

# Séries temporelles appliquées

Yves ARAGON  
aragon@cict.fr

Janvier 2008

# Table des matières

Références bibliographiques . . . . .	4
<b>1 Préliminaires sur les séries temporelles</b>	<b>5</b>
1.1 Exemples de séries temporelles . . . . .	5
1.2 Objectifs de l'analyse d'une série temporelle . . . . .	7
1.3 Tendance, Saisonnalité, Résidus . . . . .	8
1.4 Notations . . . . .	10
1.5 Un exemple : le niveau du lac Huron . . . . .	13
<b>2 ARMA - SARMA</b>	<b>19</b>
2.1 Propriétés de base des séries stationnaires . . . . .	19
2.1.1 Stationnarité . . . . .	19
2.1.2 Fonction d'autocovariance . . . . .	20
2.1.3 Qualité de l'erreur . . . . .	21
2.2 Modèles de séries stationnaires . . . . .	24
2.2.1 Série linéaire . . . . .	24
2.2.2 Processus gaussien . . . . .	25
2.3 Processus autorégressif . . . . .	25
2.3.1 Processus autorégressif d'ordre 1 . . . . .	25
2.3.2 Fonction d'autocorrélation partielle . . . . .	28
2.4 Processus Moyenne mobile . . . . .	28
2.5 Saisonnalité . . . . .	31
2.5.1 Saisonnalité multiplicative . . . . .	31
2.6 Modèles ARIMA et SARIMA . . . . .	31
2.7 Complément : non stationnarité en variance . . . . .	32
2.8 Etude empirique d'une série par une modélisation ARIMA . . . . .	33
2.8.1 La méthode MINIC . . . . .	34
2.9 Mise en pratique dans SAS . . . . .	36
2.9.1 Modélisation du niveau du Lac Huron . . . . .	39
2.10 Estimation d'un ARMA . . . . .	47
2.10.1 Fonction de vraisemblance d'un processus gaussien AR(1) . . . . .	47
2.10.2 Fonction de vraisemblance d'un processus gaussien MA(1) . . . . .	48
2.11 Prévion d'une série temporelle . . . . .	49
2.11.1 Prévion à erreur quadratique minimum pour un ARMA . . . . .	49
2.11.2 Exemple . . . . .	50
2.12 Exemple d'équations de Yule-Walker : cas de l'ARMA(2,1) . . . . .	51
<b>3 Processus non stationnaires</b>	<b>56</b>
3.1 Le problème de la racine unité . . . . .	56
3.1.1 Illustration . . . . .	57
3.2 Tests de Dickey-Fuller . . . . .	58

3.3	Tests de Dickey-Fuller augmentés (ADF)	59
3.3.1	Représentation en correction d'erreur ECM	59
3.3.2	Mise en oeuvre des tests de Dickey-Fuller augmentés	60
3.3.3	Exemples simulés	61
3.3.4	Séries réelles	63
3.4	Test d'une nonstationnarité saisonnière	69
3.5	Annexe	71
<b>4</b>	<b>Processus VAR et VARMA</b>	<b>74</b>
4.1	Introduction. Exemples	74
4.2	Série vectorielle stationnaire	75
4.2.1	Estimation de la fonction de corrélation croisée	76
4.3	Exemples de processus vectoriels stationnaires	78
4.3.1	Bruit blanc multidimensionnel	78
4.3.2	Processus VAR	79
4.3.3	Processus VMA	81
4.3.4	Processus VARMA	81
4.4	Représentation $MA(\infty)$ et réponse impulsionnelle	82
4.5	Exemple d'inférence sur un VAR	83
4.5.1	Données et analyse préliminaire	83
4.5.2	Identification	85
4.5.3	Estimation	86
4.6	Simulation	90
<b>5</b>	<b>Coïntégration</b>	<b>92</b>
5.1	Notions sur la régression d'une série sur une autre série	93
5.1.1	Exemple	93
5.2	Non stationnarité d'un VAR	95
5.2.1	Intuition de la coïntégration	95
5.2.2	Représentation en correction d'erreur	96
5.3	Coïntégration	97
5.3.1	Coïntégration avec terme déterministe	98
5.3.2	Traitement de la coïntégration par Johansen	100
5.4	Coïntégration dans SAS, proc varmax	100
5.4.1	Exemples SAS	101
5.5	Test de causalité au sens de Granger	106
5.5.1	Simulations	106
5.6	Exemples de syntaxe de la documentation SAS	110

# Bibliographie

- [1] Box G., Jenkins, Reinsel G.C. *Time Series*, , 3rd Ed. Prentice Hall.
- [2] Bourbonnais R. *Econométrie*, 3ème éd., 2000, Dunod.
- [3] Bresson G. et Pirotte A. *Econométrie des séries temporelles*, 1995 PUF.
- [4] Brockwell P.J., Davis R.A. *Introduction to Time Series and Forecasting* , 1997, Springer.
- [5] Brockwell P.J., Davis R.A. *Time Series : Theory and Methods*, 2nd edition , 1991, Springer.
- [6] Bourbonnais R., Terraza M. *Analyse des séries temporelles en économie*, 1998, PUF.
- [7] Charemza W., Deadman D.F. *New Directions in Econometric Practice*, 1992 Edward Elgar.
- [8] Chan N. H. *Time Series — Applications to finance*, 2002, Wiley.
- [9] Destandau S. et Le Guen M. *Analyse Exploratoire des Données avec SAS/INSIGHT*, 1998, INSEE Guides 7-8 .
- [10] Franses P. H. *Time series models for business and economic forecasting*, 1998, Cambridge University Press.
- [11] Gouriéroux C., Monfort A. *Séries temporelles et modèles dynamiques*, 1995, 2nd edition, Economica.
- [12] Gouriéroux C., Scaillet, Szafarz *Econometrie de la finance. Analyses historiques*, 1997, Economica.
- [13] Hamilton J.D. *Time Series Analysis*, 1994, Princeton University Press.
- [14] Jaffe J.A. *Mastering the SAS System* 2nd ed., 1996, ITP.
- [15] Kennedy P., *A guide to Econometrics*, Fifth Edition, 2003, Blackwell.
- [16] Lardic S., Mignon, V. *Econométrie des séries temporelles macroéconomiques et financières*, 2002, Economica.
- [17] Lütkepohl H., *New Introduction to Multiple Time Series*, 2005, Springer.
- [18] Mills T. C., *The Econometric Modelling of Financial Time Series* 2nd ed., 1999, Cambridge.
- [19] Reinsel G. C. *Elements of Multivariate Time Series Analysis*, 1997, Springer.
- [20] Rydberg T., Realistic statistical modelling of financial data. *International Statistical Review*, 2000, **68**, 233-258.
- [21] SAS Institute, *SAS/ETS User's Guide* Version 6 2nd ed., SAS Institute.
- [22] Shumway R., Stöffel *Time Series Analysis and its Applications*, 2000, Springer.
- [23] Tassi Ph. *Méthodes statistiques*, 1989 et nombreuses éditions, Economica.
- [24] Tsay R. S. *Analysis of Financial Time Series*, 2002, Wiley.
- [25] Wei *Time Series Analysis : Univariate and Multivariate Methods*, 1990, Addison-Wesley.
- [26] Zivot E., J. Wang *Modeling financial time series with S-Plus*, 2nd edition, 2006, Springer.

# Chapitre 1

## Préliminaires sur les séries temporelles

**Introduction** Une série temporelle univariée est une suite de données quantitatives mesurées à intervalles réguliers. Dans ce chapitre on donne quelques exemples de séries temporelles et de questions concrètes qu'on peut se poser devant une telle série. On verra notamment qu'une même série peut servir à différents usages et donc recevoir des traitements variés. Ensuite, on introduit une série de référence : le bruit blanc (BB) et l'opérateur retard qui simplifie grandement l'écriture des problèmes de séries temporelles. Enfin un exemple, le niveau du lac Huron, illustre ce que peut apporter la prise en compte de la dynamique de l'erreur dans une régression d'une variable  $y$  sur des variables  $x$  quand les observations sont indicées par le temps. Ce problème est très classique. Vous devez avoir présent à l'esprit que l'équation 1.7 est un cas particulier de 1.2 avec erreur dynamique.

### 1.1 Exemples de séries temporelles

Une série temporelle est une suite d'observations indicées par le temps, la date à laquelle l'observation est faite est une information importante sur le phénomène observé.

Examinons le chronogramme de quelques séries <sup>1</sup>et notons leurs caractéristiques les plus évidentes.

- 1 La population de la France (Fig. 1.1) et la population des Etats-Unis (Fig. 1.2) sont des séries où le temps explique bien le niveau de la série. Une fonction du temps assez lisse capte une grande part de la variabilité de la série.
  - ☐ Les démographes sont intéressés par la prévision de la population à 10 ans, à 20 ans.
- 2 Le nombre mensuel de morts et blessés graves par accident de la route en Grande-Bretagne (Fig. 1.5) est une série où la saisonnalité <sup>2</sup> contribue à expliquer le niveau. Le niveau moyen reste stable jusqu'à la fin de 1982 et il y a d'importantes fluctuations saisonnières. En février 1983 une nouvelle législation rend obligatoire le port de la ceinture de sécurité.
  - ☐ Un service de santé publique peut vouloir prédire le nombre de morts chacun des 6 prochains mois pour voir où et quand faire une campagne de prévention, mais il peut aussi vouloir une vision synthétique de la situation, un aperçu de la tendance sur l'année.
- 3 Les ventes de champagne (en millions de bouteilles) (Fig. 1.3) et les ventes de vin australien (Fig. 1.4) montrent une saisonnalité qui contribue à expliquer le niveau mais en plus la moyenne et la variabilité de ces séries augmentent avec le temps. Ce sont des séries *hétéroscédastiques* (c'est-à-dire à variance non constante) dont la variance à une date est fonction de la moyenne à cette date.
  - ☐ La chambre de commerce d'une région viticole peut être intéressée par l'examen de la tendance des ventes débarrassée de ce qui peut se passer à court-terme alors qu'un syndicat de transporteurs a besoin de savoir combien de bouteilles devront partir de la région le mois prochain. La chambre de commerce a besoin d'une représentation lissée de la série alors que le transporteur a besoin

<sup>1</sup>On appelle chronogramme d'une série, son graphique, avec le temps en abscisse.

<sup>2</sup>Le terme "saisonnalité" désigne un aspect périodique plus ou moins marqué dans certaines séries. Dans ce chapitre on ne cherche pas à lui donner un sens précis, mais seulement à l'identifier graphiquement.

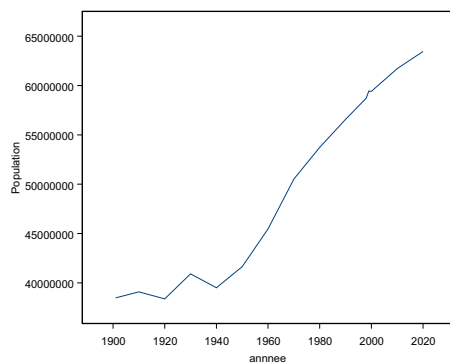


FIG. 1.1 – Population française

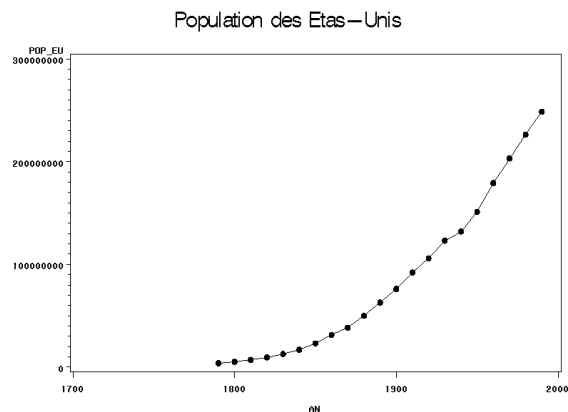


FIG. 1.2 – Population des Etats-Unis

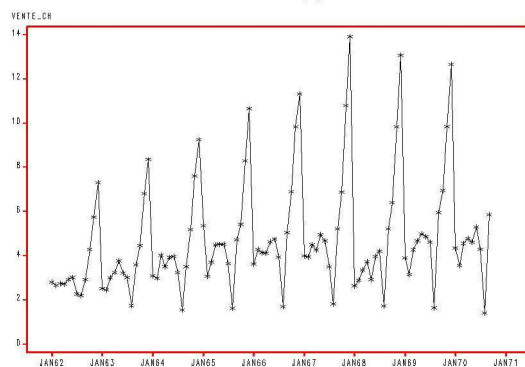


FIG. 1.3 – Ventes de bouteilles de champagne

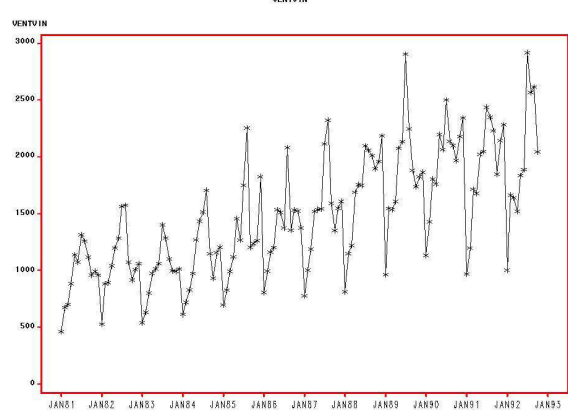


FIG. 1.4 – Ventes de vin australien

d'une prévision à un ou deux mois de la série, prévision qui, elle, devra évidemment incorporer les fluctuations saisonnières. La série lissée par la méthode X11 est représentée figure ( 1.6).

- 4 Le niveau du lac Huron (1.8) montre une tendance légèrement décroissante et on observe que les résidus de l'ajustement d'une droite au niveau, à deux dates consécutives, sont corrélés positivement (Fig. 1.15).

