \quad (22)$$

Cette opération s'effectue de manière récursive en identifiant les termes des deux cotés. On va commencer par traiter un exemple où les deux polynômes sont de degré 1 avant de donner la formule générale. On part de

$$D(L) = \frac{\beta_0 + \beta_1 L}{1 - \alpha L} = \delta_0 + \delta_1 L + \delta_2 L^2 + \dots$$

d'où l'on tire

$$\begin{aligned} \beta_0 + \beta_1 L &= (1 - \alpha L)(\delta_0 + \delta_1 L + \delta_2 L^2 + \dots) \\ &= \delta_0 + (\delta_1 - \alpha \delta_0)L + (\delta_2 - \alpha \delta_1)L^2 + \dots \end{aligned}$$

L'identification des puissances de L des deux membres va donner les équations suivantes

$$\begin{aligned} \delta_0 &= \beta_0 \\ \delta_1 - \alpha \delta_0 &= \beta_1 \\ \delta_2 - \alpha \delta_1 &= 0 \\ &\dots \end{aligned}$$

d'où l'on tire la condition initiale $\delta_0 = \beta_0$ et la solution de la récurrence devient

$$\begin{aligned}\delta_1 &= \beta_1 - \alpha\beta_0 \\ \delta_2 &= \alpha(\beta_1 - \alpha\beta_0) \\ \delta_3 &= \alpha^2(\beta_1 - \alpha\beta_0) \\ &\dots\end{aligned}$$

On en déduit la relation: $\delta_j = \alpha^{j-1}(\beta_1 - \alpha\beta_0)$.

Cette suite se généralise facilement pour $A(L)$ et $B(L)$ quelconques. Comme dans tous les cas on aura $\alpha_0 = 1$, il en résulte que l'on aura toujours la même condition initiale $\delta_0 = \beta_0$. Il vient ensuite la récurrence suivante:

$$\delta_j = \sum_{i=1}^{\min(j,r)} \alpha_i \delta_{j-i} + \beta_j \quad \text{si } 1 \leq j \leq s \quad (23)$$

$$\delta_j = \sum_{i=1}^{\min(j,r)} \alpha_i \delta_{j-i} \quad \text{si } j > s \quad (24)$$

4.2 Multiplicateurs

La notion de multiplicateur est définie dans le contexte d'un équilibre de long terme. On a

$$\bar{y} = \frac{B(1)}{A(1)} \bar{x}.$$

On envisage maintenant une perturbation unitaire de la valeur d'équilibre \bar{x} , ce qui fait que l'on passe de \bar{x} à $\bar{x} + 1$. Quel est la nouvelle valeur d'équilibre de y ? On a:

$$\bar{y}^* = \frac{B(1)}{A(1)} (\bar{x} + 1) = \bar{y} + \frac{B(1)}{A(1)}. \quad (25)$$

Donc $B(1)/A(1)$ est le multiplicateur total étant donné qu'il représente la somme de tous les coefficients δ_j . *Le multiplicateur d'impact* est le premier élément de la suite des δ , c'est à dire δ_0 . *Le multiplicateur d'intérim* à l'ordre J est défini par la somme des J premiers termes de la suite des δ_j :

$$\delta_J^* = \sum_{j=0}^J \delta_j \quad (26)$$

Il est commode de normaliser les multiplicateurs d'intérim par rapport au multiplicateur total de manière à pouvoir calculer le pourcentage de l'effet après J périodes, sachant qu'à l'infini l'effet sera total.

4.3 Retard moyen

Normalisons maintenant la suite des δ_j par leur somme. Si tous les δ_j sont positifs, alors la somme des δ_j normalisés sera égale à un. Comme il s'agit de nombres positifs et normalisés à un, on peut alors les interpréter comme des probabilités définies sur la suite

entière des retards $(0, 1, 2, \dots, \infty)$ et calculer certaines caractéristiques comme la moyenne et la médiane. Le retard moyen est alors défini par:

$$\mu_\delta = \frac{\sum_{j=0}^{\infty} j \delta_j}{\sum_{j=0}^{\infty} \delta_j} \quad (27)$$

Il faut prendre cette formule comme une définition, mais non comme une procédure de calcul, car elle nécessite le calcul de tous les δ_j . Pour calculer le retard moyen, on lui préfère l'approche initiée par Griliches (1967). Considérons formellement $D(z)$ avec $D(1) = 1$ et tous les δ_i positifs. Si $D(z)$ converge pour $-z_0 < z < z_0$, alors $D(z)$ est appelée une fonction génératrice de probabilités. Entre autres propriétés, on a que l'espérance de la variable qui est distribuée selon la distribution générée par $D(z)$ est égale à la dérivée première de $D(\cdot)$ calculée au point $z = 1$.¹ Appliquons ces formules à notre cas. Le retard moyen que met un choc sur x_t à se transmettre à y_t se calcule à partir de

$$\mu_\delta = \left. \frac{D'(L)}{D(L)} \right|_{L=1}.$$

Comme:

$$D'(L) = \frac{A(L) B'(L) - A'(L) B(L)}{A^2(L)}, \quad (28)$$

on arrive à:

$$\mu_\delta = \frac{D'(1)}{D(1)} = \frac{B'(1)}{B(1)} - \frac{A'(1)}{A(1)}. \quad (29)$$

Dans le cas du modèle d'ajustement partiel:

$$y_t = \alpha y_{t-1} + \beta x_t + \epsilon_t \quad (30)$$

on vérifie facilement à partir de (29) que le retard moyen est égal à $\alpha/(1 - \alpha)$. On appelle parfois α le coefficient d'inertie. Quand α est égal à zéro, il n'y a aucune inertie dans le modèle et l'ajustement se fait de manière instantanée. Pour $\alpha = 1$, le retard moyen devient infini.

5 Inférence dans les modèles dynamiques

L'inférence dans le modèle auto-régressif

$$y_t = \alpha y_{t-1} + u_t$$

ne pose pas de problème particulier. On peut appliquer les moindres carrés car la variable prédéterminée est indépendante des erreurs. En effet

$$\begin{aligned} \text{plim } \frac{1}{T} \sum y_{t-1} u_t &= \text{plim } \frac{1}{T} \sum (\alpha y_{t-2} + u_{t-1}) u_t \\ &= \text{plim } \frac{1}{T} \sum \alpha y_{t-2} u_t + \text{plim } \frac{1}{T} \sum u_{t-1} u_t \\ &= 0. \end{aligned}$$

¹La variance fait intervenir la dérivée seconde de $D(\cdot)$, notée $D''(\cdot)$. Son expression est donnée par $D''(1) + D'(1) - [D'(1)]^2$.

Si l'on ajoute des variables exogènes, même retardées, le problème ne change pas. Considérons par exemple le modèle

$$y_t = \alpha y_{t-1} + \beta_1 x_t + \beta_2 x_{t-2} + u_t.$$

Pour estimer ce modèle, il suffit de construire deux matrices d'observations

$$y = [y_t] \quad X = [y_{t-1}, x_t, x_{t-1}]$$

Alors l'estimateur des moindres carrés sera formé par $\hat{\gamma} = (X'X)^{-1}X'y$ où γ représente l'ensemble des paramètres de régression à estimer.

Par contre, le modèle à retards échelonnés va lui poser des problèmes d'inférence. La façon la plus simple de s'en apercevoir consiste à considérer le cas du modèle à retards géométriques

$$y_t = \beta \sum_{j=0}^{\infty} \alpha^j x_{t-j} + u_t.$$

On a vu que l'on pouvait résoudre ce modèle en utilisant un polynôme de retard en un modèle autorégressif, mais avec erreurs autocorrélées:

$$y_t = \alpha y_{t-1} + \beta x_t + v_t \quad v_t = u_t - \alpha u_{t-1}.$$

L'estimateur des OLS ne sera pas consistant car la variable endogène retardée est cette fois-ci corrélée avec le terme d'erreur. En effet:

$$\begin{aligned} \text{plim} \frac{1}{T} \sum y_{t-1} v_t &= \text{plim} \frac{1}{T} \sum y_{t-1} (u_t - \alpha u_{t-1}) \\ &= 0 - \alpha \text{plim} \frac{1}{T} \sum y_{t-1} u_{t-1} \\ &= -\alpha \text{plim} \frac{1}{T} \sum (y_{t-2} + \beta x_{t-1} + v_{t-1}) u_{t-1} \\ &= -\alpha \sigma^2, \end{aligned}$$

ce qui n'est plus égal à zéro, sauf si $\alpha = 0$, c'est à dire le cas statique. On va donc devoir utiliser une méthode d'estimation différente. La méthode la plus simple, mais la moins précise, consiste à utiliser des variables instrumentales. L'autre approche, plus précise, passe par le maximum de vraisemblance.

5.1 Variables instrumentales

On peut construire un estimateur par variable instrumentales en utilisant x_{t-1} comme instrument pour y_{t-1} . Ce n'est pas très efficace mais sert pour amorcer les itérations d'un MLE. On rappelle que quand on définit les moindres carrés, on suppose que les erreurs sont indépendantes des exogènes. C'est une condition d'identification qui dans le modèle

$$y_t = \alpha y_{t-1} + \beta x_t + v_t$$

s'exprime comme

$$\begin{aligned} E[(y_t - \alpha y_{t-1} - \beta x_t)x_t] &= 0 \\ E[(y_t - \alpha y_{t-1} - \beta x_t)y_{t-1}] &= 0. \end{aligned}$$

En résolvant ces deux équations où l'on a remplacé l'espérance par la moyenne empirique, on trouve l'estimateur des moindres carrés. On a deux équations pour deux inconnues. Si une de ces deux relations n'est plus valide, on ne peut plus calculer l'estimateur des moindres carrés. C'est le cas ici dans le modèle auto-régressif. Pour résoudre cette difficulté, on va remplacer la variable fautive par un instrument qui par conjecture sera corrélé avec y_{t-1} , mais non corrélé avec le terme d'erreur. Si x_{t-1} est choisi comme instrument, on va remplacer la deuxième condition par

$$E[(y_t - \alpha y_{t-1} - \beta x_t)x_{t-1}] = 0.$$

On va pouvoir définir l'estimateur par variable instrumentale de la manière suivante. Construisons tout d'abord les matrices d'observations:

$$Z = [x_{t-1}, x_t], \quad X = [y_{t-1}, x_t], \quad y = [y_t],$$

comportant $T - 1$ lignes. Si l'on définit maintenant le paramètre $\gamma' = [\alpha, \beta]$, alors l'estimateur par variable instrumentale est égal à

$$\hat{\gamma}_{IV} = (Z'X)^{-1}Z'y.$$

Cet estimateur représente la version la plus simple de l'estimateur à variable instrumentale, car l'on a remplacé la variable fautive par un seul instrument. Dans le cas général on aura plus d'une variable instrumentale. On ne pourra résoudre aussi facilement les deux équations car la matrice $Z'X$ ne sera plus carrée et inversible. On aura alors un problème de sur-identification et l'on fera appel à la méthode des variables instrumentales généralisées. Voir par exemple le chapitre 5 de Verbeek (2000) pour plus de détails.

5.2 La méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance consiste à partir de la densité d'une observation, à en déduire la densité d'un échantillon complet et ensuite à trouver la valeur du paramètre qui rend maximale la probabilité d'avoir effectivement obtenu cet échantillon. Dans le cas IID, si la densité d'une observation est $p(y_i; \theta)$ où θ est le paramètre qui indexe cette densité, la fonction de vraisemblance se forme en considérant le produit

$$L(y_1, y_2, \dots, y_T) = \prod_{i=1}^T p(y_i; \theta).$$

L'estimateur du maximum de vraisemblance se trouve en résolvant les conditions du premier ordre. Pour des raisons de commodité de calcul, on préfère considérer le logarithme de la vraisemblance et donc:

$$\frac{\partial \log L(y; \theta)}{\partial \theta} = 0.$$

La matrice d'information se définit comme moins l'espérance de la dérivée seconde de la log vraisemblance

$$I(\theta) = -\mathbf{E} \frac{\partial^2 \log L(y; \theta)}{\partial \theta \partial \theta'}.$$

On ne peut facilement calculer cette espérance que dans des cas simples. Dans le cas général, on va se contenter d'une approximation asymptotique

$$IA(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} I(\theta).$$

Sous certaines conditions de régularité, l'estimateur du maximum de vraisemblance sera asymptotiquement normal avec

$$\sqrt{T}(\hat{\theta} - \theta) \sim \mathbf{N}(0, IA^{-1}(\theta)),$$

ce qui signifie que pour calculer la variance de $\hat{\theta}$ on utilisera $IA^{-1}(\theta)/T$.