Si l'objectif est de prédire le niveau l'an prochain, il faut exploiter à la fois la linéarité de la tendance et la corrélation de la série à 2 dates consécutives.

- 5 Le rendement du titre Alcatel (1.7) est une série hétéroscédastique, c'est-à-dire une série dont la variance n'est pas constante au cours du temps, dont la variabilité change au cours du temps de façon non évidente. Le rendement d'une action est très souvent de moyenne nulle et les rendements à 2 dates consécutives souvent non corrélés. Il n'y a donc pas de problème de prévision du rendement mais la prévision de sa variabilité est utile.

On voit sur ces exemples qu'on ne peut pas considérer une série temporelle comme une suite d'observations indépendantes. Par conséquent tous les calculs et raisonnements de statistique mathématique basés sur l'indépendance entre les observations doivent être revus et si nécessaire, pensés autrement. Mais la connaissance de la dépendance temporelle permet en contrepartie d'améliorer la prévision de la valeur d'une série à une date future connaissant le présent et le passé.

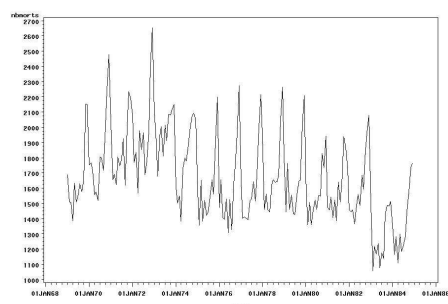


FIG. 1.5 – Nombre de morts par accident au Royaume-Uni

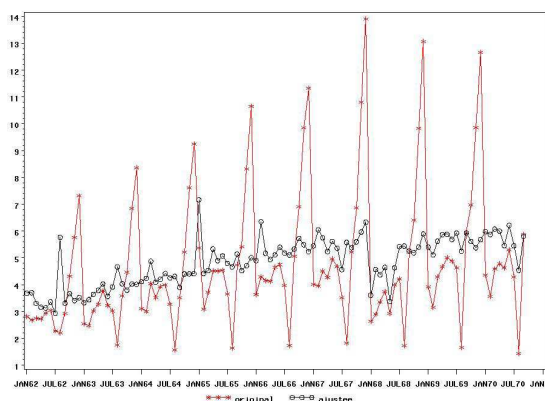


FIG. 1.6 – Série des ventes de champagne brute et lissée par X11

## 1.2 Objectifs de l'analyse d'une série temporelle

**Décrire** Quand on s'intéresse à une série temporelle, la première étape, comme pour toutes données est de décrire la série. On utilise pour ce faire un certain nombre de graphiques :

- diagramme séquentiel ou chronogramme (time plot)
- histogramme pour avoir une idée de la distribution des valeurs
- boxplot pour repérer les valeurs atypiques
- histogramme des valeurs absolues pour apprécier l'hétéroscédasticité éventuelle (notamment pour les séries de rendements de titres)
- Lag plot (diagramme retardé), c'est-à-dire diagramme de dispersion avec en abscisse la série retardée de 1, 2 ou  $k$  instants et en ordonnée la série initiale, pour comprendre la dépendance de la série par rapport à son passé. La figure (1.16) en est un exemple. La généralisation de cette représentation à plusieurs décalages consécutifs est illustrée par la figure (1.13), obtenue à l'aide de la fonction `lagplot()` de R.
- month plot. Il est intéressant pour des données présentant une saisonnalité. La figure (1.9) représente le month plot du trafic passager à Blagnac de 1993 à 2006. Il regroupe les 12 séries de trafic mensuel : une série par mois, de janvier à décembre et un point par année d'observation. Le décrochement correspond aux attentats du 11 septembre 2001.

On calcule aussi les statistiques descriptives usuelles : moyenne, variance, coefficients d'aplatissement et d'asymétrie.

Comme on l'a noté sur les exemples, une même série temporelle peut être analysée de différentes façons suivant l'objectif poursuivi.

**Résumer** Dans certains cas on veut une vue synthétique débarrassée de détails de court-terme ; c'est souvent un besoin des instituts de statistique officielle. Les méthodes qui donnent une vue d'ensemble de la série nette de fluctuations de court terme sont appelées méthodes de lissage.

**Modéliser** Expliquer le niveau ou parfois la variance du niveau, par des modèles à peu de paramètres

- Modèle sans variable explicative

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + u_t \quad (1.1)$$

- Modèle avec variable explicative

$$Y_t = f(X_t) + u_t \quad (1.2)$$

où  $u_t$  est une erreur

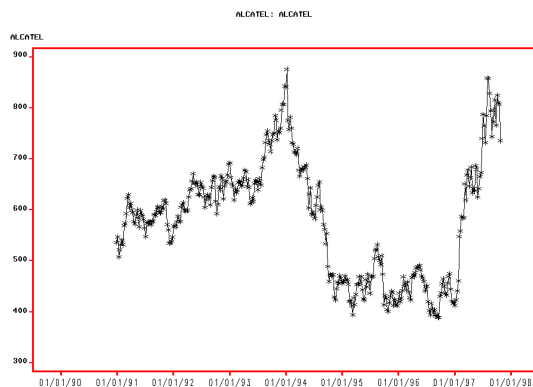


FIG. 1.7 – Cours de l'action Alcatel

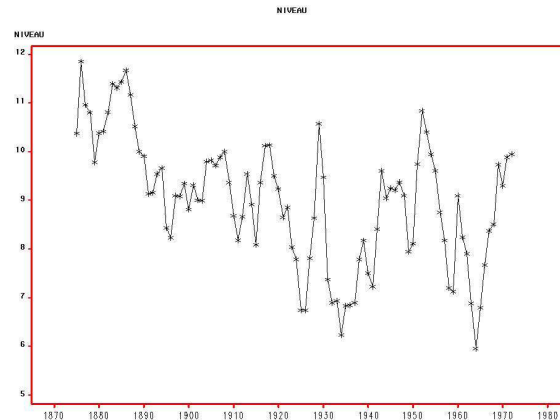


FIG. 1.8 – Niveau du lac Huron

- modèle statique :  $X_t$  ne contient pas de valeurs passées de  $\{Y_t\}$  et les  $u_t$  sont non corrélés entre eux
- modèle dynamique :  $X_t$  contient des valeurs retardées de  $y_t$  ou les  $u_t$  sont auto-corrélés. Dans ce cas la dynamique de ces erreurs doit être modélisée.

Dans tous les cas, la fonction  $f()$  doit être estimée dans une classe de fonctions. Dans ce cours on ne considère que des fonctions linéaires mais le cours de régression non et semi-paramétrique propose des classes plus vastes.

**Prédire** La prévision de valeurs à des dates futures connaissant le présent et le passé de la série peut être basée sur un modèle, ce que nous pratiquerons à travers les modèles ARMA, ARMAX... ou bien être construite sans référence à un modèle, cas du lissage exponentiel et des ses généralisations.

L'essentiel de ce cours est consacré à la modélisation. Mais un modèle ne remplace pas un graphique. Il faut toujours essayer de comprendre un phénomène par plusieurs approches : modèle et graphiques, plusieurs modèles...

### 1.3 Tendence. Saisonnalité. Résidus

Il est classique de décomposer une série temporelle  $\{Y_t, t = 1, \dots, n\}$  en tendance  $m_t$  (trend), effet saisonnier  $s_t$ , erreur  $U_t$ . On s'intéresse habituellement à un modèle additif :

$$y_t = m_t + s_t + U_t, \quad \text{où } E(U_t) = 0 \quad (*)$$

ou à un modèle multiplicatif :

$$y_t = m_t s_t U_t, \quad \text{où } E(U_t) = 1 \quad (**)$$

Par exemple, les séries montrant une saisonnalité qui a de plus en plus d'ampleur (cas des ventes de champagne, Fig. 1.3) sont souvent mieux ajustées par un modèle multiplicatif que par un modèle additif.

#### Commentaires



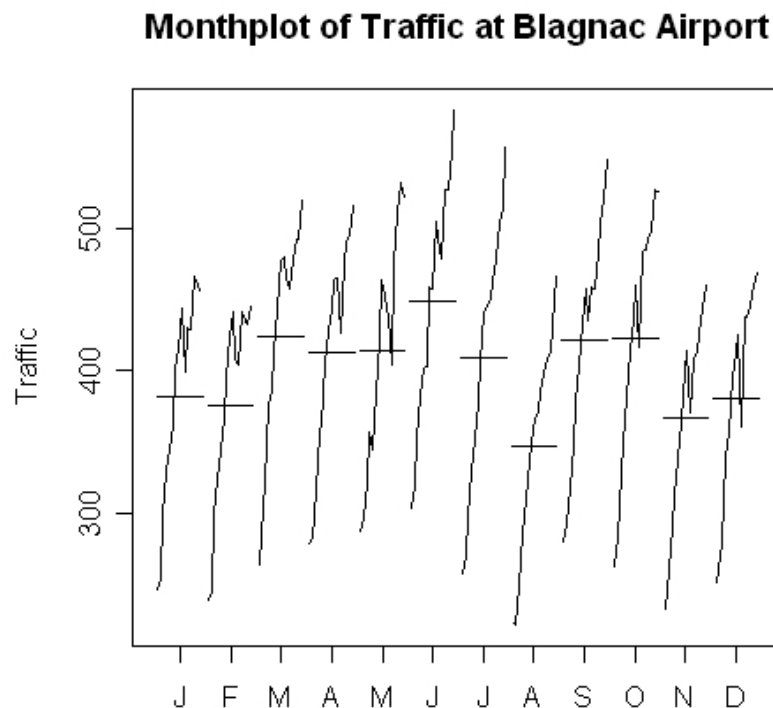


FIG. 1.9 – Month plot du trafic de l'aéroport de Toulouse-Blagnac

- Une décomposition permettant d'éliminer l'effet saisonnier est utile quand on veut examiner le comportement à moyen terme d'une série. Les services de statistique officielle, les organismes qui étudient les grandes tendances d'un phénomène, travaillent sur des séries désaisonnalisées. A titre d'exemple on consultera le site de la Sécurité routière :  
<http://www.securiteroutiere.gouv.fr/infos-ref/observatoire/conjoncture/> puis "Les méthodes de production de l'analyse conjoncturelle".
- Par contre quand c'est le comportement à court terme qui intéresse, il faut garder l'aspect saisonnier. Les gestionnaires de stocks de magasins doivent tenir compte de la saisonnalité des ventes de leurs produits quand ils passent leurs commandes.
- Le choix d'un modèle ou d'un autre, l'incorporation ou non d'une composante, peuvent s'apprécier d'après le graphique de la série et peuvent être validés par l'analyse elle-même.
- Il n'y a pas une unique façon d'obtenir une décomposition telle que (\*) ou (\*\*) pour une série particulière.

Dans R, les fonctions : `decompose()`, `stl()`, `StructTS()` donnent des décompositions saisonnières d'une série.

Dans SAS, les `proc X11` et `X12`, sont des programmations des méthodologies X11 et X12. Elles donnent notamment des outils de lissage et de désaisonnalisation. Elles sont reconnues et appliquées par les services de statistique officielle de la plupart des pays.

On considère qu'une série temporelle observée  $\{x_t, t = 1, \dots, T\}$  est la réalisation d'un processus aléatoire  $\{X_t, t = 1, \dots, T\}$ . Les  $X_t$  sont des v.a. (variables aléatoires),  $x_t$  est la valeur prise par  $X_t$ , ce qui advient à la date (ou à l'instant d'observation)  $t$ , ces v.a. à des dates différentes sont normalement

corrélées. Un cas limite est celui où ces v.a. sont i.i.d..

### Bruit blanc (BB)

**Définition 1** Un bruit blanc  $\{Z_t\}$  est une suite de v.a. non corrélées de moyenne nulle, de variance constante  $\sigma_Z^2$ .

On note  $\{Z_t\} \sim BB(0, \sigma_Z^2)$ .  $\{Z_t\} \sim BBN(0, \sigma_Z^2)$  désigne un bruit blanc gaussien, c'est-à-dire une suite de v.a. i.i.d.  $N(0, \sigma_Z^2)$ .

Le BB est une série de référence. Modéliser une série revient souvent à trouver les opérations qui la décrivent comme transformation d'un bruit blanc.

**Étapes** de l'étude d'une série sans série explicative associée.

- On commence par enlever, si nécessaire, les effets systématiques (par exemple, le nombre de jours ouvrables dans une série hebdomadaire d'une production, l'effet "jour de la semaine" dans une série quotidienne), les effets occasionnels (grèves, panne de machine).
- On impute des valeurs aux dates où manque l'observation de la série.
- Soit  $x_t, t = 1, \dots, n$ , la série obtenue après ces corrections. Si elle présente une tendance et une saisonnalité, il faut les estimer ou les éliminer suivant l'objectif recherché.
- Enfin on prédit chaque composante de la série puis la série.

Notons qu'il y a souvent plusieurs méthodes possibles pour réaliser ces étapes. De plus certaines méthodes comme celle Box-Jenkins, peuvent modéliser la saisonnalité mais ne peuvent pas isoler une composante saisonnière.

## 1.4 Notations

**Opérateur retard.** On note B(backwards) ou L (lag), l'opérateur qui fait passer de  $x_t$  à  $x_{t-1}$  :

$$Bx_t = x_{t-1}$$

On a :

$$B^2x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$$

**Opérateur différence.** La différence première est :

$$\Delta x_t = (1 - B)x_t = x_t - x_{t-1}$$

on a :

$$\Delta^2 x_t = \Delta(\Delta x_t) = x_t - 2x_{t-1} + x_{t-2}$$

**Opérateur différence saisonnière.** Pour une série mensuelle il peut être utile d'examiner ses accroissements d'une année sur l'autre (janvier sur janvier, ...). L'opérateur différence saisonnière  $\Delta_{12} = 1 - B^{12}$  est utile dans ce cas.

$$\Delta_{12}x_t = (1 - B^{12})x_t = x_t - x_{t-12}$$

Ces notations simplifient l'écriture des équations relatives aux séries.