Quand les observations sont liées dans le temps, on ne peut plus prendre un simple produit pour trouver la densité de l'échantillon complet. C'est à dire qu'il n'est plus possible d'écrire $f(y_1, y_2) = f(y_1) \cdot f(y_2)$, mais par contre on peut toujours utiliser la décomposition $f(y_1, y_2) = f(y_1|y_2) \cdot f(y_2)$. On va donc maintenant devoir écrire:

$$\log L(y; \theta) = \sum_{t=2}^T \log l(y_t|y_{t-1}, \dots, y_1) + \log l(y_1).$$

On aura donc une suite de densités conditionnelles, suivie par la densité marginale de la première observation. Harvey (1981) montre comment on peut remplacer la première partie de cette écriture par la densité des erreurs de prévision, c'est à dire la densité des résidus récurrents. Traitons par exemple le modèle autorégressif $y_t = \alpha y_{t-1} + \epsilon_t$. La densité de la première observation est

$$f(y_1) = f_N \left(y_1 | 0, \frac{\sigma^2}{1 - \alpha^2} \right),$$

et la densité des résidus récurrents e_t

$$f(e_t) = f_N(e_t | y_t - \alpha y_{t-1}, \sigma^2).$$

La log vraisemblance sera donc

$$\begin{aligned} \log L(y) &\propto \frac{1}{2} \log(1 - \alpha^2) - \frac{1 - \alpha^2}{2\sigma^2} y_1^2 \\ &\quad - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^T (y_t - \alpha y_{t-1})^2. \end{aligned}$$

On voit tout de suite que si l'on néglige la distribution de la première observation, on retombe sur la solution des moindres carrés, c'est à dire la minimisation d'une somme de carrés d'erreurs.

5.3 Première estimation par MLE

Considérons en premier le modèle à retards échelonnés de Koyck:

$$y_t = \beta \sum_{j=0}^{\infty} \alpha^j x_{t-j} + \epsilon_t$$

Les erreurs ϵ_t étant IID, l'estimation par MLE peut se faire en minimisant une somme de carrés des erreurs, comme si l'on était dans un modèle de régression linéaire à erreurs normales usuelles. Cependant, l'écriture directe est de peu d'utilité car elle fait intervenir une somme infinie de retards. On va décomposer donc la somme infinie en deux parties

$$y_t = \beta \sum_{j=0}^{t-1} \alpha^j x_{t-j} + \alpha^t [\beta \sum_{j=0}^{\infty} \alpha^j x_{0-j}] + \epsilon_t$$

Ce qui peut s'écrire encore

$$y_t = \beta x_t^*(\alpha) + \alpha^t \mathbf{E}(y_0) + \epsilon_t$$

où $x_t^*(\alpha) = \sum_{j=0}^{t-1} \alpha^j x_{t-j}$ doit être construit de manière récursive en partant de $x_1^*(\alpha) = x_1$:

$$x_t^*(\alpha) = x_t + \alpha x_{t-1}^*(\alpha).$$

Il faut régler le problème des conditions initiales, c'est à dire donner une valeur à $\mathbf{E}(y_0)$. Trois solutions sont possibles.

- On peut les fixer égales à zero. On minimise alors simplement

$$S(\alpha, \beta) = \frac{1}{T} \sum (y_t - \beta x_t^*(\alpha))^2$$

Il y a bien sûr un léger biais dans l'estimation. Celui-ci s'atténue en grand échantillon ou pour des valeurs faibles de α car les conditions initiales rentrent au moyen de α^t .

- On peut estimer les conditions initiales en les traitant comme un paramètre supplémentaire. Le problème d'optimisation comporte alors un paramètre de plus, ce qui rend la recherche de l'optimum plus problématique.
- Enfin, on peut décider de fixer les conditions initiales égales à $\mathbf{E}(y_1)$ en modifiant les indices de la sommation. Le modèle devient

$$y_t = \beta \sum_{j=0}^{t-2} \alpha^j x_{t-j} + \alpha^{t-1} \mathbf{E}(y_1) + \epsilon_t$$

On peut maintenant poser sans trop d'erreur $\mathbf{E}(y_1) = y_1$. L'inconvénient, c'est que l'on a utilisé une observation de moins dans la vraisemblance.

Dans tous les cas, il faut veiller à faire l'optimisation sous la contrainte que $|\alpha| < 1$.

5.4 Estimation du modèle général

Le modèle précédent est relativement restrictif car il impose une structure très particulière des retards. Le modèle général se

$$y_t = \frac{B(L)}{A(L)}x_t + \epsilon_t,$$

alors que le modèle de Koyck correspond à $B(L) = 1$ et $A(L) = 1 - \alpha L$. Pour estimer ce modèle général, il serait relativement difficile d'opérer comme précédemment. On préfère circonscrire la question des retards infinis en passant à la forme autorégressive au moyen de

$$A(L)y_t = B(L)x_t + A(L)\epsilon_t.$$

La procédure par maximum de vraisemblance consiste tout d'abord à construire la suite des résidus récurrents. Prenons l'exemple simple où les polynômes $B(L)$ et $A(L)$ sont tous les deux de degré 1. On aura la formule de récursion générale:

$$\hat{\epsilon}_t = y_t + \alpha y_{t-1} - \beta_0 x_t - \beta_1 x_{t-1} + \alpha \epsilon_{t-1}.$$

Il faudra amorcer cette récursion en posant par exemple $\epsilon_1 = 0$ et en la démarrant en $t = 2$ pour résoudre le problème des valeurs de départ pour x et y . On peut alors facilement évaluer la récursion suivante en prenant $\epsilon_1 = 0$:

$$\begin{aligned} \hat{\epsilon}_2 &= y_2 + \alpha y_1 - \beta_0 x_2 - \beta_1 x_1 + \alpha \epsilon_1 \\ \hat{\epsilon}_3 &= y_3 + \alpha y_2 - \beta_0 x_3 - \beta_1 x_2 + \alpha \epsilon_2 \\ \dots &= \dots \end{aligned}$$

Si on laisse tomber la question de la distribution de la première observation, la fonction de vraisemblance approchée se réduit à

$$\log L(y_T, \dots, y_2; \theta) \propto -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^T \hat{\epsilon}_t^2.$$

5.5 Polynômes d'Almon

On part maintenant du modèle avec retards en nombre finis égaux à m :

$$y_t = \sum_{j=0}^m \delta_j x_{t-j} + u_t$$

On ne peut estimer ce modèle directement à cause des problèmes de multicollinéarité. On va supposer que les $m+1$ paramètres sont contraints par un polynôme en j de degré $n < m$ (en général $n = 2, 3$), ce qui implique

$$\delta_j = \sum_{i=0}^n \gamma_i j^i = P(j)$$

On va alors remplacer les paramètres δ_j par leur valeur dans le modèle

$$y_t = \sum_{j=0}^m \left(\sum_{i=0}^n \gamma_i j^i \right) x_{t-j} + u_t = \sum_{i=0}^n \omega_{it} \gamma_i + u_t$$

avec

$$\omega_{it} = \sum_{j=0}^m j^i x_{t-j}.$$

On va donc construire un jeu de nouvelles variables exogènes par simple transformation. Si W est la matrice de ces exogènes transformées, le modèle se note

$$y = W\gamma + u$$

et l'on va estimer les $n + 1$ paramètres γ de ce modèle par OLS. On peut ensuite retrouver les paramètres de retard en utilisant le polynôme.

Comment déterminer le degré m du polynôme à utiliser? On va comparer un modèle de régression non-contraint qui comporte n retards et donc $n + 1$ régresseurs à un modèle où les coefficients des n retards sont contraints par un polynôme de degré m . On va utiliser pour cela un test en F . Appelons SSE la somme des carrés des résidus pour le modèle non-contraint et SSE_m la somme des carrés des résidus pour le modèle à polynôme d'Almon de degré m . Pour tester la restriction, il suffit de calculer

$$F = \frac{(SSE_m - SSE)/(m - n)}{SSE/(T - m - 1)}$$

Cette statistique est distribuée selon une loi en $F(m - n, T - m - 1)$.

6 Application: la relation rendement risque en finance

La finance théorique a mis en avant l'arbitrage entre rendement et risque, le paramètre central étant l'aversion au risque. Un investisseur acceptera un risque plus grand dans son choix de portefeuille si celui-ci lui rapporte en espérance un rendement supérieur. Cette relation classique est une relation statique que l'on a du mal à mettre à jour dans les études empiriques. Merton (1973) a proposé un modèle inter-temporel pour le CAPM (Capital Asset Pricing Model) qui propose la relation empirique suivante qui reste à tester:

$$E_t(R_{t+1}) = \mu + \gamma \text{Var}(R_{t+1}).$$

R_{t+1} mesure en $t + 1$ le rendement du portefeuille de marché. En finance on appelle rendement la variation du cours, c'est à dire $S_{t+1} - S_t$ si S est le prix de l'actif financier. L'espérance est calculée en t . Le coefficient γ représente l'aversion au risque. Le cours du portefeuille de marché est donné par un indice. Le problème dans l'équation du haut c'est que la variance (ou volatilité) n'est pas observée et doit être estimée à partir des rendements passés.

On utilise des observations mensuelles pour mesurer l'espérance des rendements. Utiliser une fréquence plus haute (données journalières) conduirait à trop de bruit. On a donc résolu

la question du membre gauche de l'équation. Pour estimer la variance des rendements, on se base en général sur des données journalières de la même variable (22 observations dans le mois). Appelons R_t les observations mensuelles des rendements et r_t les observations journalières de cette même variable et \hat{V}_t un estimateur de la variance future en t . L'estimateur usuel est basé sur

$$\hat{V}_t = \frac{1}{22} \sum_{d=0}^{22} r_{t-d}^2.$$

Cet estimateur est basé sur le principe que les variations journalières fournissent une bonne idée sur la variance du portefeuille de marché. Toutefois quand on l'utilise pour estimer la relation initiale, on trouve des valeurs pour γ qui sont tantôt positives, tantôt négatives et rarement significatives. Ghysel, Santa-Clara, and Valkanov (2005) ont proposé d'utiliser un estimateur basé sur les modèles à retards échelonnés. L'idée est que premièrement la variance peut être persistante et que donc un mois d'observations n'est pas suffisant et que deuxièmement le poids des observations passées doit décroître régulièrement. Ils proposent donc d'utiliser

$$V_t = 22 \sum_{d=0}^{\infty} w_d r_{t-d}^2$$

où les w_d sont des poids sur lesquels on va imposer une structure. On a vu que les retards de Koyck imposaient une structure de décroissance exponentielle au moyen d'un seul paramètre. Ghysel, Santa-Clara, and Valkanov (2005) proposent d'utiliser une structure décroissante à deux paramètres

$$w(d, \kappa) = \frac{\exp(\kappa_1 d + \kappa_2 d^2)}{\sum_{i=0}^{\infty} \exp(\kappa_1 i + \kappa_2 i^2)}$$

Ce schéma garantit des poids positifs à cause de l'exponentielle et dont la somme est égale à 1. D'autre part, les poids seront décroissants à partir d'un certain nombre de retards pour $\kappa_2 < 0$. Ghysel, Santa-Clara, and Valkanov (2005) choisissent de tronquer la suite infinie à 252, c'est à dire à environ une année d'observations journalières. On se retrouve finalement dans un modèle à retards échelonnés que l'on va estimer dans sa forme initiale en supposant que les erreurs sont indépendantes. Les auteurs font une hypothèse de normalité des erreurs, basée sur cette hypothèse d'indépendance, mais supposent que la variance des erreurs varie dans le temps selon le schéma suivant:

$$R_{t+1} \sim N(\mu + \gamma V_t(\kappa_1, \kappa_2), V_t(\kappa_1, \kappa_2))$$

ce qui correspond à un modèle de régression à retards échelonnés polynômiaux et erreurs hétérosckédastiques un peu particulier

$$R_t = \mu + \gamma 22 \sum_{d=0}^{252} w(d, \kappa) r_{t-d} + U_t$$

avec $\text{Var}(U_t) = V_t(\kappa_1, \kappa_2)$. L'indice t représente des fréquences mensuelles pour R et U , alors qu'il représente des fréquences journalières pour r . L'hypothèse de normalité permet d'écrire la fonction de vraisemblance comme:

$$\log L(R_T, \dots, R_{12}; \theta) \propto -\frac{1}{2} \sum_{t=12}^T \log V_t(\kappa_1, \kappa_2) - \sum_{t=12}^T \frac{1}{2V_t(\kappa_1, \kappa_2)} \sum_{t=12}^T \hat{U}_t^2,$$

avec

$$\hat{U}_t = R_t - \mu - \gamma \sum_{d=0}^{252} w(d, \kappa) r_{t-d}.$$

Ghysel, Santa-Clara, and Valkanov (2005) estiment ce modèle en utilisant des données américaines fournies par le Center for Research in Security Prices (CRSP) de Janvier 1928 à Décembre 2000. Toutes les estimations sont significatives comme l'indiquent les t tests