### Exemples

(a) Manipulation de l'opérateur retard.

Pour la fonction constante  $= 1 \forall t$ ,

$$(1 - 0.9B)1 = 1 - 0.9B1 = 1 - 0.9 = 0.1$$

de même

$$\frac{1}{1-0.9B}1 = (1 + 0.9B + (0.9)^2B^2 + \dots)1 = 1 + 0.9 + (0.9)^2 + \dots = \frac{1}{1-0.9} = 10.$$

Pour la fonction  $= t \forall t$

$$(1 - 0.9B)t = t - 0.9Bt = t - 0.9(t - 1) = 0.1t + 0.9$$

de même

$$\begin{aligned} \frac{1}{1-0.9B}t &= (1 + 0.9B + (0.9)^2B^2 + \dots)t = t + 0.9(t-1) + (0.9)^2(t-2) + \dots = \\ &= t \frac{1}{1-0.9} - 0.9 \frac{1}{(1-0.9)^2} = \dots = 10t - 90. \quad (1.3) \end{aligned}$$

**(b) Calcul de séries retardée et de séries différenciées dans SAS.**

```
data b;
  set donn.champa2;
  ven1 = lag(vente_ch);
  ven12 = lag12(vente_ch);
5  d1_ven = dif1(vente_ch);
  d12_ven = dif12(vente_ch);
  run;

proc print data=b (obs=14);
10 run;
```

Le Système SAS				10:00 Saturday, June 18, 2005 29			
15	Obs	VENTE_CH	DATE	ven1	ven12	d1_ven	d12_ven
	1	2.815	JAN62	.	.	.	.
	2	2.672	FEB62	2.815	.	-0.143	.
	3	2.755	MAR62	2.672	.	0.083	.
20	4	2.721	APR62	2.755	.	-0.034	.
	5	2.946	MAY62	2.721	.	0.225	.
	6	3.036	JUN62	2.946	.	0.090	.
	7	2.282	JUL62	3.036	.	-0.754	.
	8	2.212	AUG62	2.282	.	-0.070	.
25	9	2.922	SEP62	2.212	.	0.710	.
	10	4.301	OCT62	2.922	.	1.379	.
	11	5.764	NOV62	4.301	.	1.463	.
	12	7.312	DEC62	5.764	.	1.548	.
	13	2.541	JAN63	7.312	2.815	-4.771	-0.274
30	14	2.475	FEB63	2.541	2.672	-0.066	-0.197

**(c) Dans R.**

- Lecture du fichier texte champa.asc .
- Conversion en série temporelle (avec date).
- Fabrication de séries retardées. Observer que la fonction `lag()` marche à l'envers de ce qu'on attend ! On peut observer qu'elle se contente de changer la date de la série. De plus elle ne s'applique qu'à un objet de type `ts`. La fonction `Lag()` du package `Hmisc` est plus simple d'emploi.
- Illustration des fonction `ts.union`, `ts.intersection`.
- Fabrication de la série retardée de façon classique.

```
> champagne = scan('champa.asc')
Read 105 items
```

```

> champagne = ts(champagne, start=c(1962,1), frequency=12)
> chL1 = lag(champagne, -1)
5 > chL12 = lag(champagne, -12)
> dd = ts.union(champagne, chL1, chL12)
> dd
      champagne    chL1    chL12
10 Jan 1962      2.815      NA      NA
   Feb 1962      2.672      2.815      NA
   Mar 1962      2.755      2.672      NA
   Apr 1962      2.721      2.755      NA
   ...
   Oct 1962      4.301      2.922      NA
15 Nov 1962      5.764      4.301      NA
   Dec 1962      7.312      5.764      NA
   Jan 1963      2.541      7.312      2.815
   Feb 1963      2.475      2.541      2.672
   Mar 1963      3.031      2.475      2.755
20 ...
   Jul 1970      4.298      5.312      4.633
   Aug 1970      1.431      4.298      1.659
   Sep 1970      5.877      1.431      5.951
   Oct 1970      NA        5.877      6.981
25 Nov 1970      NA        NA        9.851
   Dec 1970      NA        NA        12.670
   Jan 1971      NA        NA        4.348
   Feb 1971      NA        NA        3.564
   ...
30 Aug 1971      NA        NA        1.431
   Sep 1971      NA        NA        5.877

> dd0 = ts.intersect(champagne, chL1, chL12)
> dd0
35 champagne    chL1    chL12
   Jan 1963      2.541      7.312      2.815
   Feb 1963      2.475      2.541      2.672
   Mar 1963      3.031      2.475      2.755
   ...
40 Jun 1970      5.312      4.618      4.874
   Jul 1970      4.298      5.312      4.633
   Aug 1970      1.431      4.298      1.659
   Sep 1970      5.877      1.431      5.951
>
45 > # fabrication de la série retardée, façon classique
> c.c1 = cbind(champagne, c(NA, champagne[-1]))
> dimnames(c.c1)[[2]] = c("ventes", "ventes.1")
> c.c1[1:5,]
      ventes ventes.1
50 [1,]  2.815      NA
   [2,]  2.672      2.672
   [3,]  2.755      2.755
   [4,]  2.721      2.721
   [5,]  2.946      2.946
55
> # plus simplement avec la fonction Lag de Hmisc
> require(Hmisc)
[1] TRUE
> x2 = x^2
60 > x2
   [1]  1  4  9 16 25 36 49 64 81 100
> diff(x2, 1)
   [1]  3  5  7  9 11 13 15 17 19
> diff(diff(x2, 1), 1)
65 [1] 2 2 2 2 2 2 2 2 2
> diff(x2, 1, 2)
   [1] 2 2 2 2 2 2 2 2 2
> diff(x2, 4)
   [1] 24 32 40 48 56 64
70

```

- 
- Exemple de lag plot. La série des températures mensuelles moyennes à Nottingham est fournie avec R (Fig. 1.12). Les diagrammes retardés de 1 à 12 de cette série (1.13) montrent des aspects très typés,

notamment aux décalages (ou retards) 1, 5, 6, 11, 12.

---

```
lag.plot(nottem, 12, layout = c(4,3), ask = NA, do.lines = FALSE,
        main = "Températures mensuelles moyennes à Nottingham, 1920-1939",
        col.main = "blue")
```

---

Les figures (1.10 et 1.11) représentent une série simulée suivant le mécanisme :

$$y_t = 0.9y_{t-12} + Z_t - 0.7Z_{t-1}, \quad Z_t \sim BBN(0, 1) \quad (**)$$

et son lag plot. La valeur à une date est fortement liée à la valeur 12 périodes avant. On voit sur le lag plot qu'il n'y a pas de corrélation entre des valeurs décalées de 1, 2, ..., 11.

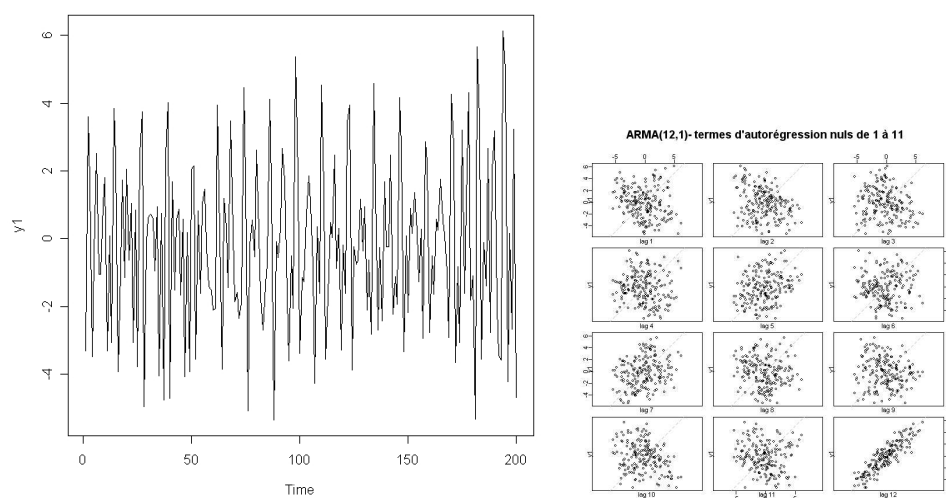


FIG. 1.10 – Chronogramme d'une série simulée FIG. 1.11 – Lag plot de la même série simulée suivant (\*\*)

**Exercice 1** Examinant la figure (1.9), (1) que peut-on dire de la saisonnalité du trafic à l'aéroport de Blagnac ? (2) Le 11 septembre a-t-il eu le même effet sur tous les mois des années suivantes (effets sur janvier, sur février, ...) ? (3) Que peut-on raconter de l'effet du 11 septembre sur la croissance de l'activité de l'aéroport ?

**Exercice 2** Lag plots. Les séries (1.10) et () ont toutes deux un fort comportement saisonnier. (1) Que traduit la forme en anneau à certains décalages pour le lag plot de la température à Nottingham ? La réponse est, pour l'instant, seulement intuitive. Il faut utiliser notre connaissance intuitive de cette série. (2) Commenter les similitudes et les différences de leurs lag plots.

## 1.5 Un exemple : le niveau du lac Huron

On a rarement à étudier une série  $y$  isolée. Souvent une autre série  $x$  non aléatoire explique une partie du niveau de  $y$ . Dans ce cas il faut :

- commencer par modéliser assez grossièrement la dépendance de la moyenne de  $y$  par rapport à  $x$ , par MCO<sup>3</sup> par exemple,
- examiner si le résidu de cette modélisation a une dynamique et, si c'est le cas, il faut également modéliser cette dynamique

---

<sup>3</sup>MCO = Moindres Carrés Ordinaires

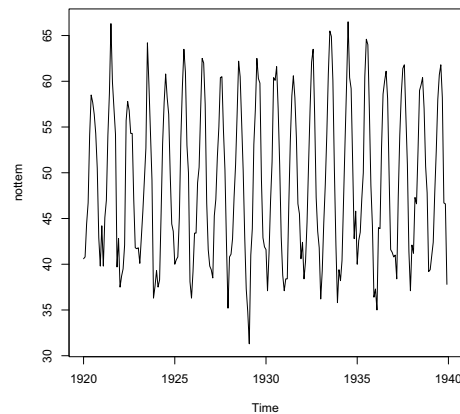


FIG. 1.12 – Température mensuelle moyenne à Nottingham

– enfin on essaie de modéliser simultanément les deux aspects : moyenne et dynamique de l’erreur. On appelle souvent un tel modèle, modèle ARMAX. On en rencontrera dans les modèles d’intervention et en TP. On en donne un petit exemple dès maintenant.

Le graphique de la série du niveau du lac Huron (1.8) montre une tendance légèrement décroissante dont la significativité doit être testée et, autour de cette tendance, une variabilité importante d’une année à l’autre. Le temps est notre variable explicative.

► La 1<sup>re</sup> étape consiste donc à faire une régression linéaire du niveau sur l’année par MCO :

$$y_t = \beta_0 + \beta_1 t + u_t, \quad t = 1, \dots, T \quad (1.4)$$

et stocker les résidus de cet ajustement.

Pour faire ce travail on a utilisé la `proc autoreg`. Comme la `proc reg` elle peut faire un ajustement par MCO mais elle dispose en plus notamment d’options pour analyser la corrélation des résidus au cours du temps (leur autocorrélation).

---

```
proc autoreg data= st6.lake;
model niveau = date;
output out= lakout residual=residmc;
run;
```

5                    Niveau du lac Huron - Residus    16:30 Sunday, June 23, 2002    9

The AUTOREG Procedure

Dependent Variable    NIVEAU

10

## Ordinary Least Squares Estimates

	SSE	122.645511	DFE	96
15	MSE	1.27756	Root MSE	1.13029
	SBC	309.266295	AIC	304.09636
	Regress R-Square	0.2725	Total R-Square	0.2725
	Durbin-Watson	0.4395		

20

	Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
	Intercept	1	8.1208	0.1864	43.57	<.0001
25	DATE	1	-0.000066	0.0000111	-6.00	<.0001

Attention ! La régression a été effectuée sur la date c'est-à-dire sur des valeurs espacées de 365 ou 366 et ne commençant pas à 1.

Le  $R^2$  n'est pas très élevé ( $R^2 = 0.2725$ ) mais la régression est significative : t Value pour  $\beta_1 = -6.00$  qui correspond à un niveau de signification empirique  $< .0001$ . Examinons maintenant le résidu de cet ajustement :  $\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 t$ . A-t-il une dynamique ?

```

title "Lac Huron - residu contre residu retarde";
proc gplot data=b;
    symbol1 v=star i= none;
5   plot resret * residmc = 1;
run;
quit;

```

On constate (graphique 1.15) que le signe du résidu est assez régulièrement alternativement positif et négatif sur plusieurs dates consécutives .

Examinons donc le diagramme de dispersion des  $(\hat{u}_{t-1}, \hat{u}_t), t = 2, \dots, T$  (graphique 1.16). Comme on pouvait s'y attendre, on observe une corrélation linéaire significative entre le résidu à une date et le résidu à la date voisine. Cette corrélation sur les résidus  $\hat{u}_t$  est le reflet d'une corrélation sur les erreurs  $u_t$  ;  $u_t$  a une dynamique.

(On comprend qu'on peut améliorer la prévision du niveau d'une année à partir de la connaissance du niveau les années antérieures en prenant également en compte de cette corrélation.) Résumé de la <sup>re</sup> étape : nous avons estimé le niveau moyen du lac et avons constaté que l'erreur est autocorrélée, qu'elle a une dynamique.