Table 1: Estimation par maximum de vraisemblance du modèle risque/rendement

Sample	μ	γ	$\kappa_1 * 10^3$	$\kappa_2 * 10^5$
1928:01 - 2000:12	6.430	2.606	-5.141	-10.580
	[11.71]	[6.71]	[-4.53]	[-5.24]
1928:01 - 1963:12	11.676	1.547	-0.909	-10.807
	[5.89]	[3.38]	[-3.77]	[-2.11]
1964:01 - 2000:12	3.793	3.748	-6.336	-18.586
	[5.67]	[8.61]	[-7.86]	[-7.71]

Les statistiques de Student sont données entre crochets

donnés entre crochets. L'ordonnée à l'origine μ capture l'effets des facteurs non présents dans l'équation. Le coefficient de risque γ est relativement stable à travers les deux périodes considérées. Les poids décroissent lentement dans le temps. Ils représentent 33% d'effet après un mois, 56% après deux mois et 75% après trois mois. On est donc loin de la pratique courante qui consiste à ne prendre qu'un mois d'observations journalières pour estimer la volatilité. Il y a un phénomène de persistance qui est pris en compte par cette approche MIDAS (mixed data sampling).

7 Modèles ARE et modèles à corrections d'erreurs

On recherche une classe de modèles qui soit à la fois riche sur le plan dynamique e facile à estimer,c'est à dire qu'elle nerequière pas l'utilisation du maximum de vraisemblance. Les modèles à retards échelonnés ne sont pas commodes sur ce plan car ils sont à la fois restrictifs comme le montre l'écriture autorégressive suivante

$$A(L)y_t = \beta x_t + A(L)\epsilon_t,$$

et compliqués à estimer à cause des problèmes d'autocorrélation. Les économètres ont eu tendance à privilégier dans cet esprit une classe de modèles que l'on va maintenant détailler. Il s'agit du modèle ARE, auto-régressif à retards échelonnés. Ce modèle est particulièrement commode car il permet une estimation par simple OLS et une stratégie de

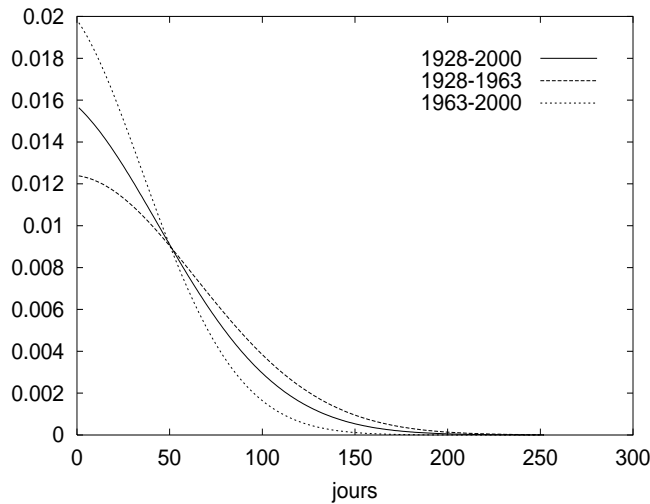


Figure 1: Poids des retards dans l'estimation de la volatilité

recherche de spécification dite du général au particulier dont on reparlera dans la section suivante. Le modèle se note:

$$A(L)y_t = \mu + B(L)x_t + u_t \quad (31)$$

où A et B sont des polynômes de retard de degré r et s respectivement. Ce modèle est général car les polynômes $A(L)$ et $B(L)$ ne comportent aucune restriction, à part celles sur $A(L)$ qui garantissent la stationnarité. La stabilité du modèle implique que les racines du polynôme $A(z)$ soient toutes situées à l'extérieur du cercle unité. Nous allons montrer que ce modèle très simple englobe en fait comme cas particulier toute une série de modèles particuliers. Dans cet exemple, il n'y a qu'une variable exogène pour la commodité de l'exposé. Mais on considérera dans la pratique des modèles du type:

$$A(L)y_t = \mu + B_1(L)x_{t1} + \dots + B_k(L)x_{tk} + u_t \quad (32)$$

donc à k variables exogènes.

7.1 Généralité du modèle

Hendry, Pagan, and Sargan (1984) partent de l'équation simplifiée suivante:

$$y_t = \mu + \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t \quad (33)$$

pour détailler tous les modèles particuliers qui se rattachent au modèle ARE.

- *Modèle de régression statique* pour $\alpha = \beta_1 = 0$, ce qui donne:

$$y_t = \mu + \beta_0 x_t + u_t \quad (34)$$

Celui-ci correspond à la forme qui est usuellement donnée par la théorie économique. Mais ce genre de modèle n'est jamais bien justifié sur le plan statistique et ne peut pas être utilisé

directement, sauf exceptions. Il permet par exemple de mettre en lumière le problème des régressions factices qui illustre le paradoxe statistique expliqué dans Granger and Newbold (1974). Il peut aussi être utilisé dans une procédure d'estimation pour des relations de cointégration comme dans Engle and Granger (1987).

- *Modèle de régression en différence* pour $\alpha = 1$ et $\beta_0 = -\beta_1$.

$$\Delta y_t = \mu + \beta_0 \Delta x_t + u_t. \quad (35)$$

Ce modèle est en général utilisé pour résoudre le problème des régression "spurieuses" en stationnarisant les séries par différence. Mais il interdit toute forme de solution de long terme dans le niveau des variables telle que représentée par le modèle statique.

- *Modèle AR(1)* des séries temporelles sans variable exogène pour $\beta_0 = \beta_1 = 0$:

$$y_t = \mu + \alpha y_{t-1} + u_t. \quad (36)$$

- *Modèle avec erreurs auto-corrélées* à l'ordre 1 pour $\beta_1 = -\alpha\beta_0$:

$$(1 - \alpha L)y_t = \mu + \beta_0(1 - \alpha L)x_t + u_t, \quad (37)$$

ce qui donne:

$$y_t = \frac{\mu}{1 - \alpha} + \beta_0 x_t + \frac{u_t}{(1 - \alpha L)}. \quad (38)$$

Cette restriction correspond à ce que l'on appelle un facteur commun, c'est à dire que la dynamique sur y et la dynamique sur x se ressemblent au point que l'on peut factoriser une racine commune dans les deux polynômes de retard et ainsi la faire apparaître dans le terme d'erreur. Dans ce cas particulier, l'autocorrélation des résidus apparaît comme un moyen commode d'estimer un modèle de façon plus parcimonieuse et n'est donc pas un inconvénient, d'où le titre de l'article de Hendry and Mizon (1978).

- *Modèle d'ajustement partiel* pour $\beta_1 = 0$:

$$y_t = \mu + \alpha y_{t-1} + \beta_0 x_t + u_t \quad (39)$$

On peut montrer que ce type d'ajustement correspond à l'utilisation d'une fonction de perte quadratique sur les coûts d'ajustement.

- *Modèle à correction d'erreur* pour $\alpha + \beta_0 + \beta_1 = 1$.

$$\Delta y_t = \mu + \beta_0 \Delta x_t + (\alpha - 1)(y_{t-1} - x_{t-1}) \quad (40)$$

On détaillera le fonctionnement de ce modèle dans la sous section suivante vu son importance pour la suite. Il faut souligner au passage que la contrainte imposée d'élasticité unitaire n'est en rien obligatoire et que l'on appelle plutôt ECM tout modèle présentant une reparamétrisation combinant différences et niveaux.

D'autres types de modèle sont encore possibles, mais ce que l'on a voulu montrer c'est que par un choix de restrictions qui sont tout à fait testables on peut arriver plusieurs types de spécifications en partant d'un simple modèle ARE. Ce point montre bien comment ce modèle se prête particulièrement bien à une procédure de recherche de spécification.

7.2 Modèles à correction d'erreur et solution de long terme

On va maintenant considérer un modèle ARE écrit sur des logarithmes pour faciliter son interprétation:

$$\log y_t = \mu + \alpha \log y_{t-1} + \beta_0 \log x_t + \beta_1 \log x_{t-1} + u_t \quad (41)$$

Si \bar{x} est la valeur constante d'équilibre de x_t , on peut définir la solution d'équilibre du modèle ARE en supprimant le terme d'erreurs:

$$\log \bar{y} = \frac{\mu}{1 - \alpha} + \frac{\beta_0 + \beta_1}{1 - \alpha} \log \bar{x} \quad (42)$$

ou encore:

$$\log \bar{y} = \frac{\mu}{1 - \alpha} + \lambda \log \bar{x} \quad (43)$$

Cette dernière écriture montre que $\lambda = (\beta_0 + \beta_1)/(1 - \alpha_0)$ est l'élasticité de long terme de y par rapport à x . Elle n'a de sens que si $|\alpha| < 1$. On peut également réinterpréter cette relation dans un cadre de croissance équilibrée quand les variables croissent à taux constant. Une relation d'équilibre entre y et x se note alors $y = Kx^\lambda$ ou bien:

$$\log y = \log K + \lambda \log x \quad (44)$$

C'est à ce type de relation que s'intéresse en général la théorie économique. Nous allons montrer maintenant comment une simple reparamétrisation du modèle ARE permet de retrouver cette relation d'équilibre et comment peut se définir un processus d'ajustement par rapport à cette relation d'équilibre. On commence par réarranger le modèle ARE selon:

$$\Delta \log y_t = \mu + (\alpha - 1) \log y_{t-1} + \beta_0 \Delta \log x_t + (\beta_0 + \beta_1) \log x_{t-1} + u_t \quad (45)$$

En mettant $(\alpha - 1)$ en facteur on arrive à:

$$\Delta \log y_t = \mu + \beta_0 \Delta \log x_t + (\alpha - 1) \left[\log y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \alpha} \log x_{t-1} \right] + u_t \quad (46)$$

Le terme entre crochets représente la solution d'équilibre. Tant que le modèle est sur un sentier d'équilibre y et x seront très colinéaires et le terme entre crochets ne va pas beaucoup varier. On pourra donc le confondre avec le terme constant. Le modèle peut donc s'approcher par un modèle en différences seules. Mais pour peu que l'on s'écarte de la solution d'équilibre, les niveaux de y et x vont diverger temporairement. Si y est supérieur à son niveau d'équilibre donné par x^λ le terme entre crochets va être positif. Comme le modèle est stable, $(\alpha - 1)$ est négatif et $\Delta \log y_t$ va être rappelé vers le bas. On a donc une correction de l'erreur qui a été commise, d'où le nom de modèle à correction d'erreur pour cette paramétrisation particulière du modèle ARE. Mais il faut noter qu'il s'agit d'une simple reparamétrisation et donc que ce mécanisme correcteur d'erreur est déjà implicitement contenu dans tout modèle ARE.

Quel est le retard moyen d'ajustement de y_t par rapport à la solution d'équilibre. Celle ci est représentée par le grand crochet. Le retard moyen est alors $\alpha/(1 - \alpha)$. Dans la

paramétrisation retenue, le coefficient de régression de y_{t-1} que l'on appellera a donné par un programme de régression linéaire est égal à $\alpha - 1$. Le retard moyen devient $-(a + 1)/a$. Pour $a = -1$, il n'y a plus de dynamique dans le modèle, et l'ajustement à la solution de long terme est instantané. Pour $a = 0$, le délai d'ajustement devient infini.

Pour terminer notons que certains auteurs imposent une restriction sur la solution de long terme en posant $\lambda = 1$. Dans ce cas l'expression $\log y_{t-1}/x_{t-1}$ est appelée terme correcteur d'erreurs. Cette contrainte correspond à la restriction $\alpha + \beta_0 + \beta_1 = 1$. Elle peut facilement se tester au moyen de la reparamétrisation suivante:

$$\Delta \log y_t = \mu + \beta_0 \Delta \log x_t + (\alpha - 1) \log y_{t-1}/x_{t-1} + (\beta_0 + \beta_1 + \alpha - 1) \log x_{t-1} + u_t \quad (47)$$

Dans cette régression il suffit de tester la nullité du coefficient de régression de x_{t-1} .

8 Stratégies de modélisation

L'approche Box-Jenkins des séries temporelles, mise au point au début des années 1970, s'attache à modéliser une variable unique, mais effectue une recherche de spécification pour le comportement dynamique de cette variable en distinguant trois étapes: identification, estimation, évaluation. L'économétrie, en définissant des stratégies de recherche de spécification, suit une route comparable, mais pour des modèles multivariés et conditionnels. Dans cette section, nous allons introduire une méthodologie particulière, dite de Hendry, du nom de l'auteur qui l'a le plus popularisée, connue aussi comme la méthode du général au particulier. L'idée générale remonte à Sargan (1964), et a ensuite circulé à la LSE. On la retrouve exposée dans divers papiers de Hendry, surtout Hendry and Richard (1982) et Hendry and Richard (1983) et Hendry, Pagan, and Sargan (1984). On présentera tout d'abord les fondements théoriques de cette approche, on décrira ensuite son implémentation pratique.

8.1 Fondements théoriques

La construction d'un modèle empirique se fait selon plusieurs étapes au moyen d'hypothèses simplificatrices dont on pourra ensuite tester certains aspects.

Le PGD. Les données observées ($w_1 \dots w_T$) sont supposées provenir d'un mécanisme économique dynamique inconnu (le processus de génération des données ou PGD) que l'on peut représenter par une distribution jointe conditionnelle aux conditions initiales w_0 et indexée par un vecteur de paramètres qui dépend possiblement du temps:

$$F_W(w_1 \dots w_T | w_0, \psi) \quad (48)$$

Cette distribution est très générale et doit être réduite pour être utilisable. La recherche de spécification d'un modèle est décrite par le processus de réduction² du PGD.

Marginalisation. La première opération de réduction consiste à marginaliser ce processus de manière à en extraire la distribution de x_t qui est un sous ensemble de w_t . Elle est indexée par θ , paramètre issu de ψ .

²On trouvera une théorie des réductions par exemple dans Florens and Moucahrt (1985).

Conditionnement séquentiel. Il est commode pour la suite d'opérer une factorisation de cette densité en un produit de densités conditionnelles séquentielles. Si l'on note l'échantillon joint selon X_1^T , la factorisation recherchée est la suivante:

$$F_X(x_1 \dots x_T | x_0, \theta) = \prod_{t=1}^T F_x(x_t | X_0^{t-1}, \lambda) \quad (49)$$

Cette opération ne comporte aucune restriction.

Normalité et linéarité. On doit maintenant supposer que la variable x_t a été transformée de telle sorte qu'elle puisse être exprimée comme une espérance conditionnelle dans un processus Normal. On peut être amené pour cela à prendre le logarithme de x_t , à introduire des ratios des variables originales. La distribution de x_t s'écrit alors:

$$x_t | X_0^{t-1} \sim N(\mu_t, \Omega_t) \quad (50)$$

Cette hypothèse est relativement contraignante dans la mesure où elle impose de résumer la distribution par les deux premiers moments de X_1^T .

On définit ϵ_t , les innovations du processus, par la différence $x_t - \mu_t$. Ce sont donc des bruits blancs par hypothèse.

L'hypothèse de linéarité implique que μ_t est linéaire en X_1^T . On peut facilement la relâcher, mais il est difficile, sauf cas particulier, d'introduire une forme fonctionnelle particulière, à moins qu'elle ne soit fortement suggérée par la théorie.

Hypothèse de Markov. L'espérance μ_t de x_t dépend pour le moment de tout le passé de la variable. Il est nécessaire de limiter la mémoire du processus par une hypothèse de Markov à l'ordre ℓ . Aussi on écrira la régression :

$$x_t = \pi_{1t}x_{t-1} + \dots + \pi_{\ell t}x_{t-\ell} + \epsilon_t \quad \epsilon_t \sim N(0, \Omega_t) \quad (51)$$

Cette écriture représente un modèle VAR (vector auto-regressive) non contraint. Ses paramètres sont indexés par le temps.

Modèle conditionnel et exogénéité faible. Le problème avec les modèles VAR, c'est que bien souvent leur paramètres ne sont pas constants dans le temps, bien que certaines fonctions de ces paramètres puissent l'être. C'est ce qui arrive quand une équation structurelle de comportement n'est pas stable, tous les paramètres de la forme réduite qui sont des fonctions des paramètres de cette équation, deviennent instables. Une approche classique consiste à partitionner x_t en (y_t', z_t') et à effectuer la factorisation de (8.1) en:

$$F_x(x_t | X_0^{t-1}, \lambda_t) = F_{y|z}(y_t | z_t, X_0^{t-1}, \lambda_{1t}) \times F_z(z_t | X_0^{t-1}, \lambda_{2t}) \quad (52)$$

On a regroupé tous les invariants dans y_t et rejeté dans z_t la partie instable du processus. On dira que z_t est faiblement exogène pour λ_{1t} si on peut négliger le processus marginal de z_t pour l'inférence sur λ_{1t} .

Constance des paramètres. Enfin on va supposer que les paramètres λ_{1t} sont constants dans le temps et donc laisser tomber l'indice t . Cette hypothèse permet au modèle, si elle vérifiée, de bien prédire. Ce qui fait qu'avec ces deux dernières hypothèses réunies, le modèle pour l'inférence devient:

$$F_{y|z}(y_t | z_t, X_0^{t-1}, \lambda_1) \quad (53)$$

8.2 La méthode du général au particulier

Nous venons d'expliquer sur le plan théorique comment arriver à un modèle de régression. Mais cela ne nous dit toujours pas comment, dans la pratique, spécifier une équation de régression. Il est possible de faire la synthèse des travaux empiriques réalisés par David Hendry et ses co-auteurs et ainsi de dégager une méthodologie. L'idée générale de la méthodologie du général au particulier consiste à démarrer une recherche de spécification à partir d'un modèle surparamétrisé qui englobe le processus qui a généré les données et à le réduire séquentiellement. Pagan (1987) relate la méthode en la décomposant en quatre étapes.

- *Un large modèle de départ.* La première étape consiste à définir le modèle en sélectionnant les variables qui entrent dans la relation. Généralement la théorie économique fournit une relation d'équilibre statique, donc essentiellement une liste de variables, mais pas de dynamique. Autour de ce noyau, on va définir un modèle de départ en prenant un grand nombre de retards sur chacune des variables. Prenons l'exemple d'une fonction de consommation usuelle:

$$c_t = c_0 + c_1 r d_t \quad (54)$$

On a deux variables, la consommation des ménages et le revenu disponible, les deux pris en logarithme. Le modèle de départ sera:

$$A(L) c_t = c_0 + B(L) r d_t + u_t \quad (55)$$

où $A(L)$ et $B(L)$ sont deux polynômes de retard de degré r et s avec $a_0 = 1$. Dans le cas d'observations trimestrielles, on prendra par exemple $r = s = 5$. Il ne faut pas choisir des valeurs trop élevées car d'une part on risque de manquer de degrés de liberté pour l'estimation³ et d'autre part une surparamétrisation trop importante risque d'introduire une auto-corrélation artificielle dans les résidus. On va commencer par estimer cette relation surparamétrisée et s'assurer que le modèle de départ est déjà correct, c'est à dire que la théorie retenue est adéquate pour expliquer les données. Cela peut ne pas être le cas. Par exemple la fonction de consommation retenue suppose que la propension moyenne à épargner est constante dans le temps. Si elle a varié, comme certaines études le laissent supposer (chute au milieu des années 80), le modèle de départ n'est déjà pas correct. Un test d'auto-corrélation permet en général de détecter cela.

L'estimation directe de cette forme non contrainte permet déjà d'obtenir certaines caractéristiques du modèle concernant sa solution de long terme et la vitesse d'ajustement à cette solution. Mais il faut pour cela calculer la somme des coefficients et cette paramétrisation ne permet pas de l'obtenir directement. Aussi on passe à une seconde étape.

- *Reparamétrisation.* On va reparamétriser le modèle non contraint de départ de manière à obtenir deux choses: des paramètres qui soient directement interprétables, des régresseurs qui soient à peu près orthogonaux entre eux. Cette dernière exigence permet d'éviter une multicolinéarité trop grande entre les variables retardées. L'interprétabilité des coefficients réclame plus d'attention. Un modèle ARE en niveau peut toujours être reparamétrisé en

³Si k est le nombre d'explicatives et r et s le degré des polynômes, il est bon de se limiter à $2.5 \times T > r + k \times s$ comme le suggère Ericsson and Campos (1990).

un mélange de différences et de niveaux. Cela vient du fait que l'on peut toujours avoir la factorisation suivante du polynôme $A(L)$:

$$A(L) = A(1)L + (1 - L)(1 - A^*(L)) \quad (56)$$

où $A^*(L)$ est de degré $r - 1$ sans terme constant. On peut de même factoriser le polynôme $B(L)$ selon:

$$B(L) = B(1)L + (1 - L)(\beta_0 + B^*(L)) \quad (57)$$

où $B^*(L)$ est de degré $s - 1$ sans terme constant. Partant de $A(L)y_t = B(L)x_t + \epsilon_t$ on arrive à:

$$\begin{aligned} \Delta y_t = & \alpha_1^* \Delta y_{t-1} + \dots + \alpha_{r-1}^* \Delta y_{t-r+1} - (A(1)y_{t-1} - B(1)x_{t-1}) \\ & + \beta_0 \Delta x_t + \beta_1^* \Delta x_{t-1} + \dots + \beta_{s-1}^* \Delta x_{t-s+1} \end{aligned} \quad (58)$$

L'avantage de ce type de reparamétrisation est que l'on fait clairement apparaître les effets de court terme qui sont représentés par les régresseurs en différence et la solution de long terme qui vient des niveaux. On a déjà mis en évidence cela pour un modèle ARE simple dans la section précédente.

- *Simplification.* Dans une recherche de spécification on recherche toujours une certaine forme de parcimonie qui a pour conséquence de fournir des estimateurs plus efficaces et des prévisions meilleures. Il s'agit donc de simplifier le modèle précédent. Deux voies sont possibles: éliminer simplement les variables qui apparaissent avec un t de Student trop faible; chercher de nouvelles combinaisons de paramètres en imposant des restrictions d'égalité par exemple. Cette partie de la méthode est certainement celle qui fait le plus appel à l'intuition économétrique de celui qui la pratique. De plus l'ordre dans lequel sont effectuées les simplifications peut avoir une certaine importance qu'il est difficile d'apprécier. Chaque test de simplification est en effet conditionnel aux tests précédents. Ceci peut introduire des biais car le niveau de significativité des tests est alors changé (voir Giles and Giles (1993)).

Quand doit on s'arrêter dans la simplification? Ou quel est le nombre maximal de retards que l'on doit considérer au début de la recherche de spécification? Il s'agit de pouvoir choisir la taille d'un modèle. Si l'on se fit au seul maximum de vraisemblance ou $\hat{\sigma}^2$, on aura tendance à choisir le modèle le plus gros. D'où l'idée d'avoir recours à un critère d'information que l'on va chercher à minimiser. Le critère d'Akaike (1974) propose de minimiser:

$$AIC(k) = \log \hat{\sigma}_k^2 + \frac{2k}{T} \quad (59)$$

Ce critère a tendance à surestimer la taille du modèle. Un autre critère, proposé par Schwarz (1978) sur des critères Bayésiens ne présente pas ce défaut et conduit à minimiser:

$$SC(k) = \log \hat{\sigma}_k^2 + \frac{k+1}{T} \log T \quad (60)$$

Dans chaque cas, le deuxième terme joue le rôle d'une pénalité sur la taille du modèle. Ces critères permettront de comparer le modèle initial au modèle parcimonieux sélectionné.

- *Evaluation par des tests.* Comment va-t-on juger le produit final? Le modèle obtenu est-il correct et donne-t-il une bonne description du comportement des données? Cette opération de vérification finale s'effectue au moyen d'une batterie de tests de mauvaise spécification. C'est certainement cette dernière étape qui assure la différence entre la méthodologie de Hendry et une pratique routinière de l'économétrie. Les points dont on veut s'assurer sont les suivants:

- le modèle prédit bien
- ses coefficients sont stables dans le temps
- les résidus ne comportent pas de facteur systématique

Un dernier point que l'on peut vérifier c'est que le modèle obtenu enveloppe les modèles précédents de la littérature, en ce sens qu'il permet d'expliquer leurs résultats (test d'encompassing).