► La 2<sup>e</sup> étape consiste à modéliser la dynamique de l'erreur. Utilisant les outils du prochain chapitre, l'examen des résidus de l'ajustement par MCO suggère que l'erreur à une date  $t$  dépend linéairement des erreurs aux dates  $t - 1$  et  $t - 2$ . Autrement dit, l'erreur est un processus autorégressif d'ordre 2.

Sans entrer dans le détail de cet exemple on en donne maintenant quelques aspects graphiques. On superpose sur le graphique (1.17) : (1) la série du niveau, (2) sa prévision à l'horizon 1 *sans* tenir compte de l'autocorrélation des erreurs ( droite des moindres carrés) puis (3) en modélisant l'erreur par un processus autorégressif d'ordre 2 (AR(2)) :

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + Z_t$$

où  $Z_t$  est un BB. On peut observer une très nette amélioration de l'ajustement au prix de 2 paramètres supplémentaires.

Ecriture synthétique du modèle ( $x_t \equiv t$ ) :

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad (1.5)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + Z_t, \quad Z_t \text{ Bruit blanc} \quad (1.6)$$

Utilisant l'opérateur retard on écrit

$$(1 - \phi_1 B - \phi_2 B^2)u_t = Z_t \text{ ou } u_t = \frac{1}{1 - \phi_1 B - \phi_2 B^2} Z_t$$

et le modèle de  $y_t$  s'écrit :

$$y_t = \beta_0 + \beta_1 x_t + \frac{1}{1 - \phi_1 B - \phi_2 B^2} Z_t. \quad (1.7)$$

Enfin il faut vérifier que le modèle ajusté est satisfaisant. En particulier il faut tester la blancheur de  $Z_t$ . Un tel test sera présenté au prochain chapitre.

Résumons. Il est utile de modéliser le mécanisme d'évolution de l'erreur  $u_t$  dans cette série pour au moins deux raisons :

1. Si notre objectif est la prévision du niveau l'année suivante, on prédira  $y_{T+1}$  par :

$$y_{T+1} = \hat{\beta}_0 + \hat{\beta}_1(T+1) + \text{Pred}(u_{T+1} | \hat{u}_1, \hat{u}_2, \dots, \hat{u}_T)$$

Dans cette expression  $\text{Pred}(\cdot | \cdot)$  désigne la prévision de ce qui est avant le  $|$  sachant ce qui est après.

2. Si l'objectif est d'avoir une bonne estimation de  $\beta_1$  il faut modéliser le mécanisme de l'erreur et estimer la régression par Moindres Carrés Généralisés (MCG). On montre en effet, voir par exemple ([2], chapitre 5), que si l'erreur est autocorrélée, l'estimateur des MCO de  $\beta$  n'est pas efficace.

Nous allons diviser les difficultés : dans le chapitre suivant nous considérons des séries dont la moyenne est constante (l'étape 1 ci-dessus est donc très simple) et nous présentons les modèles ARMA. Ce sont des modèles assez généraux pour l'erreur  $u_t$  dans (1.7). En fin de chapitre nous traiterons en détail la modélisation du niveau du lac Huron comme exemple de modèle ARMAX (la moyenne de la variable dépendante est une fonction linéaire de variables explicatives  $x$  et l'erreur suit un processus ARMA).



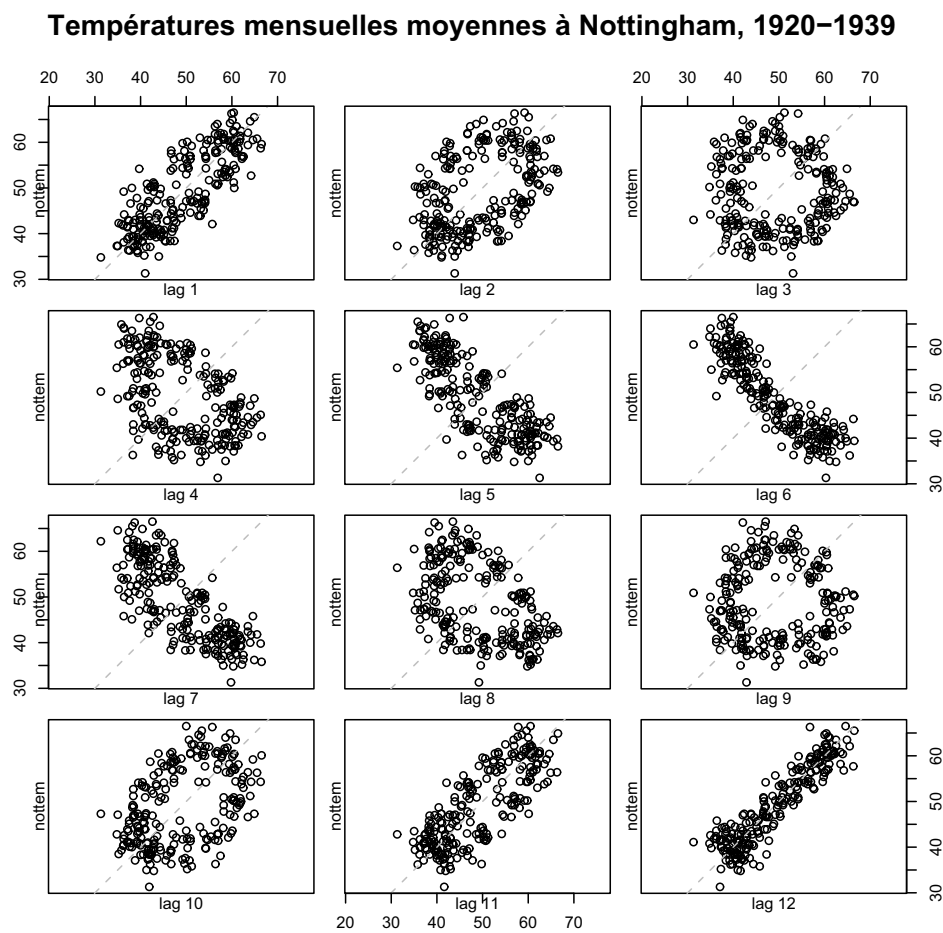


FIG. 1.13 – Température à Nottingham - Diagrammes retardés

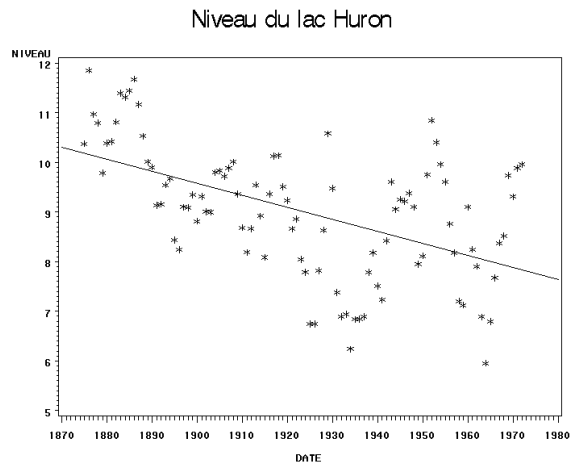


FIG. 1.14 – Niveau du lac Huron - ajustement d'un polynome de degré 1

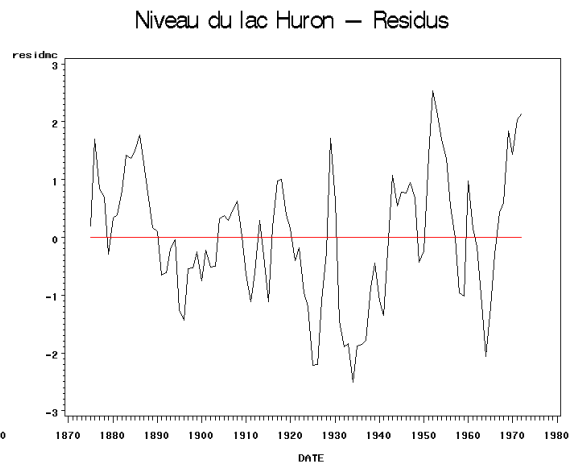


FIG. 1.15 – Niveau du lac Huron - résidus en fonction du temps

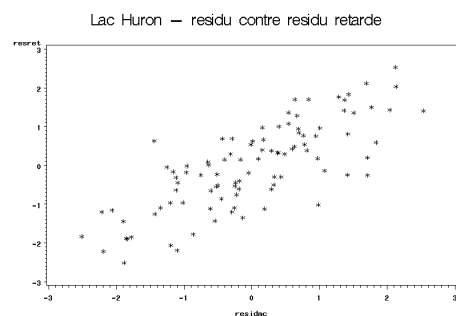


FIG. 1.16 – Lac Huron - résidu contre résidu retardé

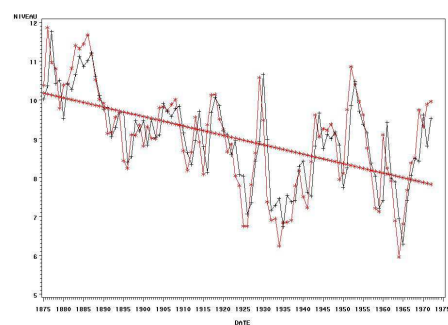


FIG. 1.17 – Lac Huron - (1) série, (2) ajustement par MCO et (3) ajustement d'une droite avec erreur AR(2)

## Chapitre 2

# ARMA - SARMA

**Introduction** Dans ce chapitre on commence par donner la définition de la stationnarité (faible). Ensuite on définit la fonction d'autocorrélation simple (ACF) d'une série stationnaire et son estimation. On voit ensuite le test du portemanteau, c'est un test de blancheur d'une série stationnaire. Il est très utile car si une série est un bruit blanc (BB), elle se traite comme une suite d'observations non corrélées de variance constante. On introduit ensuite les modèles ARMA. Ce sont des modèles qui conviennent à beaucoup de séries stationnaires de moyenne constante. Un modèle  $\text{ARMA}(p, q)$  pour une série a deux paramètres à fixer avant toute estimation :  $p$  l'ordre de l'autorégression (AR) et  $q$  l'ordre de la moyenne mobile (MA). On introduit une autre fonction décrivant l'autocorrélation d'une série stationnaire : la fonction d'autocorrélation partielle (PACF). Or la PACF d'un  $\text{AR}(p)$  est nulle à partir de l'ordre  $p + 1$  de même l'ACF d'un  $\text{MA}(q)$  est nulle à partir de l'ordre  $q + 1$ . Etant donné une série, découvrir des valeurs possibles pour  $p$  et  $q$  s'appelle *identifier* la série. Or les *fonctions d'autocorrélation empiriques*, simple et partielle, estiment correctement ces mêmes *fonctions d'autocorrélation théoriques*. Des valeurs possibles pour  $p$  et  $q$  sont donc suggérées par les fonctions d'autocorrélation empiriques. On voit enfin comment on identifie puis estime un ARMA avec SAS.

Finalement, si une série est stationnaire, les modèles ARMA sont une classe assez vaste pour essayer de modéliser la série. Pour vérifier que le modèle est correct il faut notamment vérifier que le résidu prédit est un BB. Et que fait-on avec les séries non stationnaires ? Nous verrons au chapitre suivante des tests de non stationnarité. D'autre part, si on a une série non stationnaire il faut trouver une ou des transformations qui la ramènent à une série stationnaire à laquelle on peut appliquer les techniques du présent chapitre. Ce chapitre est très important, car on retrouvera les notions qu'il contient dans différentes circonstances. Durée : 6h sans compter le temps de manipulation de séries sur logiciel.

## 2.1 Propriétés de base des séries stationnaires

### 2.1.1 Stationnarité

La stationnarité est la clef de l'analyse des séries temporelles. Une série  $\{Y_t\}$  est dite *strictement stationnaire* si la distribution conjointe de  $(Y_{t_1}, \dots, Y_{t_k})$  est identique à celle de  $(Y_{t_1+t}, \dots, Y_{t_k+t})$ , quel que soit  $t, k$  entier positif arbitraire et  $(t_1, \dots, t_k)$  liste de  $k$  entiers positifs arbitraires. Autrement dit, la stationnarité stricte dit que la distribution conjointe de  $(Y_{t_1}, \dots, Y_{t_k})$  est invariante quand on fait glisser le temps. Cette condition est difficile à vérifier et on utilise une version plus faible de stationnarité.

On dit qu'une série temporelle  $\{Y_t\}$  est *faiblement stationnaire* si la moyenne de  $Y_t$  ne dépend pas de  $t$  et si la covariance entre  $Y_t$  et  $Y_{t-l}$  ne dépend que de  $l$  et non de  $t$ . Précisément,

**Définition 2**  $\{Y_t\}$  est dite *faiblement stationnaire* si :

- $E(Y_t) = \mu_Y$  où  $\mu_Y$  est une constante indépendante de  $t$ ,
- $\text{cov}(Y_t, Y_{t-l})$  ne dépend que de  $l$ , entier et dans ce cas elle est notée :

$$\gamma_Y(l) = \text{cov}(Y_t, Y_{t-l}).$$

**Remarque.** On a distingué stationnarité stricte et stationnarité faible et la plupart des modèles que nous allons examiner concernent des séries normalement distribuées ; pour elles, les deux notions coïncident. Mais on verra que le rendement du cours d'une action est souvent faiblement stationnaire sans l'être strictement (voir le chapitre consacré aux modèles hétéroscédastiques). La suite du chapitre concerne les séries faiblement stationnaires.

**Considérations pratiques pour apprécier la stationnarité d'une série.** Il est important d'arriver à se faire une idée de la stationnarité d'une série par l'observation de son chronogramme. La stationnarité dit que la moyenne et la variance de la série sont constantes quelque soit la date. Elle implique donc que le graphe de la série en fonction du temps montre des fluctuations qui se ressemblent (en niveau et en amplitude), quelque soit la date autour de laquelle on examine la série.

A Une série stationnaire a une moyenne constante. Considérons le cours de l'action Alcatel (Figure 1.7 de l'ethème 1). Imaginons un intervalle d'une année environ et faisons glisser cet intervalle. De 1991 à 1993, le niveau moyen sur un an reste à peu près constant. On remarque la même chose de 1995 à 1997. Mais la fluctuation se fait autour d'un niveau beaucoup plus faible que pour la période précédente. Il est manifeste donc que pour cette série, la moyenne dépend de  $t$  et donc la série n'est pas stationnaire. En résumé, si le niveau d'une série fluctue peu autour d'un niveau moyen sur un petit intervalle de temps mais que ce niveau moyen change avec la position de l'intervalle, on peut conclure que la série n'est pas stationnaire.

**Question** Quelles séries de l'ethème 1 peuvent être reconnues comme non stationnaire au seul examen de leur niveau moyen ?

B Une série stationnaire a une variance constante. Ce qui veut dire que l'ampleur de la fluctuation de la série reste la même sur un petit intervalle de temps quelque soit cet intervalle. La population des Etats-Unis entre dans cette catégorie. Mais cette série n'est pas stationnaire car sa moyenne n'est pas constante.

**Question** Quelles séries de l'ethème 1 montrent une variance non constante et ne sont donc pas stationnaires ?

C On est maintenant équipé pour examiner la série des températures de l'air à Nottingham.

**Question** Le chronogramme de la série est très régulier. Chaleur en été, froid en hiver... La moyenne sur 6 mois consécutifs change selon que ces mois sont centrés sur l'été ou sur l'hiver. Régularité veut-il dire stationnarité ? Examiner également le lag plot. Comment raconter le graphe pour le décalage 3 ?

## 2.1.2 Fonction d'autocovariance

La covariance  $\gamma_l = \text{cov}(Y_t, Y_{t-l})$  est appelée autocovariance d'ordre (ou de décalage)  $l$  (lag- $l$  autocovariance). Pour chaque décalage  $l$ , il y a une autocovariance.

**Définition 3** La fonction :  $l \mapsto \gamma_l$  est la fonction d'autocovariance de  $\{Y_t\}$ .

Cette fonction a deux propriétés importantes :

### Propriété 1

(a)  $\gamma_0 = \text{var}(Y_t)$ ,

(b)  $\gamma_l = \gamma_{-l}$ , car :

$$\gamma_{-l} = \text{cov}(Y_t, Y_{t-(-l)}) = \text{cov}(Y_{t-(-l)}, Y_t) = \text{cov}(Y_{t+l}, Y_t) = \text{cov}(Y_{t+l}, Y_{(t+l)-l}) = \gamma_l$$

Autre notation. On écrit aussi  $\gamma_Y(l)$ , en particulier pour distinguer la fonction d'autocovariance d'une série  $Y$ , de celle d'une autre série.

**Fonction d'autocorrélation (ACF)** Par définition, le coefficient d'autocorrélation d'ordre  $l$  est

$$\rho_l = \frac{\text{cov}(Y_t, Y_{t-l})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t-l})}}. \quad (2.1)$$

Mais  $\text{var}(Y_{t-l}) = \text{var}(Y_t) = \gamma_0$  donc :

$$\rho_l = \frac{\text{cov}(Y_t, Y_{t-l})}{\text{var}(Y_t)} = \frac{\gamma_l}{\gamma_0}. \quad (2.2)$$

Enfin en terme d'espérance mathématique et notant que par la stationnarité :  $E(Y_t) = \mu$  indépendant de  $t$ , on a :

$$\rho_l = \frac{E[(Y_t - \mu)(Y_{t-l} - \mu)]}{E[(Y_t - \mu)^2]}. \quad (2.3)$$

$\rho_l$  est une mesure de la dépendance de la valeur  $Y$  en une date par rapport à sa valeur à une date décalée de  $l$  intervalles de temps.

**Définition 4** La fonction :

$$l \rightsquigarrow \rho_l, l = 0, 1, 2, \dots$$

est appelée fonction d'autocorrélation (théorique), **FAC** (ou **ACF** en anglais) de la série  $\{Y_t\}$ .

De la définition on voit que :  $\rho_0 = 1, -1 \leq \rho_l \leq 1$ .

Etant donné un échantillon  $y_t, t = 1, \dots, T$ , de  $\{Y_t\}$  stationnaire, notons la moyenne empirique,  $\bar{y} = \sum_{t=1}^T y_t / T$ . Le coefficient d'autocorrélation empirique d'ordre 1 est :

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (2.4)$$

Le coefficient d'autocorrélation empirique d'ordre  $l \geq 1$  est

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^T (y_t - \bar{y})(y_{t-l} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, 0 \leq l \leq T-1 \quad (2.5)$$

Sous des conditions générales, voir par exemple Brockwell et Davis ([4]),  $\hat{\rho}_l$  est un estimateur convergent de  $\rho_l$ .

**Définition 5** La fonction :

$$l \rightsquigarrow \hat{\rho}_l, l = 0, 1, 2, \dots$$

est appelée fonction d'autocorrélation empirique de la série  $\{Y_t\}$ .

### 2.1.3 Qualité de l'erreur

Tous les modèles que nous envisagerons expriment une série temporelle  $y_t$  comme la somme d'une fonction du passé ( $y_{t-1}, y_{t-2}, \dots$ ) de la série et d'une erreur bruit blanc qu'on notera  $Z_t$ . Etant donné une série observée :  $(y_1, y_2, \dots, y_T)$ , le statisticien cherche un modèle de dépendance par rapport au passé dans la classe des modèles à sa disposition. Il ajuste le modèle et se retrouve avec des résidus :  $\hat{Z}_1, \hat{Z}_2, \dots, \hat{Z}_T$ . Si le modèle a été correctement choisi, cette série doit ressembler à un bruit blanc. si ce n'est pas le cas il faut corriger le modèle. La normalité du résidu doit également être testée. Si la série n'est pas normale, les tests d'égalité d'un paramètre à une valeur particulière, les intervalles de prédiction, fournis par les logiciels ne sont pas corrects.

*Tests de normalité*

Il y a plusieurs tests de normalité d'une distribution. La statistique de Jarque-Bera est couramment utilisée dans les études économétriques. C'est la méthode mise en œuvre dans la `proc autoreg` de SAS.

Principe : Supposons une série de  $n$  observations indépendantes d'une v.a.  $Y, y_1, \dots, y_n$ , de moyenne 0, de variance 1. Les 4 premiers moments empiriques sont

$$m_1 = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$m_q = \frac{1}{n} \sum_{i=1}^n (y_i - m_1)^q, \quad q = 2, 3, 4$$

L'asymétrie (skewness) empirique est

$$S = \frac{m_3}{m_2^{3/2}}$$

et l'aplatissement (kurtosis) empirique est

$$K = \frac{m_4}{m_2^2}$$

Si  $Y$  suit une loi normale on peut montrer qu'alors :

$$S \sim AN(0, \frac{6}{n}), \quad K \sim AN(3, \frac{24}{n}), \quad S \text{ et } K \text{ sont approximativement indépendants}$$

Ces résultats permettent de considérer :  $n \frac{S^2}{6}, n \frac{(K-3)^2}{24}$ , comme statistique pour tester la normalité, sous cette hypothèse en effet, chacune suit approximativement une loi de  $\chi_2^2$  à 1 degré de liberté.

La statistique de Jarque-Bera combine ces deux statistiques. Elle vaut :

$$N = n \left\{ \frac{S^2}{6} + \frac{(K-3)^2}{24} \right\}$$

Si  $Y$  de moyenne 0, de variance 1 est normalement distribuée de moyenne 0, alors  $N$  suit approximativement une loi de  $\chi_2^2$ . On rejette l'hypothèse nulle de normalité, pour de grandes valeurs de la statistique.

*Tests de non corrélation*

Pour un processus linéaire (voir le paragraphe suivant pour la définition et [5] Th. 7.2.2 pour les conditions exactes) dont la FAC vérifie :  $\rho(k) = 0, k > m$ , on a pour  $k > m$  :

$$\hat{\rho}(k) \sim AN(0, \text{var}(\hat{\rho}(k))) \quad AN \text{ signifie : approximativement normal} \quad (2.6)$$

$$\text{var}(\hat{\rho}(k)) \simeq \frac{1}{T} (1 + 2\rho(1)^2 + \dots + 2\rho(m)^2) \quad (2.7)$$

Application de ce résultat. Si  $\{Y_t\}$  est une suite de v.a. i.i.d., de moment d'ordre 2 fini,  $E(Y_t^2) < \infty$ , alors (2.6) se réduit à

$$\text{var}(\hat{\rho}(k)) \simeq \frac{1}{T} \quad (2.8)$$

On montre d'autre part que les coefficients d'autocorrélation  $\hat{\rho}(l)$  sont approximativement indépendants et normalement distribués (de moyenne 0, de variance  $1/T$ ). Ce résultat permet de tester la nullité d'un  $\rho(k)$  particulier. De plus étant donnée une réalisation  $y_1, \dots, y_T$  d'une série vérifiant (2.8), on devrait avoir environ 95% des coefficients d'autocorrélation empirique dans l'intervalle  $\pm 1,96/\sqrt{T}$ . Si la proportion observée s'éloigne de cette valeur, on peut conclure que les observations ne sont pas indépendantes.

**Test du Portemanteau** Si l'on veut tester que  $\{Y_t\}$  est un bruit blanc, on a évidemment intérêt à considérer simultanément la nullité de plusieurs coefficients d'autocorrélation. La statistique

$$Q(h) = T \sum_{j=1}^h \hat{\rho}(j)^2 \quad (2.9)$$

où  $h$  est un décalage qu'on choisit permet de tester :

$$H_0^h : \rho_1 = \rho_2 = \dots = \rho_h$$

Sous l'hypothèse que  $\{Y_t\}$  est un bruit blanc,  $Q$  suit approximativement une loi de  $\chi^2$  à  $h$  degrés de liberté.  $Q$  est appelée statistique de Box-Pierce. On rejette l'hypothèse  $H_0^h$  pour les grandes valeurs de  $Q(h)$ .

Quand on veut tester la blancheur du résidu d'un ajustement on utilise la statistique de Ljung-Box :

$$Q^*(h) = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}(k)^2}{T-k}.$$

Sous  $H_0 : \rho(1) = \dots = \rho(h) = 0$   $Q^*(h) \simeq \chi^2(h-m)$ ,  $m$  est le nombre de paramètres estimés dans le modèle.

Un exemple de sortie SAS de ce test figure à la sous-section(2.3.1).

**Test de Durbin-Watson.** Le test de Durbin-Watson est un test de blancheur de série particulier. Il s'intéresse à la situation :

$$y_t = \mathbf{x}_t' \beta + u_t, \quad t = 1, \dots, T \quad (2.10)$$

$$u_t = \rho u_{t-1} + Z_t \quad (2.11)$$

où :  $\mathbf{x}_t$  est un vecteur de  $p+1$  variables explicatives (dont la constante),  $Z_t \sim BB$ . Il teste  $H_0 : \rho = 0$ . La statistique de Durbin-Watson est

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2},$$

où  $\hat{u}_t$  est le résidu de l'ajustement par moindres carrés ordinaires de  $y_t$  sur  $x_t$ . En développant numérateur et dénominateur on voit que

$$DW \simeq 2(1 - \hat{\rho}) \in (0, 4],$$

où  $\hat{\rho} = \frac{\sum_{t=2}^T \hat{u}_{t-1} \hat{u}_t}{\sum_{t=1}^T \hat{u}_t^2}$ . Les valeurs de DW proches de 0 indiquent une autocorrélation proche de 1.

Pour le test de  $H_0 : \rho = 0$  contre  $H_1 : \rho > 0$ , la région critique (RC) correspond à de fortes valeurs de DW (DW sensiblement inférieur à 2, et la RC pour  $H_1 : \rho < 0$ , correspond à de faibles valeurs de DW. Pratiquement une statistique de DW  $\ll 2$  peut être le signe d'une mauvaise spécification du modèle (par exemple, tendance ajustée linéaire alors que la tendance réelle est quadratique).

Ce test étant effectué par la plupart des logiciels, que les observations soient ou ne soient pas une série temporelle, il ne faut pas oublier d'éliminer son résultat des sorties, quand les données ne sont pas indicées par le temps. Dans SAS il est programmé ainsi que certaines extensions, dans la proc autoreg.

**Notes.**

1. Plus le décalage  $l$  est grand, moins il y a d'observations pour estimer  $\rho_l$  dans (2.5). On s'arrête habituellement à  $l \simeq T/4$ .
2. Observons que l'on peut calculer (2.5) pour toute série, stationnaire ou non. Si la série n'est pas stationnaire, (2.5) a un usage purement empirique. On montre que pour une série stationnaire, le *corrélogramme empirique*, graphe de  $l \mapsto \hat{\rho}(l)$  décroît exponentiellement vers 0, avec éventuellement des oscillations. Inversement, un graphe du corrélogramme empirique qui ne décroît pas très rapidement est l'indice d'une non stationnarité de la série. Ainsi l'examen du corrélogramme empirique d'une série permet de se faire une idée de sa stationnarité. Cet examen complète celui du chronogramme.
3. Dans la formule (2.7) on ne connaît pas les  $\rho_k$ . On les remplace par leurs estimations dans la formule de la variance.
4. A partir des formules (2.6), (2.7) ou (2.8), on peut construire des intervalles autour de 0, par exemple à 95%. SAS base ces intervalles sur (2.6) et (2.7) alors que R utilise (2.8).
5. Il est très pratique de se souvenir de  $Q(h)$  comme d'une somme de  $h$  variables approximativement indépendantes, carrés de v.a.  $N(0, 1)$  :

$$Q(h) = T \sum_{j=1}^h \hat{\rho}(j)^2 = \sum_{j=1}^h \left( \frac{\hat{\rho}(j) - 0}{1/\sqrt{T}} \right)^2$$

or (1) le carré d'une variable  $N(0,1)$  suit une loi  $\chi_1^2$  (2) la somme de deux v.a. indépendantes et distribuées suivant des lois  $\chi_{n_1}^2$  et  $\chi_{n_2}^2$  suit une loi  $\chi_{n_1+n_2}^2$ . La loi de  $Q(h)$  est donc bien  $\chi^2$  sous l'hypothèse nulle.