Les logiciels actuels d'économétrie permettent en général d'effectuer tous ces tests. Il faut citer par exemple PCGIVE, logiciel issu d'Oxford et du groupe de recherche de David Hendry, TSP, EVIEWS qui est une version PC de TSP, RATS, STATA et encore SAS.

9 Tests de mauvaise spécification

9.1 La théorie usuelle des tests d'hypothèses

La théorie usuelle des tests est due à Neyman et Pearson. Considérons un modèle paramétrique caractérisé par une densité indexée par un paramètre θ . On veut tester l'hypothèse ponctuelle $\theta = \theta_0$, dite *hypothèse nulle* contre l'hypothèse ponctuelle $\theta = \theta_1$, dite *hypothèse alternative*. On observe un échantillon Y de T observations à partir duquel on va calculer une statistique de test. On aura au préalable déterminé la distribution de cette statistique sous l'hypothèse nulle et défini deux régions dans cette distribution: une région d'acceptation et une région de rejet. Selon la valeur de la statistique de test par rapport à ces deux régions, on va accepter ou rejeter l'hypothèse nulle en commettant deux types d'erreurs:

- Type I : on rejette à tort H_0 . On minimise cette erreur en choisissant une taille ou un seuil pour le test, habituellement 5% sans qu'il faille attacher une valeur magique à ce chiffre.
- Type II : on accepte à tort l'hypothèse nulle. Pour obtenir le meilleur test possible, on va chercher à maximiser sa puissance, c'est à dire $1 - \text{Prob}(\text{erreur type II})$.

Pour deux hypothèses ponctuelles, le lemme de Neyman-Pearson montre que le test le plus puissant définit une zone critique de taille α au moyen d'une constante κ et du rapport des vraisemblances:

$$L(\theta_0)/L(\theta_1) \leq \kappa \quad L(\theta_0)/L(\theta_1) > \kappa.$$

On rejettera H_0 quand le rapport de vraisemblance sera plus grand que κ . On ne rejettera pas H_0 dans le cas contraire.

9.1.1 Le test du rapport de vraisemblance

Le lemme de Neyman-Pearson suggère une méthode de construction de test (le rapport de vraisemblance), mais ne dit rien quand l'hypothèse alternative n'est plus ponctuelle. Or dans la pratique, l'hypothèse alternative n'est que très rarement ponctuelle et le cas standard consistera à tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. On arrivera en général à construire des tests consistants, c'est à dire des tests où la probabilité de l'erreur de type I est bien égale au seuil choisi. Mais, en général, le test ainsi construit ne sera pas, sauf exception, un test optimal, c'est à dire un test qui a la puissance maximum.

Dans le modèle linéaire, l'approche de Neyman et Pearson permet de construire un test exact pour tester l'inclusion d'un groupe de m régresseurs supplémentaires. Le test se construit en considérant la somme des carrés des erreurs SSE_0 dans le cas contraint sous H_0 et cette même somme SSE_1 dans le cas non contraint. On ne connaît pas la distribution du rapport $(SSE_0/SSE_1)^{-T/2}$, mais on connaît celle d'une transformation de ce rapport qui correspond à un test en F :

$$\frac{SSE_0 - SSE_1}{SSE_1} / \frac{m}{T - k} \sim F(m, T - k).$$

Quand $m = 1$, ce test est équivalent à un test de Student.

Dans les modèles non-linéaires, il va falloir se contenter de résultats de grand échantillon. On va estimer le modèle sous H_0 , ce qui va conduire à définir $L(\theta_0)$. Sous l'hypothèse alternative, θ est totalement libre. On calculera donc la fonction de vraisemblance en son maximum pour l'échantillon considéré. Alors la statistique suivante aura une distribution du χ^2 avec:

$$-2 \log \frac{L(\theta_0)}{L(\hat{\theta})} \sim \chi^2(m),$$

où m est la taille de la restriction. Ce test n'est en général pas optimal, mais il est consistant.

9.1.2 Le test de Wald

L'approche du rapport de vraisemblance impose que l'on estime le modèle à la fois sous l'hypothèse nulle et sous l'alternative. Il est des cas où l'estimation sous l'hypothèse nulle est difficile. On va alors considérer un test de Wald qui est basé sur la remarque suivante concernant l'estimateur du maximum de vraisemblance

$$\sqrt{T}(\hat{\theta} - \theta) \rightarrow N(0, IA^{-1}(\theta)).$$

Si par exemple le modèle contraint implique une restriction linéaire du type $H_0 : R\theta = r$ où R est une matrice $m \times k$ fixe de contraintes et r un vecteur connu de dimension m , alors on a par le théorème de transformation

$$\sqrt{T}R(\hat{\theta} - \theta) = \sqrt{T}(R\hat{\theta} - r) \rightarrow N(0, RIA^{-1}(\theta)R'),$$

ce qui fait que la statistique suivante a une distribution du χ^2 :

$$T(R\hat{\theta} - r)'[RIA^{-1}(\theta)R']^{-1}(R\hat{\theta} - r) \sim \chi^2(m).$$

Ce test se généralise au cas de contraintes non-linéaires.

9.1.3 Le test de Lagrange

Le test de Lagrange correspond au cas exactement symétrique du test de Wald. Le modèle a été estimé sous H_0 et l'estimer sous l'hypothèse alternative paraît difficile. Appelons $\hat{\theta}_0$ l'estimation par maximum de vraisemblance de θ . Si H_0 est vraie, alors la valeur $\hat{\theta}$ sera proche de l'estimation de θ sous l'hypothèse alternative. Supposons que l'on ait l'expression de la fonction de vraisemblance du modèle non contraint que l'on va noter $L(\theta)$. Alors on aura

$$\frac{\partial \log L(\theta)}{\partial \theta} \Big|_{\hat{\theta}_0} = D \log L(\hat{\theta}_0) \simeq 0.$$

Comment tester que cette quantité est effectivement proche de 0? On va former la statistique de test suivante sur la base d'une forme quadratique

$$(D \log L(\hat{\theta}_0))' I^{-1}(\hat{\theta}_0) D \log L(\hat{\theta}_0)$$

où I est la matrice d'information. En utilisant un développement de Taylor, on peut montrer que cette statistique a la même distribution asymptotique que la statistique de test du rapport de vraisemblance, c'est à dire une χ^2 à m degrés de libertés, où m est le nombre de contraintes. Voir par exemple Harvey (1981), page 168 pour une preuve.

Dans beaucoup de problèmes et sous l'hypothèse de normalité, la fonction de vraisemblance peut se réduire à un facteur additif près à:

$$\log L = -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum e_t^2,$$

ce qui fait que

$$\frac{\partial \log L(\theta, \sigma^2)}{\partial \theta} = \sigma^{-2} \sum \frac{\partial e_t}{\partial \theta} e_t.$$

Après quelques calculs, on peut montrer que la statistique LM s'écrit dans ce cas:

$$LM = \sigma_0^{-2} (\sum z_t e_t)' (\sum z_t z_t')^{-1} (\sum z_t e_t),$$

avec $z_t = \text{et } \hat{\sigma}_0^2 = T^{-1} \sum e_t(\theta_0)$. Cette expression peut se réinterpréter comme T fois le coefficient de corrélation de la régression linéaire de $e_t(\theta_0)$ sur $\frac{\partial e_t}{\partial \theta_0}$. Le test LM consistera alors à tester si TR^2 est supérieur à la valeur critique d'une χ^2 à m degrés de liberté ou non.

9.1.4 Relation entre les trois tests

Si pour un même problème on calcule les trois statistiques de test, W , LR et LM , on va en général trouver que

$$W \geq LR \geq LM.$$

Si le modèle est une régression linéaire, les trois statistiques seront alors strictement équivalentes et se calculent en général sous la forme d'un test en F . Le test de Wald sera en général un peu plus puissant que le test LM, mais au prix d'une distorsion de taille en échantillon

fini. Par contre, en grand échantillon, les trois tests ont la même distribuion asymptotique sous l'hypothèse nulle et sont donc équivalents.

Cette équivalence est utile à connaître, mais on ne peut en déduire des considérations sur la puissance des trois tests, car de toute façon sont asymptotiquement équivalents.

On retiendra que le test LR est sans doute le plus intuitif, mais il nécessite l'estimation du modèle sous les deux hypothèses. Le test W ne nécessite que l'estimation du modèle sous l'alternative alors que le test LM n'a besoin que de l'estimation sous l'hypothèse nulle. C'est ce type de test que l'on va utiliser pour investiguer la bonne spécification d'un modèle.

9.2 Tests de mauvaise spécification

La construction d'un modèle de régression (ou de tout autre modèle plus complexe) met en jeu diverses hypothèses simplificatrices dont il s'agit de mesurer la validité au moyen de tests de mauvaise spécification. Ces tests sont puissants pour plusieurs directions de mauvaise spécification. Si l'hypothèse nulle est rejetée, cela ne conduit pas du tout à accepter l'alternative, car le test a pu rejeter H_0 pour une raison différente que celle contenue dans H_1 . Ainsi quand un test détecte par exemple de l'auto-corrélation dans les résidus, cela ne veut pas dire qu'il faille corriger l'estimation par une procédure de Cochrane-Orcut si populaire pourtant dans les logiciels. Le problème peut très bien venir, et c'est en général le cas, de variables manquantes⁴.

Il est possible de donner une vision unifiée de ces tests en distinguant selon la terminologie de Davidson and MacKinnon (1985) les tests qui peuvent s'exprimer comme selon une direction de régression et ceux qui s'expriment selon une direction de moments d'ordre supérieur. Pour la première catégorie de tests, l'hypothèse alternative peut être construite comme une autre formulation de l'espérance conditionnelle de y_t . La seconde catégorie de tests concerne la violation d'hypothèses faites sur les moments d'ordre deux ou supérieur du terme d'erreur, comme la normalité et l'hétéroscédasticité. Pour implémenter ces tests, il suffit de construire une régression augmentée pour la première catégorie et en général une régression auxiliaire pour la seconde catégorie. L'hypothèse nulle de bonne spécification correspondra alors à la nullité d'un sous ensemble de coefficients testée au moyen d'un test en F . Pagan (1984) a systématisé ce type d'approche. Si le modèle de régression initial se note:

$$y = X\beta + \epsilon$$

avec y vecteur d'observations sur l'endogène et X matrice d'observations sur les exogènes, une régression augmentée se définit alors au moyen de:

$$y = X\beta + A\gamma + \epsilon \tag{61}$$

où A est une matrice d'observations définissant la direction d'augmentation pour l'espérance conditionnelle de y . Le modèle sera bien spécifié si on n'a pas besoin de A pour modéliser l'espérance conditionnelle de y , donc si un test permet d'accepter $\gamma = 0$. Si l'on fait

⁴Le seul cas où il est intéressant d'avoir un modèle à erreurs auto-corrélées, c'est quand on a réussi à mettre en évidence une restriction de facteur commun.

l'hypothèse que les erreurs ϵ_t sont $N(0, \sigma^2)$ et indépendantes, alors l'hypothèse nulle pourra se tester au moyen d'un test en F exact

$$\frac{(SS0 - SS1)/m}{SS1/(T - k)} \sim F(m, T - k - m).$$

Quand il s'agit de tester une hypothèse sur le moment d'ordre deux de y_t , la régression auxiliaire sera de la forme:

$$\hat{\epsilon}_t^2 = \sigma^2 + A_t' \gamma + u_t, \quad (62)$$

où $\hat{\epsilon}_t^2$ représente le résidu estimé de l'équation linéaire initiale. Là encore, le modèle sera bien spécifié si on peut accepter l'hypothèse que $\hat{\gamma} = 0$.

Les tests ainsi conçus appartiennent à la classe des tests dits de Lagrange car le modèle sous l'hypothèse alternative n'est jamais estimé. Dans un test de Lagrange, le score de la vraisemblance sous l'hypothèse alternative est évalué en se servant de la valeur des paramètres sous l'hypothèse nulle. Si l'hypothèse nulle est correcte, ce score, qui est nul sous H_1 , le sera aussi sous H_0 . L'avantage de ce test, c'est qu'il permet de n'avoir à estimer le modèle que sous la contrainte de H_0 . De plus si le modèle est linéaire, on se ramène au test en F habituel, d'où la correspondance entre ce test et l'approche par régression augmentée.