(cf Formule de Bartlett).

6. La loi de la statistique DW dépend des variables explicatives  $x$ . Durbin et Watson ont trouvé des encadrements aux quantiles de cette loi qui ne dépendent que du nombre de variables explicatives, en plus de la constante. C'était très utile dans les années 50, quand les ordinateurs étaient rares.
7. Le modèle (2.10) est de même nature que celui retenu pour le lac Huron au chapitre précédent.

## 2.2 Modèles de séries stationnaires

### 2.2.1 Série linéaire

Une série  $Y_t$  est dite *linéaire* si elle peut s'écrire :

$$Y_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i} \quad (2.12)$$

où  $Z_t$  est un  $BB(0, \sigma_Z^2)$ ,  $\psi_0 = 1$  et la suite  $\{\psi_i\}$  est *absolument sommable*, c'est-à-dire  $\sum_i |\psi_i| < \infty$ . On admettra qu'une série linéaire est stationnaire.

Une série est dite *linéaire et causale* si elle est linéaire et  $\psi_i = 0$ ,  $i < 0$ , autrement dit elle ne dépend que du BB passé. Pour une discussion sur le sens de l'écriture ci-dessus, on peut voir par exemple Hamilton ([13]), paragraphe 3.3.



Si  $Y_t$  est linéaire et causal on obtient :  $EY_t = \mu$ ,  $\text{var}(Y_t) = \sigma_Z^2 \sum_{i=0}^{\infty} \psi_i^2$  L'autocovariance d'ordre  $k$  est :

$$\gamma_k = \text{cov}(Y_t, Y_{t-k}) = E\left[\sum_{i=0}^{\infty} \psi_i Z_{t-i}, \sum_{j=0}^{\infty} \psi_j Z_{t-k-j}\right] \quad (2.13)$$

$$= E\left(\sum_{i,j=0}^{\infty} \psi_i \psi_j Z_{t-i} Z_{t-k-j}\right) \quad (2.14)$$

$$= \sum_{j=0}^{\infty} \psi_{j+k} \psi_j E(Z_{t-k-j}^2) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \quad (2.15)$$

Si la série est linéaire et causale et si de plus  $\psi_i = 0$  pour  $i > q$  on dit que  $Y_t$  est une moyenne mobile d'ordre  $q$  (**MA(q)**). Une série linéaire causale est un **MA( $\infty$ )**.

#### Notes.

1. La définition d'une série linéaire a un intérêt théorique car on peut lui associer une caractérisation de la stationnarité.
2. On peut observer que les calculs d'autocovariances sont particulièrement simples dans cette écriture.
3. Il peut sembler peu pratique d'écrire un processus à l'aide d'une infinité de paramètres mais les modèles ARMA que nous allons voir : (a) sont des cas particuliers à peu de paramètres de séries linéaires et causales (b) peuvent être de bonnes approximations de telles séries.

### 2.2.2 Processus gaussien

Nous admettons sans justification, la définition non formalisée suivante. Un *processus*  $\{Y_t\}$  est *gaussien* si :

- $\forall k$ , un nombre de dates,
- $\forall (t_1, t_2, \dots, t_k)$ , un choix de  $k$  dates,

le vecteur aléatoire :  $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})$  est normalement distribué. On peut observer qu'un processus linéaire à bruit blanc gaussien est gaussien.

### 2.3 Processus autorégressif

Les calculs d'autocovariances sont simples sur un processus linéaire, mais cette écriture en terme d'un bruit non observé n'est pas très pratique. On préfère avoir une expression d'un processus explicitement en fonction de son passé. C'est le cas des processus autorégressifs. Mais nous verrons aussi à quelles conditions il est possible d'écrire un processus linéaire comme une autorégression. C'est la notion d'inversibilité.

#### 2.3.1 Processus autorégressif d'ordre 1

On dit que  $\{Y_t\}$  est un processus autorégressif d'ordre 1 s'il obéit à une équation :

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + Z_t, \quad t \in Z \quad (2.16)$$

##### Moments d'ordres 1 et 2 d'un AR(1)

Supposons  $\{Y_t\}$  dans (2.16) stationnaire alors, sa moyenne  $\mu$ , est constante et prenant l'espérance mathématique des deux côtés on obtient

$$\mu = \phi_0 + \phi_1 \mu$$

et si  $\phi_1 \neq 1$  :

$$E(Y_t) = \mu = \frac{\phi_0}{1 - \phi_1}.$$

Comme alors,  $\phi_0 = \mu(1 - \phi_1)$  on a :

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + Z_t$$

On pose  $\dot{Y}_t = Y_t - \mu$ . C'est le processus centré. Avec l'opérateur retard, on a :

$$(1 - \phi_1 B)\dot{Y}_t = Z_t$$

Par substitutions successives on obtient que  $\dot{Y}_t$  peut être exprimé comme une moyenne mobile infinie :

$$\dot{Y}_t = Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2} + \dots \quad (2.17)$$

pourvu que  $-1 < \phi_1 < 1$ . Cette condition est suffisante pour que le processus soit stationnaire. Dans ce cours nous supposons toujours qu'un processus qui vérifie (2.16) est stationnaire si et seulement si  $-1 < \phi < 1$ .

On appelle (2.17) la représentation MA( $\infty$ ) de  $Y_t$ . L'écriture de  $Y_t$  comme une somme de v.a. non corrélées permet de calculer facilement les variance et autocovariances comme nous le voyons maintenant. Elevons au carré les deux côtés de (2.17), il vient :

$$\begin{aligned} \text{var}(Y_t) &= \sigma_Z^2(1 + \phi^2 + \phi^4 + \dots) \\ &= \frac{\sigma_Z^2}{1 - \phi^2} \end{aligned}$$

Enfin écrivons (2.17) en  $t - k$  et calculons les espérances des deux côtés de :

$$Y_t Y_{t-k} = (Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \dots)(Z_{t-k} + \phi Z_{t-k-1} + \phi^2 Z_{t-k-2} + \dots)$$

où,  $Z_t$  étant un BB,  $E(Z_t Z_{t-m}) = 0, m \neq 0$ . On obtient pour  $k > 0$

$$\gamma_k = (\phi^k + \phi^{k+2} + \phi^{k+4} + \dots)\sigma_Z^2 = \phi^k \gamma_0$$

La fonction d'autocorrélation de l'AR(1) est :

$$\rho_k = \phi^k, k = 0, 1, 2, \dots \quad (2.18)$$

### Exemples

**AR(1) simulé.** On a simulé 150 observations de  $y_t$  obéissant à

$$y_t = 4 - 0.8 y_{t-1} + Z_t, \quad Z_t \sim BBN(0, 1)$$

On a obtenu une moyenne empirique de 2.271132 et un écart-type 1.371441.

Retard	FAC théorique	FAC empirique
0	1	1.00000
1	-0.8	-.72267
2	0.64	0.46219
3	-0.512	-.30436
4	0.4096	0.22636
5	-0.32768	-.13674
6	0.26214	0.06958

Le test de blancheur de la série est :

	To	Chi-		Pr >						
	Lag	Square	DF	ChiSq	-----Autocorrelations-----					
5	6	138.90	6	<.0001	-0.723	0.462	-0.304	0.226	-0.137	0.070
	12	156.79	12	<.0001	-0.136	0.168	-0.172	0.162	-0.090	-0.017
	18	160.94	18	<.0001	0.067	-0.032	0.045	-0.080	0.084	-0.058
	24	164.51	24	<.0001	0.067	-0.057	0.014	0.025	-0.057	0.091

Ces résultats ont été obtenu par la syntaxe (la série simulée  $y$  est dans la table a) :

```
proc arima data = a; i var= y; run; quit;
```

**Processus AR(2)** Soit  $Y_t$  stationnaire, obéissant à l'équation :

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t \quad (2.19)$$

Prenant l'espérance des deux côtés on obtient :

$$E(Y_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

pourvu que  $1 \neq \phi_1 + \phi_2$ . Comme  $\phi_0 = \mu(1 - \phi_1 - \phi_2)$ , on a :

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + Z_t$$

et le processus centré :  $\dot{Y}_t = Y_t - \mu$  vérifie :

$$\dot{Y}_t = \phi_1 \dot{Y}_{t-1} + \phi_2 \dot{Y}_{t-2} + Z_t. \quad (2.20)$$

Nous obtenons :

$$\gamma(l) = \phi_1 \gamma(l-1) + \phi_2 \gamma(l-2), l > 0 \quad (2.21)$$

On appelle (2.21), l'équation de moments d'un AR(2). La fonction d'autocorrélation d'un AR(2) est :

$$\begin{aligned} \rho(1) &= \frac{\phi_1}{1 - \phi_2} \\ \rho(l) &= \phi_1 \rho(l-1) + \phi_2 \rho(l-2), l > 1 \end{aligned} \quad (2.22)$$

**Condition de stationnarité d'un AR(2)** Il faut pouvoir l'écrire comme un MA( $\infty$ ). Précisons. Un processus qui vérifie (2.19) est stationnaire si et seulement si les racines du polynôme caractéristique  $1 - \phi_1 B - \phi_2 B^2$  sont  $> 1$  en module.

Explication. Soit  $s_1$  et  $s_2$  les racines de  $1 - \phi_1 z - \phi_2 z^2 = 0$ . On a donc  $(1 - z/s_1)(1 - z/s_2) = 0$  et on peut développer en série  $\frac{1}{1 - \phi_1 z - \phi_2 z^2}$  ssi  $|1/s_1| < 1$  et  $|1/s_2| < 1$ .

**Processus autorégressif d'ordre p** Une extension immédiate du modèle AR(1) est le modèle AR(p). Soit  $\{Z_t\}$  un BB. Un processus  $\{Y_t\}$  est dit autorégressif d'ordre  $p$  s'il s'écrit :

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + Z_t$$

Avec l'opérateur retard on peut écrire cette autorégression à l'ordre  $p$  comme :

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t &= \phi_0 + Z_t \\ \Phi(B) Y_t &= \phi_0 + Z_t \end{aligned}$$

**Condition de stationnarité d'un processus autorégressif d'ordre p.** Nous inspirant de ce qu'on a obtenu pour un AR(2), on admettra qu'un processus autorégressif d'ordre  $p$  est stationnaire si les racines de l'équation :  $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$  sont en module  $> 1$ .

### 2.3.2 Fonction d'autocorrélation partielle

Considérons une série stationnaire  $\{X_t\}$  et ses régressions linéaires sur son passé :

$$\begin{aligned} X_t &= \phi_{0,1} + \phi_{1,1} X_{t-1} + u_{1t} \\ X_t &= \phi_{0,2} + \phi_{1,2} X_{t-1} + \phi_{2,2} X_{t-2} + u_{2t} \\ X_t &= \phi_{0,3} + \phi_{1,3} X_{t-1} + \phi_{2,3} X_{t-2} + \phi_{3,3} X_{t-3} + u_{3t} \\ &\vdots \end{aligned}$$

Par exemple,  $\phi_{0,2} + \phi_{1,2} X_{t-1} + \phi_{2,2} X_{t-2}$  désigne l'espérance conditionnelle linéaire de  $X_t$  sachant  $X_{t-1}, X_{t-2}$ , voir par exemple ([11] Chap. 5, ou [13]...). Les  $\phi_{k,k}, k = 1, 2, \dots$  forment la fonction d'autocorrélation partielle (PACF) et les estimations  $\hat{\phi}_{k,k}, k = 1, 2, \dots$  forment la *fonction d'autocorrélation partielle empirique*.  $\phi_{k,k}$  a la même interprétation que les coefficients d'une régression linéaire classique :  $\phi_{k,k}$  représente l'apport d'explication de  $X_{t-k}$  à  $X_t$  étant donné qu'on a déjà régressé sur  $X_{t-1}, \dots, X_{t-k+1}$  (ou toutes choses égales par ailleurs).

Supposons en particulier que  $X_t$  soit autorégressif, un AR(3) pour fixer les idées, alors il est clair que  $X_{t-4}$  n'apporte rien de plus que  $X_{t-1}, X_{t-2}, X_{t-3}$  pour expliquer  $X_t$  et on montre en effet que pour un AR(3)  $\phi_{k,k} = 0, k > 3$ . De même la PACF d'un AR(p) est nulle à partir de l'ordre  $p + 1$ .

Estimation de la PACF. De même que dans la méthode des MCO les  $\hat{\beta}$  s'expriment en fonction des moments d'ordre 2 des observations, la PACF d'une série stationnaire s'exprime en fonction de son ACF (cette dépendance est décrite par l'algorithme de Durbin-Levinson).

D'un point de vue pratique on pensera qu'une série suit un AR(p) si les  $\hat{\phi}_{k,k} \simeq 0, k > p$ . Précisément, si  $X_t$  est un AR(p), on a les propriétés :

- $\hat{\phi}_{p,p}$  converge vers  $\phi_{p,p}$  quand  $n \rightarrow \infty$ ,
- $\hat{\phi}_{l,l}, \forall l > p$  converge vers 0 quand  $n \rightarrow \infty$ ,
- $\text{var}(\hat{\phi}_{l,l}) \simeq 1/n \forall l > p$ .

La PACF empirique permet donc d'identifier un processus comme un AR d'un certain ordre  $p$ . Mais on peut également identifier l'ordre d'un processus qu'on sait AR, par un critère d'information (AIC, SBIC...).