9.3 Tests selon une direction de régression

9.3.1 Auto-corrélation

La présence d'auto-corrélation des résidus dans les modèles ARE entraîne une liaison entre la variable endogène retardée et le terme d'erreur, ce qui rend les moindres carrés inconsistants. De plus la présence d'auto-corrélation dans les résidus d'une régression est très souvent le signe de l'absence d'une variable. Soit une variable économique pertinente n'a pas été introduite dans le modèle, soit le modèle ne présente pas suffisamment de retards pour les variables existantes, ou il en présente trop ce qui induit une surparamétrisation. En présence d'une variable endogène retardée, le test de Durbin et Watson n'est plus applicable. Sa valeur est biaisée vers 2. La présence de l'endogène retardée rend inapplicable la théorie asymptotique utilisée pour déterminer la distribution de ce test⁵. Le test de Lagrange modifié suggéré par Durbin (1970) est basé sur une régression augmentée construite en ajoutant à la régression étudiée les résidus estimés retardés de un. Quand cette régression augmentée est estimée par moindres carrés, la statistique de Student associée

⁵Durbin (1970) a proposé de modifier ce test de manière à tenir compte de la présence de l'endogène retardée. Ce test est le "Durbin h test" défini par:

$$h = \sqrt{T} \frac{1 - 0.5DW}{\sqrt{1 - T \text{Var}(\hat{\alpha})}}.$$

Sous l'hypothèse nulle d'absence d'auto-corrélation des résidus, le test est distribué selon une loi normale de moyenne nulle et de variance unité. Pour tester une auto-corrélation positive au seuil de 5% on prendra $h > 1.645$ comme région critique. Ce test n'est malheureusement pas très puissant. De plus il n'est pas toujours possible de l'utiliser. Il faut pour cela que $T \text{Var}(\hat{\alpha}) < 1$.

au coefficient de régression des résidus retardés sert de test asymptotique pour détecter la présence d'auto-corrélation à l'ordre un.

Nous allons montrer maintenant comment construire la régression augmentée qui permet d'avoir une expression commode du test. Soit donc le modèle de régression avec erreurs auto-corrélées à l'ordre un suivant:

$$y_t = x_t' \beta + u_t \quad (63)$$

$$u_t = \rho u_{t-1} + \epsilon_t \quad (64)$$

avec $\epsilon_t \sim IN(0, \sigma^2)$. x_t représente l'observation de k variables exogènes. Ce vecteur peut très bien contenir des retards de l'endogène sans qu'ici le problème soit changé. Avec l'opérateur retard L on a $u = \epsilon / (1 - \rho L)$ pourvu que $|\rho| < 1$. Alors on peut réécrire le modèle en une seule équation sous la forme:

$$y_t = x_t' \beta + \rho y_{t-1} - x_{t-1}' \beta \rho + \epsilon_t. \quad (65)$$

Cette forme n'est pas très pratique car elle comporte une non linéarité en $\beta \rho$. On va choisir de la linéariser par un développement de Taylor au premier ordre autour de deux valeurs pour β et ρ que l'on va temporairement désigner par $\hat{\beta}$ et $\hat{\rho}$:

$$\beta \rho \simeq \hat{\beta} \hat{\rho} + (\rho - \hat{\rho}) \hat{\beta} + (\beta - \hat{\beta}) \hat{\rho} \quad (66)$$

Comme on s'intéresse à un test de Lagrange, on va choisir pour valeurs de $\hat{\beta}$ et $\hat{\rho}$ les valeurs qui correspondent à l'estimateur sous H_0 , c'est à dire les OLS pour $\hat{\beta}$ et zéro pour $\hat{\rho}$. Donc il reste:

$$y_t = x_t' \beta + (y_{t-1} - x_{t-1}' \hat{\beta}) \rho + \epsilon_t \quad (67)$$

$$y_t = x_t' \beta + \hat{u}_{t-1} \rho + \epsilon_t, \quad (68)$$

où \hat{u}_{t-1} représente les résidus de la régression sous H_0 . Le test de Lagrange est maintenant équivalent à un test en F sur la nullité du coefficient de \hat{u}_{t-1} avec $(1, T - k - 1)$ degrés de liberté. On peut étendre ce test à un ordre d'auto-corrélation supérieur en incluant r retards de \hat{u}_t . Le test est alors un F à $(r, T - k - r)$ degrés de liberté.

9.3.2 Le test du RESET ou test de Ramsey

Il s'agit d'un test général de variables manquantes. Très souvent il est difficile d'avoir une idée sur la nature de ces variables. L'approche de Ramsey (1969) suppose que l'effet des variables omises dans le modèle peut être approché par une fonction polynomiale de $X\beta$. Si l'on remplace β par son estimateur OLS, la régression initiale sera alors augmentée par des puissances de \hat{y} avec

$$A_t \gamma = \sum_{i=1}^p \hat{y}^{i+1} \gamma_i$$

Des expériences de Monte Carlo ont montré que la performance de ce test était optimale avec $p = 3$.

La question avec ce test est de savoir si des puissances de $X\beta$ sont de bonnes approximations pour la forme fonctionnelle que l'on a omise. Thursby and Schmidt (1977) suggèrent d'utiliser des puissances des régresseurs individuels au lieu des puissances de \hat{y} . Le test construit de cette façon semble supérieur au test du RESET.

9.3.3 Test sur la forme fonctionnelle

Bien souvent, un modèle est spécifié de manière linéaire parce que le modèle alternatif à estimer paraît trop compliqué. On peut citer par exemple la fonction de production Cobb-Douglas qui s'estime par moindres carrés après passage aux logarithmes alors que la fonction de production CES oblige à une estimation non-linéaire. Mais le modèle linéaire est-il toujours une simplification acceptable? Harvey (1981) page 172 cite également un modèle de demande de monnaie présentant une trappe à liquidité

$$\log m_t/p_t = \log y_t/p_t \beta_1 + \beta_2/(r_t - \gamma) + \epsilon_t.$$

Comment tester que $\gamma = 0$ dans ce modèle?

Les deux modèles précédemment cités sont des modèles qui deviennent linéaires sous H_0 en imposant que l'élasticité de substitution soit égale à 1 dans un cas ou que $\gamma = 0$ dans l'autre cas. On peut formaliser la situation en écrivant

$$y = f(X, \beta, \theta) + \epsilon \quad (69)$$

avec $\theta = \theta_0$ sous H_0 . Un développement de Taylor de la forme fonctionnelle autour de $\theta = \theta_0$ permet d'écrire

$$y = f(X, \beta, \theta_0) + \frac{\partial f(X, \hat{\beta}, \theta_0)}{\partial \theta} (\theta - \theta_0) + \epsilon \quad (70)$$

ou encore

$$y = X\beta + \frac{\partial f(X, \hat{\beta}, \theta_0)}{\partial \theta} (\theta - \theta_0) + \epsilon \quad (71)$$

d'où l'on déduit que

$$A = \frac{\partial f(X, \hat{\beta}, \theta_0)}{\partial \theta} \quad (72)$$

9.3.4 Test de stabilité

L'hypothèse de stationnarité conditionnelle implique que les coefficients de régression ne varient pas dans le temps. On peut tester cette hypothèse en postulant que sous l'hypothèse alternative les coefficients de régression sont variables et suivent un modèle du type:

$$\beta_t = \beta + z_t \gamma \quad (73)$$

où z_t contient les variables explicatives qui conditionnent l'évolution de β_t . Pour construire la régression de test, il suffit de remplacer β_t par sa valeur dans la régression initiale, ce qui donne:

$$y_t = x_t' \beta + x_t' z_t \gamma + u_t \quad (74)$$

La variable A dans (9.2) sera alors égale à $x_t' z_t$. Il faut noter que ce test est sensible aux problèmes d'hétéroscédasticité.

Le cas le plus simple est celui où z_t représente une variable muette qui vaut 0 avant une certaine date et 1 ensuite, ce qui fait qu'il y a T_1 observations dans un premier régime et

T_2 dans le second. Le coefficient de régression change donc à partir d'une date fixe. C'est le test de Chow (1960). Il est intéressant de présenter le problème sous forme matricielle. Sous l'hypothèse nulle, le modèle s'écrit

$$y = X\beta + u$$

et le paramètre de régression β sera le même pour tout l'échantillon. Si par contre on suppose une rupture à une date donnée, le modèle va s'écrire sous l'hypothèse alternative:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Ce modèle ne peut s'estimer que s'il y a suffisamment d'observations dans chaque régime. Le test s'effectue en comparant la somme de résidus au carré sous chaque hypothèse au moyen de

$$\frac{(u'u - (u_1'u_1 + u_2'u_2))/k}{(u_1'u_1 + u_2'u_2)/(t - 2k)} \sim F(k, T - 2k)$$

Dans le cas où il n'y a pas suffisamment d'observations dans un des régimes, il faut utiliser une forme alternative qui correspond à

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} 0 \\ X_2 \end{pmatrix} (\beta - \beta_2) \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

9.4 Postsample prediction failure

Pendant toute la procédure de recherche de spécification, on va laisser de côté T_2 observations en fin d'échantillon. On prendra en général $T_2 = 48$ ou $T_2 = 8$ dans le cas d'observations trimestrielles, c'est à dire une ou deux années. Une fois retenue la spécification du modèle, on va le tester pour savoir s'il prédit bien les valeurs de l'échantillon laissées de côté et donc tester la nullité des erreurs de prévision, ou bien le fait que la variance des erreurs de prévision n'est pas supérieure à la variance estimée des termes d'erreurs. Il existe plusieurs façon d'effectuer ce test. Si SSE_1 est la somme des carrés des erreurs obtenue pour le premier échantillon, on va ensuite estimer le modèle sur l'échantillon complet pour obtenir SS_0 et tester l'égalité de ces deux sommes au moyen d'un test en F:

$$\frac{(SS_0 - SS_1)/T_2}{SS_1/(T_1 - k)} \sim F(T_2, T_1 - k)$$

On peut également procéder par regression augmentée en introduisant la variable

$$A = \begin{pmatrix} 0 \\ I_2 \end{pmatrix}$$

où I_2 est une matrice identité de taille T_2 dans une regression portant sur l'échantillon complet. Le paramètre γ est de dimension T_2 et représente les erreurs de prédiction du modèle. Un test de nullité de γ est alors un test de non significativité des erreurs de prédiction, test en t cette fois-ci.

9.4.1 Test d'exogénéité

L'hypothèse testée ici consiste à savoir si les variables exogènes sont vraiment indépendantes du terme d'erreur ϵ_t . Le problème a été étudié de manière extensive dans la littérature (voir les références dans Godfreys (1988), page 145 ou Spanos (1986)). On peut poser le problème de la manière suivante. Soit le modèle de régression suivant:

$$y_t = x_t'\beta_1 + Y_t'\beta_2 + \epsilon_t. \quad (75)$$

On suspecte Y_t de ne pas être exogène. Cela signifie en fait que l'on a négligé le modèle qui génère Y_t quand on a estimé le premier modèle, alors que ce modèle était nécessaire pour l'inférence sur β_2 . Il faut donc réintroduire ce modèle explicatif auxiliaire de Y_t et on lui donnera la forme:

$$Y_t = W_t'\delta + v_t \quad (76)$$

où W_t est un ensemble de variables exogènes contenant x_t , mais également d'autres variables exogènes en nombre suffisant, c'est à dire qu'il faut qu'il y en ait au moins un nombre égal à la taille de Y_t' . A partir de ce modèle auxiliaire, on va calculer les résidus estimés \hat{v}_t pour les introduire ensuite dans la régression initiale. Si le coefficient associé à cette variable supplémentaire est non-significatif, alors il n'y a pas de problème d'exogénéité. Ce test a été initialement proposé par Hausman (1978).

9.5 Augmentation de la variance conditionnelle

Après s'être intéressés à une mauvaise spécification de l'espérance conditionnelle, nous allons maintenant examiner le moment d'ordre deux des résidus. Les deux problèmes ne sont pas indépendants. Il est possible par exemple de montrer qu'un problème de non-constance des paramètres peut se traduire exactement en un problème d'hétéroscédasticité, c'est à dire de non-constance de la variance des résidus dans le temps.

L'hétéroscédasticité rend l'estimateur des moindres carrés inefficace en ce sens que l'écart type des coefficients estimé sera mal estimé. L'hétéroscédasticité va entraîner des problèmes au niveau des tests classiques de spécification (nullité d'un coefficient de régression). Mais également certains tests de mauvaise spécification, comme le test du RESET, ne vont plus avoir la même distribution en présence d'hétéroscédasticité.