## 2.4 Processus Moyenne mobile

**Processus MA(1)** On dit que  $(Y_t)$  est un processus moyenne mobile d'ordre 1 (MA(1)), s'il obéit à une équation :

$$Y_t = \mu + Z_t - \theta Z_{t-1} \quad (2.23)$$

où  $Z_t \sim \text{BB}(0, \sigma_Z^2)$ . Cette équation s'écrit encore :

$$Y_t - \mu = (1 - \theta B) Z_t$$

**Moments d'ordres 1 et 2 d'un MA(1)** En prenant l'espérance mathématique des deux côtés de (2.23), on voit que

$$E(Y_t) = \mu$$

La variance de  $Y_t$  est la variance d'une combinaison affine de variables non corrélées donc :  $\text{var}(Y_t) = (1 + \theta^2)\sigma_Z^2$ . De même,  $\text{cov}(Y_t, Y_{t-1}) = \text{cov}(\mu + Z_t - \theta Z_{t-1}, \mu + Z_{t-1} - \theta Z_{t-2}) = -\theta\sigma_Z^2$ . On voit que  $\text{cov}(Y_t, Y_{t-k}) = 0, k > 1$ . En résumé,  $\forall \theta$ , le processus MA(1) défini par (2.23) est stationnaire, de moyenne  $\mu$ , de fonction d'autocorrélation :

$$\rho(k) = \begin{cases} 1 & \text{si } k = 0, \\ \frac{-\theta}{1+\theta^2} & \text{si } k = 1, \\ 0 & \text{si } k > 1 \end{cases}$$

Etudiant la fonction  $\mathbf{R} \ni \theta \rightarrow \frac{-\theta}{1+\theta^2}$  on constate que  $|\frac{-\theta}{1+\theta^2}| \leq 0.5$  ainsi :

►le processus MA(1) ne peut pas décrire des phénomènes à forte autocorrélation. ◀

D'autre part, on aimerait pouvoir exprimer le processus en fonction de son passé (observé) et pas seulement en fonction d'un bruit.

**Inversibilité.** Introduisons le processus centré,  $\dot{Y}_t = Y_t - \mu$ , correspondant à (2.23). On voit que si  $|\theta| < 1$ , on peut développer  $(1 - \theta B)^{-1}$  en série entière. Ceci nous amène à une définition.

On dit qu'un processus est *inversible* si on peut l'écrire comme une autorégression infinie. Ainsi, un MA(1) est inversible si la racine de l'équation  $1 - \theta z = 0$  est  $> 1$  en module. On observe que la condition d'inversibilité d'un MA(1) est techniquement parallèle à la condition de stationnarité d'un autorégressif d'ordre 1.

**Processus MA(q)** Un processus  $(Y_t)$  est dit **processus moyenne mobile** d'ordre  $q$  ( $MA(q)$ ) si :

$$Y_t = \mu + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \dots - \theta_q Z_{t-q}$$

où  $Z_t \sim \text{BB}(0, \sigma_Z^2)$ .

On peut noter de façon équivalente :

$$\begin{aligned} Y_t &= \mu + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t \\ &= \mu + \Theta(B) Z_t \end{aligned}$$

Un  $MA(q)$  est une série linéaire.

Propriétés.

1. Un  $MA(q)$  est un processus stationnaire.
2. La fonction d'autocorrélation d'un processus MA(q) est nulle à partir de l'ordre  $q + 1$ .

Cette dernière propriété est utile pour deviner (identifier) l'ordre de moyenne mobile convenable pour modéliser une série. En effet, en présence d'un corrélogramme empirique non significativement différent de 0 à partir d'un certain ordre ( $k$ ), on pensera à modéliser la série correspondante par un  $MA(k - 1)$ .

**Inversibilité d'un MA(q).** Un MA(q) est inversible si les racines de  $1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0$  sont, en module,  $> 1$ .

**Processus ARMA(p,q)**  $\{Y_t\}$  est un processus ARMA(p,q) s'il est stationnaire et obéit à une équation :

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p} = \theta_0 + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \dots - \theta_q Z_{t-q} \quad (2.24)$$

où  $Z_t \sim \text{BB}(0, \sigma_Z^2)$ . Observons que (2.24) s'écrit aussi :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t,$$

ou encore :

$$Y_t = \frac{\theta_0}{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p} + \frac{1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p} Z_t, \quad (2.25)$$

Examinons le 1<sup>er</sup> terme du côté droit avec, pour simplifier l'écriture,  $p = 2$ . C'est une fraction rationnelle en  $B$ . Le numérateur est une constante et le polynôme  $1 - \phi_1 B - \phi_2 B^2$  se factorise en termes de degré 1, réels ou complexes :  $1 - \phi_1 B - \phi_2 B^2 = (1 - d_1 B)(1 - d_2 B)$  où les  $d_i$  sont les inverses des racines du polynôme.

La stationnarité de  $Y_t$  implique que  $|d_i| < 1, i = 1, 2$ . Alors  $(1 - d_1 B)^{-1} = 1 + d_1 B + d_1^2 B^2 + \dots$  et  $(1 - \phi_1 B - \phi_2 B^2)^{-1} = (1 + d_1 B + d_1^2 B^2 + \dots)(1 + d_2 B + d_2^2 B^2 + \dots)$  et donc le processus peut s'écrire comme un MA(*infinity*) et est donc bien stationnaire.

Maintenant, si  $c$  est constante dans le temps,  $(1 + d_1 B + d_1^2 B^2 + \dots)c = (1 + d_1 + d_1^2 + \dots)c = (1/(1 - d_1))c$ . Finalement (2.25) revient à :

$$Y_t = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p} + \frac{1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p} Z_t \quad (2.26)$$

D'autre part  $E(\frac{1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p} Z_t) = 0$ , en utilisant le même argument que ci-dessus.

D'où :

$$\mu = E(Y_t) = \frac{\theta_0}{1 - \phi_1 - \dots - \phi_p}.$$

La fonction d'autocovariance d'un ARMA( $p, q$ ) se calcule par des équations de récurrence (voir 2.22 pour un AR(2)).

### Notes

- 1 Les équations qui relient les paramètres AR et MA aux coefficients d'autocorrélation (voir (2.21, pour le cas d'un AR(2)) sont les *équations de Yule-Walker* : étant donné les coefficients auto-régressifs et de moyenne mobile, on peut trouver les coefficients d'autocorrélation. Inversement, étant donné une fonction d'autocorrélation empirique et des ordres  $p$  et  $q$ , on peut obtenir par ces équations des estimations des paramètres AR et MA. Ces estimations servent de valeurs initiales dans l'estimation MV <sup>1</sup> des paramètres.
- 2 Un ARMA( $p, q$ ) stationnaire et inversible peut s'écrire comme un AR( $\infty$ ). Mais très souvent un AR( $p_\epsilon$ ) d'ordre  $p_\epsilon$  pas très supérieur à  $p$  en donne une très bonne approximation : les paramètres sont certes mal estimés mais la série est bien prédite et son erreur correctement reconstituée. C'est l'idée qu'exploite la méthode d'identification MINIC que nous verrons plus loin.
- 3 On trouvera en annexe de ce chapitre, un exemple d'équations de Yule-Walker, à bien comprendre, ainsi qu'une présentation de l'estimation des AR et des MA qu'il faut au moins parcourir pour voir que l'estimation d'un MA est assez difficile.
- 4 On trouvera encore en annexe, une présentation de la prévision, qu'il suffira de parcourir pour comprendre le principe, mais sur laquelle il n'est pas utile de se pencher longuement.

<sup>1</sup>MV = Maximum de Vraisemblance

## 2.5 Saisonnalité

### 2.5.1 Saisonnalité multiplicative

Décrivons brièvement la modélisation de la saisonnalité dans l'approche de Box-jenkins. Soit une série mensuelle observée (pour simplifier) sur un nombre entier d'années, à partir d'un mois de janvier. On note  $y_{ij}$  l'observation du mois  $j$  de l'année  $i$ ;  $j = 1, \dots, 12, i = 1, \dots, N$ .

Rangeons les valeurs observées en une table de Buys-Ballot : une ligne par an, une colonne par mois :

janv.	févr.	mars	...	déc.
$y_{11}$	$y_{12}$	$y_{13}$	...	$y_{1,12}$
$y_{21}$	$y_{22}$	$y_{23}$	...	$y_{2,12}$
...	...	...	...	...
$y_{N1}$	$y_{N2}$	$x_{13}$	...	$y_{N,12}$

Supposons qu'on modélise la dépendance d'un mois sur un ou deux mois précédents (sans s'occuper de l'effet saisonnier) et qu'on adopte un ARMA(p,q) :

$$\Phi(B)Y_t = c + \Theta(B)b_t$$

Il est fort probable, si la série présente une saisonnalité, que le résidu  $\hat{b}_t$  ne sera pas blanc mais aura une structure de corrélation saisonnière. On peut envisager deux traitements de cette "non blancheur". Ou bien on ajoute des termes de retard dans les polynômes  $\Phi$  et  $\Theta$ , ou bien on modélise  $b_t$  par un ARMA dont l'unité de temps est l'année :

$$b_t = \frac{\Theta_s(B^s)}{\Phi_s(B^s)} z_t$$

où  $s$  désigne la période (ici,  $s = 12$ ). Ce qui donne :

$$\Phi_s(B^s)\Phi(B)Y_t = c_1 + \Theta(B)\Theta_s(B^s)Z_t \quad (2.27)$$

avec  $Z_t \sim \text{BB}$ , où  $\Phi(B)$ ,  $\Phi_s(B^s)$ ,  $\Theta(B)$ ,  $\Theta_s(B^s)$  sont respectivement des polynômes de degrés  $p, q$  en  $B$  et  $P, Q$  en  $B^s$ . On dit que  $Y_t$  est un SARMA( $p, q$ )( $P, Q$ ) $_s$  s'il vérifie (2.27) et est stationnaire.

Les conditions de

- stationnarité de  $Y_t$  sont : les racines des polynômes  $\Phi(B)$  et  $\Phi_s(B^s)$  sont en module  $> 1$ .
- inversibilité de  $Y_t$  sont : les racines des polynômes  $\Theta(B)$  et  $\Theta_s(B^s)$  sont en module  $> 1$ .

## 2.6 Modèles ARIMA et SARIMA

Un processus ARIMA(p,d,q) est un processus dont la différence d'ordre  $d$  est un ARMA(p,q). Par exemple  $Y_t$  vérifiant :

$$(1 - B)Y_t = c + \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B - \phi_3 B^3} Z_t$$

est un ARIMA(3,1,2),

$Y_t$  vérifiant :

$$(1 - B)^2 Y_t = c + (1 - \theta_1 B) Z_t$$

est un ARIMA(0,2,1).

Un ARIMA est autorégressif mais son polynôme d'autorégression a 1, racine  $d$  fois.

Ainsi l'ARIMA(3,1,2) ci-dessus, peut s'écrire

$$(1 - (1 + \phi_1)B + \phi_1 B^2 - \phi_3 B^3 + \phi_3 B^4)Y_t = d + (1 - \theta_1 B - \theta_2 B^2)Z_t$$

il y a bien une partie autorégressive mais son polynôme d'autorégression a 1 comme racine et la série est non stationnaire. On n'écrit pas que c'est un AR(4,2) mais un ARIMA(3,1,2).

La série différenciée  $(1 - B)Y_t$  est la série des accroissements de  $Y_t$ , alors que la série différenciée deux fois  $(1 - B)^2 Y_t$  est la série des accroissements des accroissements. L'opération inverse de la différenciation est l'intégration d'où le I dans "ARIMA".

Un processus  $Y_t$  est un SARIMA( $p, d, q$ )( $P, D, Q$ ) $_s$  s'il obéit à

$$(1 - B)^d (1 - B^s)^D Y_t = c + \frac{(1 - \theta_1 B \cdots - \theta_q B^q)(1 - \Theta_1 B^s - \cdots - \Theta_Q B^{sQ})}{(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP})} Z_t \quad (2.28)$$

$$(1 - B)^d (1 - B^s)^D Y_t = c + \frac{\Theta(B)\Theta_s(B^s)}{\Phi(B)\Phi_s(B^s)} Z_t$$

où

$$\Theta_s(B^s) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{sQ}$$

et

$$\Phi_s(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP}$$

Par exemple un SARIMA(0, 1, 1)(1, 1, 0) $_{12}$  obéit à :

$$(1 - B)(1 - B^{12})Y_t = c + \frac{1 - \theta_1 B}{1 - \Phi_1 B^{12}} Z_t \quad (2.29)$$

Nous verrons au chapitre (3), les tests de racine unité permettant de décider s'il faut différencier une série.

## 2.7 Complément : non stationnarité en variance

On envisage le cas d'une série  $\{Y_t\}$  dont la moyenne,  $\mu_t$ , varie avec le temps de façon déterministe et dont la variance dépend du niveau moyen :

$$Y_t = \mu_t + U_t$$

avec  $\text{var}(U_t) = h^2(\mu_t)\sigma^2$  pour une certaine fonction  $h$ . Pour traiter cette situation d'hétéroscédasticité, on cherche une transformation  $g$  telle que  $\text{var}(g(Y_t)) \simeq \text{constante}$ . C'est la technique dite de *stabilisation de la variance*.

Par linéarisation on a :

$$g(Y_t) \simeq g(\mu_t) + (Y_t - \mu_t)g'(\mu_t)$$

et

$$\text{var}(g(Y_t)) \simeq [g'(\mu_t)]^2 \text{var}(Y_t)$$

On cherche donc  $g$  telle que  $g'(x) = 1/h(x)$ . Par exemple, pour  $h(x) = x$ ,  $g'(x) = 1/x$  et donc  $g(x) = \log(x)$ .

Soit une série  $Y_t$  qui montre à la fois une croissance dans le temps :  $\mu_t = E(Y_t)$  fonction croissante de  $t$  et une variabilité qui augmente avec  $t$  :  $\text{var}(Y_t)$  fonction croissante de  $t$ . Si on modélise ce phénomène par :  $\text{var}(Y_t) = c\mu_t^2$  on voit que la transformation en log stabilisera la variance. Cette situation est classique dans les séries de stock. Les exemples "ventes de champagne", "ventes de vin" entrent dans ce cadre. La série de la consommation britannique utilisée plus loin dans ce chapitre en est un autre exemple. En présence de telles séries on commence par les transformer en log avant de faire une modélisation ARIMA ou autre.