Là encore un test d'homoscédasticité va pouvoir détecter une mauvaise spécification dans une direction différente de celle pour laquelle il est conçu. Ce type de test est puissant contre toute une variété d'alternatives incluant l'omission de variables ou une forme fonctionnelle incorrecte. On ne pourra donc corriger le modèle en suivant le résultat du test (en modelant par exemple une forme particulière de variance non constante).

Une façon à peu près générale de modéliser une variance non constante dans le temps consiste à utiliser la formulation de Breuch and Pagan (1979):

$$\begin{cases} y_t = x_t'\beta + u_t \\ \sigma_t^2 = \sigma^2 h(z_t'\gamma) \\ u_t \sim NID(0, \sigma_t^2) \end{cases}$$

La variance σ_t^2 dépend des variables exogènes z_t (possiblement reliées aux x_t) par l'intermédiaire de la fonction positive $h(\cdot)$ et du paramètre γ avec $h(0) = 1$. Donc pour $\gamma = 0$, on retrouvera l'homoscédasticité. Notons que pour la matrice d'information ne soit pas singulière sous H_0 , il faut que $h'(0) \neq 0$. Cette formulation recouvre les cas d'hétéroscédasticité multiplicative avec $h(z_t'\gamma) = \exp(z_t'\gamma)$ et d'hétéroscédasticité additive avec $h(z_t'\gamma) = (1 + z_t'\gamma)$. Notons toutefois que dans la mesure où $h'(0)$ est une constante, le test de Lagrange associé à l'hypothèse nulle $\gamma = 0$ sera invariant par rapport à la forme spécifique de $h(\cdot)$ [voir Godfreys (1988), page 126]. On basera le test sur une régression auxiliaire de la forme:

$$\hat{u}_t^2 = \sigma^2 + z_t'\gamma + v_t \quad (77)$$

où les \hat{u}_t^2 représentent le carré des résidus de la régression initiale.

9.5.1 Variances inégales

On suppose simplement que σ_t^2 ne prend que deux valeurs selon que $t \leq t_0$ ou $t > t_0$. z_t dans ce cas n'est qu'une variable indicatrice qui vaut 1 avant t_0 et zéro au delà.

$$\hat{u}_t^2 = \sigma^2 + d_t\gamma + v_t \quad (78)$$

où d_t est une variable muette qui vaut 0 si $t < t_0$ et 1 sinon. Il faut bien sûr avoir une idée a priori pour choisir la valeur de t_0 . Ce test est asymptotiquement équivalent au test de Test de Goldfeld and Quandt (1965).

9.5.2 Arch

Engle (1982) a introduit un modèle où la variance des erreurs dépend de la taille des erreurs passées. Ce modèle est fort utilisé pour modéliser la variabilité des séries financières. Concrètement la variance σ_t^2 est modélisée selon:

$$\sigma_t^2 = \sigma^2 + \sum_{i=1}^r \alpha_i u_{t-i}^2 \quad (79)$$

Ce test est plus facile à mettre en oeuvre que le précédent, car on n'a qu'à choisir le nombre de retards dans la régression suivante

$$\hat{u}_t^2 = \omega + \sum_{i=1}^r \alpha_i \hat{u}_{t-i}^2 + v_t. \quad (80)$$

On n'aura aucun effet ARCH si les α_i sont tous égaux à 0.

9.6 Augmentation des moments d'ordre supérieur

Il s'agit essentiellement du test de normalité des résidus. Les arguments asymptotiques permettraient de penser que la normalité des résidus est sans importance. Pourtant la normalité est nécessaire pour pouvoir utiliser un test en F ou en t en petit échantillon. D'autre part, certains tests asymptotiques nécessitent la normalité des résidus. Enfin, les OLS ne

produisent pas une estimation robuste si les résidus s'écartent très fort de la normalité (d'où d'autres estimateurs basés sur des termes d'erreur Student, ou des fonctions de perte non quadratiques).

Un test de normalité a été proposé par Jarque and Bera (1980). Pour introduire ce test de manière simple, on peut remarquer que la distribution normale a des propriétés très particulières quant à ses moments. *Premièrement*, le moment centré d'ordre trois est nul (ainsi que tous les moments centrés impairs). C'est d'ailleurs une caractéristique de toutes les distributions symétriques. Pour mesurer l'asymétrie d'une distribution, on calculera le ratio μ^3/σ^3 . Ce ratio n'a pas d'échelle. *Deuxièmement*, on peut mesurer l'aplatissement d'une distribution par le ratio μ^4/σ^4 . Si l'on calcule ce ratio pour la Normale, on s'aperçoit qu'il est égal à 3. Le test de normalité de Jarque et Bera teste de manière conjointe la conformité à deux caractéristiques de deux statistiques sur les résidus. Pour cela, on va estimer les quantités suivantes

$$\mu_k = \frac{1}{T} \sum (\hat{u}_t - \bar{u}_t)^k$$

A partir de ces quantités, on va calculer deux ratios correspondant à la skewness et à la kurtosis, ratios qui vont avoir chacun une distribution asymptotique normale. Pour le coefficient empirique d'asymétrie, on a:

$$S_T^{1/2} = \frac{\mu_3}{\mu_2^{3/2}} \rightarrow N(0, \sqrt{\frac{6}{T}}).$$

Pour le coefficient d'aplatissement, on obtient:

$$K_T = \frac{\mu_4}{\mu_2^2} \rightarrow N(3, \sqrt{\frac{24}{T}}).$$

Le test de Jarque and Bera (1980) s'obtient en prenant la somme de chacune de ces statistiques au carré. Le résultat final sera donc distribué selon une $\chi^2(2)$ sous l'hypothèse nulle:

$$JB = \frac{T}{6} S_T^2 + \frac{T}{24} (K_T - 3)^2 \rightarrow \chi^2(2). \quad (81)$$

Il faut toutefois noter qu'en échantillon fini les deux parties du test ne sont pas indépendantes, alors qu'asymptotiquement elles le sont. Les valeurs critiques asymptotiques ne donnent donc qu'une idée approximative des valeurs critiques de petit échantillon. Ce test a été développé en supposant que les observations étaient indépendantes. Il existe un test alternatif développé par Bai and Ng (2005) qui tient compte de la dépendance des observations, mais il ne semble pas encore beaucoup employé dans la pratique.

10 La demande de monnaie en Belgique

On va maintenant traiter un petit exemple empirique qui concerne la demande de monnaie $M1$ en Belgique. Les données sont annuelles et couvrent la période 1953 – 1982. La

théorie économique fournit une relation d'équilibre qui relie le logarithme de la quantité réelle de monnaie LMP au logarithme du revenu réel LYP et au taux d'intérêt R . On a pris les définitions suivantes pour les variables. $M1$ est la masse monétaire $M1$, YP est le revenu disponible réel des particuliers. R est le taux sur les certificats de trésorerie à trois mois. Afin de conserver une relation en logarithme, on va prendre la transformation $LR = \log(1 + r_t/100)$. On va alors partir d'un modèle dynamique général surparamétré qui comprendra deux retards de chaque variable et dont l'estimation par OLS donne:

$$\begin{aligned}
 LMP = & \underset{[2.56]}{0.56} LMP1 + \underset{[1.19]}{0.26} LMP2 + \underset{[2.07]}{0.69} LYP + \underset{[0.00]}{0.01} LYP1 - \underset{[-1.50]}{0.55} LYP2 \\
 & - \underset{[-2.69]}{0.96} LR - \underset{[-0.94]}{0.52} LR1 + \underset{[1.25]}{0.67} LR2 - \underset{[-1.00]}{0.70} \\
 R^2 = & 0.99 \quad \hat{\sigma} = 0.023 \quad ss = 0.0103 \\
 \chi^2(1)_{\text{corr}} = & 1.83 \quad \chi^2(2)_{\text{norm}} = 0.60 \quad \chi^2(2)_{\text{arch}} = 0.74
 \end{aligned}$$

Cette équation semble un bon départ de modélisation car elle passe les tests de mauvaise spécification effectués (on rappelle que la valeur critique à 5% du test du χ^2 est de 3.84 pour 1 degré de liberté et de 5.99 pour 2 degrés de liberté). Mais ses coefficients sont mal estimés et peu interprétables. Au moyen d'un petit calcul qui consiste à prendre la somme des coefficients des polynômes de retard, on peut pourtant faire apparaître la solution de long terme qui est:

$$0.18LMP = 0.15LYP - 0.81LR - 0.70$$

Nous allons maintenant reparamétriser cette équation de manière à faire apparaître directement la solution de long terme:

$$\begin{aligned}
 \Delta LMP = & - \underset{[-1.08]}{0.18} LMP1 - \underset{[-1.19]}{0.26} \Delta LMP1 + \underset{[2.07]}{0.68} \Delta LYP + \underset{[1.03]}{0.15} LYP1 + \underset{[1.50]}{0.55} \Delta LYP1 \\
 & - \underset{[-2.69]}{0.96} \Delta LR - \underset{[-1.36]}{0.80} LR1 - \underset{[1.25]}{0.68} \Delta LR1 - \underset{[-1.00]}{0.70} \\
 R^2 = & 0.73 \quad \hat{\sigma} = 0.023
 \end{aligned}$$

A part le R^2 qui dépend de la moyenne de l'endogène, les caractéristiques de l'équation n'ont pas changé. On peut remarquer que les coefficients des niveaux sont identiques à ceux trouvés précédemment pour la solution de long terme. Si maintenant on simplifie l'équation en éliminant les variables qui ne sont pas significatives, on arrive à:

$$\begin{aligned}
 \Delta LMP = & - \underset{[-2.16]}{0.26} LMP1 + \underset{[3.52]}{0.94} \Delta LYP + \underset{[2.43]}{0.23} LYP1 - \underset{[-4.60]}{1.30} \Delta LR - \underset{[-3.08]}{1.21} LR1 - \underset{[-2.42]}{1.17} \\
 R^2 = & 0.69 \quad \hat{\sigma} = 0.023 \quad ss = 0.0119 \\
 \chi^2(1)_{\text{corr}} = & 1.28 \quad \chi^2(2)_{\text{norm}} = 0.88 \quad \chi^2(2)_{\text{arch}} = 1.45
 \end{aligned}$$

qui passe encore les tests de mauvaise spécification. Son écart type n'a pas changé, ce qui laisse supposer que les restrictions imposées sont acceptables. Un test en F formel pour m restrictions par rapport aux k variables initiales donne:

$$\frac{(ss_0 - ss)/m}{ss/(T - k)} = \frac{(0.0119 - 0.0103)/3}{0.0103/19} = 0.98 \sim F(3, 19)$$

avec une valeur critique à 5% de 3.13. Testons maintenant le modèle en prévision en ne retenant que la période 1954 – 1980 pour l'estimation. Le test de "prédictive failure" sur les deux dernières observations est $\chi^2(2) = 1.71$. On a donc une équation parcimonieuse qui se comporte relativement bien.

On va maintenant tester la restriction de l'égalité à un de l'élasticité de long terme de la monnaie par rapport au revenu. On définit pour cela la variable $LMPY = LPM - LYP$ qui est le logarithme de l'inverse de la vitesse de circulation de la monnaie. On peut procéder de deux façons, soit en estimant l'équation non contrainte suivante:

$$\Delta LMP = - \underset{[-2.16]}{0.26} LMPY1 + \underset{[3.52]}{0.94} \Delta LYP - \underset{[-0.74]}{0.03} LYP1 - \underset{[-4.60]}{1.30} \Delta LR - \underset{[-3.08]}{1.21} LR1 - \underset{[-2.42]}{1.17}$$

$$R^2 = 0.69 \quad \hat{\sigma} = 0.023 \quad ss = 0.0119$$

On observe alors que le coefficient de $LYP1$ n'est pas statistiquement différent de zéro. On peut estimer directement l'équation contrainte:

$$\Delta LMP = - \underset{[-2.39]}{0.20} LMPY1 + \underset{[3.63]}{0.84} \Delta LYP - \underset{[-4.72]}{1.32} \Delta LR - \underset{[-3.19]}{1.24} LR1 - \underset{[-2.63]}{1.00}$$

$$R^2 = 0.69 \quad \hat{\sigma} = 0.023 \quad ss = 0.0122$$

Le test en F pour une contrainte est le suivant:

$$\frac{(0.0123 - 0.0119)/1}{0.0119/23} = 0.46 \sim F(1, 23)$$

avec une valeur critique à 5% de 4.38, ce qui fait que la restriction est aussi acceptée par cette méthode. Aux erreurs d'arrondi près cette statistique devrait être égale au carré du test de Student associé à $LYP1$ calculé dans la régression précédente. La solution de long terme cette fois-ci est:

$$LMP = LYP - 6.20LR - 5.00$$

que l'on peut encore écrire avec M et Y représentant des quantités nominales:

$$\frac{M}{Y} = e^{-5} \cdot (1 + r/100)^{-6.20}$$

Cette application a permis d'illustrer les différentes étapes de recherche de spécification avec un modèle ECM. Elle illustre également un certain état de l'art prédominant avant l'introduction dans la littérature des concepts de racine unitaire et de cointégration.

11 Conclusion

Dans ce chapitre, on a vu comment modéliser une seule variable y_t conditionnellement à son passé et à d'autres variables dites exogènes et dont le DGP demeure non spécifié. On verra dans un chapitre ultérieur comment modéliser de façon simultanée plusieurs variables et l'on retrouvera cette opposition apparente entre modélisation statistique et

modélisation économétrique ou conditionnelle. Cette opposition n'est en fait qu'apparente car on montrera comment l'une peut se déduire de l'autre au moyen d'hypothèses spécifiques.

Même dans un contexte dynamique cette opposition entre les deux approches de modélisation (séries temporelles et modèles économétriques) n'est pas si forte que l'on veut bien le dire car les deux types d'approches s'intéressent aux caractéristiques dynamiques des processus. Ainsi dans un AR(1) tout simple, le coefficient de régression de y_{t-1} mesure l'inertie ou la mémoire du système. Dans un modèle de régression, on retrouve ces mêmes notions, mais cette fois-ci conditionnellement aux variables exogènes et l'on s'intéressera à des retards moyens de réaction de y_t par rapport à une variation des exogènes.

Parler de l'inertie d'un système conduit naturellement à se poser la question de la persistance des chocs. Cette notion sera traitée de façon détaillée dans le dernier chapitre. Mais remarquons déjà que dans tout modèle stationnaire, les chocs ont tendance à s'amortir et à l'infini leur impact est nul. Dans les chapitres suivants, nous lèverons cette hypothèse de stationnarité et nous verrons alors comment les propriétés dynamiques des modèles et statistiques des estimateurs en sont bouleversées.

12 Lectures additionnelles

On pourra consulter avec profit l'ouvrage de Harvey (1981) ainsi que le survey de Hendry, Pagan, and Sargan (1984). Pour les tests de spécification, Godfreys (1988) présente bien l'approche des test de Lagrange. Des tests sont également présentés dans Lütkepohl and Krätzig (2004).

13 Exercices

13.1 Davidson, Hendry, Srba, Yeo (1978): Examen 2006

Cette première série d'exercices demande de lire l'article de Davidson, Hendry, Srba, Yeo (1978): *The Economic Journal* 88, pages 661-672. La lecture du présent chapitre est aussi essentielle bien sûr.

1. Quelles sont les variables que l'on trouve dans une fonction de consommation classique? Que dit la théorie économique et que ne dit-elle pas en terme de solution de long terme et d'ajustement dynamique?
2. Les données utilisées dans cet article sont trimestrielles. Pourquoi les auteurs considèrent-ils une différenciation à l'ordre 4 et quels bénéfices en tirent-ils?
3. Considérez l'équation

$$x_t = \beta_1 w_t + \beta_2 w_{t-4} + \beta_3 x_{t-4} + v_t$$

Différenciez à l'ordre 4 cette équation en appliquant le filtre $(1 - L^4)$. Repartez du modèle initial et maintenant imposez les contraintes $\beta_1 = -\beta_2$ et $\beta_3 = 1$. Factorisez ce qui peut l'être et comparez ces deux équations (l'équation différenciée et l'équation avec contraintes). Que pouvez vous conclure et quelles sont les différences?

4. Le modèle de départ d'une fonction de consommation en niveau est une relation de la forme

$$C_t = K Y_t$$

Quelles sont les opérations qu'il faut effectuer pour obtenir une relation économétriquement estimable (logarithmes, dynamique,...). Exprimez le modèle économétrique sous forme à correction d'erreurs. Quelle est la solution de long terme?

13.2 Exercice 2: Examen 2007

1. Considérez le modèle ARE suivant:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$$

Écrivez ce modèle sous la forme d'un modèle à correction d'erreurs. Quelle restriction doit-on imposer pour que l'élasticité de long terme soit égale à 1. Pour que l'on puisse parler commodément d'élasticité faut-il que les variables soient en niveau ou en logarithme?

2. On a estimé l'équation de demande de monnaie suivante

$$\Delta LMP = - \underset{[-2.16]}{0.26} LMP1 + \underset{[3.52]}{0.94} \Delta LYP + \underset{[2.43]}{0.23} LYP1 - \underset{[-4.60]}{1.30} \Delta LR - \underset{[-3.08]}{1.21} LR1 - \underset{[-2.42]}{1.17}$$

$$R^2 = 0.69 \quad \hat{\sigma} = 0.023 \quad ss = 0.0119$$

$$\chi^2(1)_{\text{corr}} = 1.28 \quad \chi^2(2)_{\text{norm}} = 0.88 \quad \chi^2(2)_{\text{arch}} = 1.45$$

Donnez la solution de long terme de ce modèle. Quelle est l'élasticité de la demande de monnaie par rapport au revenu? Comment testeriez vous la restriction que cette élasticité est égale à un. Donnez deux méthodes pour cela.

13.3 Exercice 3

Soit le polynôme $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_r L^r$ et sa factorisation polynômiale

$$A(L) = A(1)L + (1 - L)(1 - A^*(L))$$

où le polynôme $A(L)$ est de degré r et le polynôme $A^*(L)$ de degré $r - 1$ sans terme constant. Posez $r = 2$ et écrivez la factorisation correspondante. Trouvez la valeur des coefficients de $A^*(L)$ en fonction des coefficients de $A(L)$.

Considérez maintenant le polynôme de retards $B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \dots + \beta_s L^s$ et sa factorisation

$$B(L) = B(1)L + (1 - L)(\beta_0 + B^*(L))$$

Posez $s = 2$ et faites la même opération que précédemment. Appliquez vos résultats au modèle

$$A(L)y_t = B(L)x_t + \epsilon_t$$

et écrivez le développement de la forme à correction d'erreurs.

13.4 Exercice 4

Soit le modèle de demande de monnaie suivant présentant une trappe à liquidité

$$\log m_t/p_t = \log y_t/p_t \beta_1 + \beta_2/(r_t - \gamma) + \epsilon_t.$$

- Ecrivez le test de Lagrange de l'hypothèse nulle $H_0 : \gamma = 0$.
- Développez le test de la forme fonctionnelle au moyen d'un développement de Taylor autour de H_0 .
- **Intuition** : regardez la page 172 de Harvey (1981). Pensez à remplacer les variables économiques par des symboles plus commodes pour les calculs.

13.5 Exercice 5

Considérez la relation usuelle du CAPM qui relie les rendements R_{t+1} au risque V_t :

$$R_{t+1} = \mu + \gamma V_t + u_t :$$

- Quelle est l'interprétation du coefficient γ ?
- Comment peut-on estimer la volatilité V_t ?
- Quel intérêt a-t-on à utiliser des données de fréquence d'observation différentes pour estimer cette relation?
- Comment formuleriez vous le modèle à retards échelonnés correspondant?

References

- AKAIKE, H. (1974): "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.
- ALMON, S. (1965): "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica*, 33, 178–196.
- BAI, J., AND S. NG (2005): "Tests for skewness, kurtosis, and normality for time series data," *Journal of Business and Economic Statistics*, 23(1), 49–60.
- BARTEN, A., G. D'ALCANTARA, AND G. CARRIN (1976): "COMET, A Medium Term Macroeconomic Model for the European Economic Community," *European Economic Review*, 7, 63–115.
- BREUCH, T., AND A. R. PAGAN (1979): "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287–1294.
- CAGAN, P. (1956): "The Monetary Dynamics of Hyperinflation," in *Studies in the Quantity Theory of Money*, ed. by M. Friedman. Chicago.

- CHOW, G. C. (1960): "Tests of Equality between Sets of Coefficients in two Linear Regressions," *Econometrica*, 28, 591–605.
- DAVIDSON, R., AND J. MACKINNON (1985): "The Interpretation of Tests Statistics," *Canadian Journal of Economics*, 18, 38–57.
- DURBIN, J. (1970): "Testing for Serial Correlation when in least-Squares Regression when Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410–421.
- ENGLE, R. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation.," *Econometrica*, 50(4), 987–1007.
- ENGLE, R. F., AND C. W. GRANGER (1987): "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.
- ERICSSON, NEIL, R., AND J. CAMPOS (1990): "PC GIVE and David Hendry's Econometric Methodology," *Revista de Econometria*, 10, 7–117.
- FLORENS, J.-P., AND M. MOUCAHRT (1985): "Conditioning in Dynamic Models," *Journal of Time Series Analysis*, 6, 15–34.
- GHYSEL, E., P. SANTA-CLARA, AND R. VALKANOV (2005): "There is a risk-return trade-off after all," *Journal of Financial Economics*, 76, 509–548.
- GILES, J., AND D. GILES (1993): "Pre-Test Estimation and Testing in Econometrics: Recent Developments," *Journal of Economic Surveys*, 7, 145–197.
- GODFREYS, L. (1988): *Misspecification Tests in Econometrics*. Cambridge, Econometric Society Monographs.
- GOLDFELD, S., AND R. E. QUANDT (1965): "Some Tests for Homoskedasticity," *Journal of the American Statistical Association*, 60, 539–547.
- GRANGER, C. W., AND P. NEWBOLD (1974): "Spurious Regression in Econometrics," *Journal of Econometrics*, 26, 1045–1066.
- GRILICHES, Z. (1967): "Distributed lags: a Survey," *Econometrica*, 35, 16–49.
- HARVEY, A. C. (1981): *The Econometric Analysis of Time Series*. Philip Allan, Oxford.
- HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–1271.
- HENDRY, D., AND G. MIZON (1978): "Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England," *Economic Journal*, 88, 549–563.
- HENDRY, D. F. (1995): *Dynamic Econometrics*, Advanced Texts in Econometrics. Oxford University Press, Oxford.

- HENDRY, D. F., A. R. PAGAN, AND D. SARGAN (1984): "Dynamic Specification," in *The Handbook of Econometrics*, ed. by Z. Griliches, and M. Intriligator, p. chap. 18. Amsterdam.
- HENDRY, D. F., AND J.-F. RICHARD (1982): "On the Formulation of Empirical Models in Econometrics," *Journal of Econometrics*, 20, 3–33.
- (1983): "The Econometric Analysis of Economic Time Series," *International Statistical Review*, 51, 111–163.
- JARQUE, C., AND A. BERA (1980): "Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals," *Economic Letters*, 6, 255–259.
- LÜTKEPOHL, H., AND M. KRÄTZIG (eds.) (2004): *Applied Time Series Econometrics, Themes in Modern Econometrics*. Cambridge University Press, Cambridge.
- MERTON, R. (1973): "An intertemporal capital asset pricing model," *Econometrica*, 41(5), 867–887.
- MUTH, J. F. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29, 315–335.
- NERLOVE, M. (1958): *The Dynamics of Supply: Estimation of Farmers' Response to Price*. Baltimore.
- PAGAN, A. (1984): "Model Evaluation by Variable Addition," in *Econometrics and Quantitative Economics*, ed. by D. F. Hendry, and K. F. Wallis, pp. 103–133. Oxford.
- (1987): "Three Econometric Methodologies: a Critical Appraisal," *Journal of Economic Surveys*, 1, 3–24.
- RAMSEY, J. (1969): "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society, B* 31, 350–371.
- SARGAN, D. (1964): "Wages and Prices in the United Kingdom: a Study in Econometric Methodology," in *Econometrics and Quantitative Economics*, ed. by D. F. Hendry, and K. F. Wallis. Oxford.
- SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.
- SPANOS, A. (1986): *Statistical Foundations of Econometric Modelling*. Cambridge.
- THURSBY, J., AND P. SCHMIDT (1977): "Some Properties of Tests for Specification Error in a Linear Regression Model," *Journal of the American Statistical Association*, 72, 635–641.
- VERBEEK, M. (2000): *A Guide to Modern Econometrics*. John Wiley and Sons.