**Complément.** On appelle *transformation de Box-Cox*, la transformation :

$$Y_t \mapsto \frac{Y_t^\lambda - 1}{\lambda}$$

où  $\lambda$  est un paramètre positif à choisir. Quand  $\lambda \rightarrow 0$  la transformation de Box-Cox revient à la transformation en log. Ces transformations peuvent ramener des distributions vers la normalité.



## 2.8 Etude empirique d'une série par une modélisation ARIMA

Etant donné une série sans saisonnalité et qu'on veut essayer de modéliser par un ARIMA on doit choisir 3 paramètres :  $p$ ,  $d$  et  $q$  puis on estime le modèle enfin on prédit la série. La validation du modèle se fait à 2 niveaux :

1. à la fin de l'estimation on doit tester la blancheur du résidu  $\hat{Z}_t$ , la significativité des coefficients estimés, voir s'il y a des messages du logiciel indiquant que l'algorithme d'estimation n'a pas convergé.
2. à la prévision : on dessine un intervalle de prévision pour une partie de série observée et on regarde si la proportion de points dans l'intervalle est proche, ni trop supérieure ni trop inférieure, du niveau choisi pour l'intervalle.

Si le résultat n'est pas satisfaisant on recommence avec d'autres valeurs de  $p$ ,  $d$  ou  $q$ .

Pour une série présentant une saisonnalité, et qu'on veut essayer de modéliser par un SARMA il y a donc 6 paramètres :  $p$ ,  $d$  et  $q$  et  $P$ ,  $D$ ,  $Q$  à choisir.

On commence toujours par examiner le diagramme séquentiel de la série.

Pour le choix de  $p$ ,  $d$  et  $q$  on s'aide des 3 fonctions : ACF, PACF et IACF (*fonction d'autocorrélation inverse*). L'IACF d'un ARMA( $p, q$ ) est définie comme l'ACF de l'ARMA( $q, p$ ) de la série obtenue en permutant les rôles des coefficients AR et MA dans son modèle. SAS est un des rares logiciels à fournir cette fonction.

Fonction	MA( $q$ )	AR( $p$ )	ARMA( $p, q$ )	bruit blanc
ACF	Nul( $q$ )	Exp	Exp	0
PACF	Exp	Nul( $p$ )	Exp	0
IACF	Exp	Nul( $p$ )	Exp	0

où :

Nul( $q$ ) veut dire, nulle à partir du décalage  $q+1$ , Exp, décroissance exponentielle et 0, que la fonction est nulle à tous les décalages.

TAB. 2.1 – Comportement des 3 fonctions d'autocorrélation, suivant le modèle de la série.

**Étape d'identification** (au sens de l'automatique) Il faut, avant l'estimation, choisir les paramètres  $p$ ,  $d$  et  $q$ . En examinant l'aspect de la FAC (Fonction d'autocorrélation) empirique, on se fait une idée de la stationnarité (atteinte ou non) de la série. On considère qu'une série est stationnaire quand sa FAC empirique décroît suffisamment vite vers 0, mais nous verrons dans un autre chapitre les tests de racine unité (il y a racine unité dans l'autorégression si  $d > 0$ ). On examine ensuite l'aspect des ACF et PACF de la série. On se fait ainsi une idée des ordres  $p$  et  $q$  possibles. Des méthodes d'aide au choix de  $p$  et  $q$  ont été proposées. SAS en intègre plusieurs. Dans une sous-section, nous décrivons rapidement la méthode MINIC.

**Étape d'estimation** Une fois  $p$  et  $q$  choisis, on estime le modèle. On teste d'abord que le résidu  $\hat{Z}_t$  est un bruit blanc. Si on doit rejeter cette hypothèse, il faut revoir le choix des ordres  $p$  et  $q$ , et de la transformation initiale. Si on peut considérer que le résidu est un bruit blanc, on analyse les résultats de l'estimation suivant les mêmes principes qu'en régression linéaire : significativité des  $\phi_i$  et  $\theta_j$  d'après les t-statistiques, corrélations entre les estimateurs de ces paramètres.

Si on rejette la blancheur du résidu, l'examen de son ACF ou PACF aide à corriger le modèle. On en a vu un exemple avec les SARMA.

Remarque. Les logiciels qui ajustent un modèle ARMA d'ordres  $p$  et  $q$  donnés, à une série supposée stationnaire, fournissent une représentation inversible : les racines de  $\Phi(B) = 0$  et  $\Theta(B) = 0$  sont  $> 1$  en module. Quand on essaie d'ajuster un modèle ARMA à une série non stationnaire, on obtient souvent un message d'avertissement ou d'erreur de la part du logiciel car les procédures

numériques d'optimisation mises en œuvre pour l'estimation ne convergent pas ou convergent mal dans un tel cas.

**Étape de prévision** On utilise enfin le modèle estimé pour prédire la série. Il est recommandé de n'utiliser qu'une partie de la série pour estimer le modèle, de façon à pouvoir comparer ensuite pour un même intervalle de temps, réalisations et prévisions. L'étape de prévision fournit le résidu  $Z_t, t = 1, \dots, T$  et si on indique un horizon  $h$  de prévision, les prévisions  $y_T(k), k = 1, \dots, h$  et, sous l'hypothèse de normalité de la série, les intervalles de prévision, basés sur des variances d'erreur de prévision du type de (2.44) et à un niveau qu'on peut choisir.

L'échec de la modélisation peut venir de causes variées dont la non stationnarité de la série.

### Considérations pratiques pour la modélisation ARIMA

A Parcimonie. On ne doit pas accumuler des paramètres. C'est un principe général en statistique. Dans une modélisation ARIMA il ne faut donc pas ajouter des paramètres pour capter la moindre trace d'autocorrélation dans les résidus d'un ajustement antérieur.

B On a rarement un modèle satisfaisant dès la 1<sup>re</sup> identification. A partir d'un premier modèle il faut alors chercher des voies d'amélioration.

- Aller du simple au compliqué. On commence par capter les principales caractéristiques par l'examen des ACF et PACF. On examine ensuite les ACF et PACF des résidus du modèle ajusté et on corrige le modèle à estimer ... et on recommence. C'est sur cette idée qu'on a introduit les modèles SARMA.
- Simplifier. On commence par un modèle avec (relativement) beaucoup de paramètres. Si on accepte la blancheur de son résidu, on essaie de le simplifier en examinant les t-statistiques des coefficients et la matrice des coefficients de corrélation entre les estimateurs des paramètres (des coefficients supérieurs à 0.9 indiquent souvent une mauvaise spécification).

C Toutes les séries ne se prêtent pas à une modélisation ARIMA et une modélisation ARIMA ne permet pas certaines opérations telle que la mise en évidence d'une composante saisonnière.

D Prévision. Le temps qu'on peut consacrer à la prévision d'une série guide le choix des méthodes. Si l'on doit prédire quelques centaines de séries en une journée, on utilisera des méthodes de prévision (souvent du lissage exponentiel) qui choisissent automatiquement un modèle sur un critère de qualité de la prévision à un certain horizon.

Voir sur ces questions la documentation de la `proc forecast` de SAS ou le package `forecast` de R :

[www-personal.buseco.monash.edu.au/~hyndman/Rlibrary/forecast/](http://www-personal.buseco.monash.edu.au/~hyndman/Rlibrary/forecast/)

### 2.8.1 La méthode MINIC

La méthode MINIC pour "Minimum Information Criterion" aide à identifier les ordres  $p$  et  $q$  pour une série stationnaire susceptible de recevoir une modélisation ARMA.

#### Principe

Soit une série  $Y_t, t = 1, \dots, n$ , centrée *stationnaire* et *invertible* suivant un modèle ARMA, d'ordres  $p$  et  $q$  inconnus.

$$Y_t = \frac{1 - \theta_1 B - \dots - \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} Z_t \quad (2.30)$$

ou

$$\Phi_{p,q}(B)Y_t = \Theta_{p,q}(B)Z_t \quad (2.31)$$

où  $Z_t \sim \text{BB}(0, \sigma_Z^2)$ . Rappelons que  $Z_t$  est le résidu de la projection de  $Y_t$  sur son passé :  $Y_{t-1}, Y_{t-2}, \dots$ . L'expression qu'on peut choisir pour cette projection n'en change pas la valeur.

► Comme le processus est supposé inversible on peut le représenter comme un  $\text{AR}(\infty)$  et en choisissant un ordre d'autorégression suffisamment grand on peut toujours approcher le modèle ARMA de cette série par un modèle AR. Précisément :

$$\Phi_{p_\epsilon}(\text{B})Y_t = \tilde{Z}_t \simeq Z_t \quad (2.32)$$

pour  $p_\epsilon$  suffisamment grand, où les paramètres  $\Phi$  sont estimés par les équations de Yule-Walker.

► D'autre part le critère BIC (Bayesian Information Criterion) pour un ajustement d'un  $\text{ARMA}(m, j)$  gaussien vaut :

$$\text{BIC}(m, j) = n \ln(s^2) + 2(m + j) \ln(n) \quad (2.33)$$

où l'estimation de la variance de l'erreur du bruit blanc, est

$$s^2 = \frac{1}{n} \sum_{t=\max(m, j)}^n \hat{\tilde{Z}}_t^2.$$

On peut observer que le BIC dépend du nombre de paramètres et, seulement indirectement, de leurs valeurs par l'intermédiaire de  $s^2$  estimation de la variance de l'innovation. Comme pour l'AIC rappelé un peu plus loin, entre plusieurs modèles non nécessairement emboîtés, on retient le modèle de BIC minimum.

► Si on a ajusté un AR d'ordre  $p_\epsilon$  suffisamment élevé, la projection de  $Y_t$  sur son passé est correctement évaluée et le résidu  $\hat{\tilde{Z}}_t$  est alors une *bonne prédiction de l'erreur de prévision*,  $Z_t$ . Pour estimer alors un  $\text{ARMA}(m, j)$  on ajuste le modèle :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_m Y_{t-m} - \theta_1 \hat{\tilde{Z}}_{t-1} - \theta_2 \hat{\tilde{Z}}_{t-2} - \dots - \theta_j \hat{\tilde{Z}}_{t-j} + u_t, \quad (2.34)$$

c'est un modèle linéaire en  $\phi$  et  $\theta$  et donc d'estimation rapide. Le résidu  $\hat{u}_t$  de cet ajustement est une bonne approximation de  $Z_t$  et on peut donc estimer  $\text{var}(Z_t)$  par  $s^2 = (1/N) \sum_t \hat{u}_t^2$  et le BIC(2.33).

► L'application de MINIC fonctionne donc ainsi :

- L'utilisateur fixe un ordre maximum d'autorégression  $p_\epsilon$  (l'option par défaut de SAS n'est pas pertinente pour toutes les séries). L'ajustement (2.32) est effectué.
- Pour une série de valeurs  $(m, j)$  l'ajustement (2.34) est effectué.
- L'utilisateur retient le couple  $(m, j)$  qui minimise le BIC.

### Notes.

- Parfois le BIC est défini comme (2.33) divisée par  $n$ .
- Le modèle AR d'ordre élevé qu'on ajuste peut donner un ajustement parfait (sur ajustement) quand la série est courte. On soupçonne une telle situation si le BIC prend des valeurs très négatives.
- L'application de cette méthode ne dispense pas d'essayer de simplifier le paramétrage, notamment en cas de saisonnalité.

### Exemple

Simulation d'un  $\text{ARMA}(1, 2)$  :

---

```
data st9.arma1_2 (keep= u date);
  um1 = 0;
  z2=0;
  z1 = 0;
5 do t = -50 to 200;
```

```

    z = 2*rannor( 1436 );
    u = .9 *um1 + z -.1*z1+.9*z2;
    if t > 0 then
    do;
10  date = intnx('month', '31dec1960'd,t);
    format date monyy.;
    output;
    end;
    z2=z1;
15  z1 = z;
    um1=u;
    end;
    run;

```

Etape d'identification avec l'option `minic`, où on précise les ordres extrêmes des autorégressions auxiliaires.

```
proc arima data=st9.armal_2;
i var=u minic perror=(1:11); run;
```

5	Le Système SAS							
	The ARIMA Procedure							
	Critère d'Information Minimum							
10	Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	
	AR 0	1.918625	1.567837	1.577507	1.587355	1.574758	1.591561	
	AR 1	1.706186	1.594322	1.54492	1.562619	1.575613	1.599728	
15	AR 2	1.565707	1.575859	1.55939	1.577493	1.597215	1.620847	
	AR 3	1.579385	1.600155	1.578792	1.601501	1.619983	1.643739	
	AR 4	1.570294	1.595204	1.604108	1.625797	1.646315	1.669871	
	AR 5	1.593425	1.619477	1.627451	1.643971	1.653991	1.678334	
20	Error series model: AR(9)							
	Minimum Table Value: BIC(1,2) = 1.54492							

Une étape d'estimation avec les valeurs suggérées  $p=1$  et  $q=2$ , confirme ensuite (a) la blancheur du bruit et (b) qu'il n'y a pas de paramètre non significatif.

```
e p=1 q=2; run;
```

**Question** Que se passe-t-il si l'on remplace dans le code ci-dessus, la ligne

```
u = .9 *um1 +z -.1*z1+.9*z2;
```

par la ligne

$$u = .9 * um1 + z - .1 * z1 - .9 * z2; ?$$

Il n'est pas nécessaire d'avoir un ordinateur pour répondre à cette question. Mais il est intéressant d'illustrer la réponse en programmant la modification.

## 2.9 Mise en pratique dans SAS

L'identification d'une série ARMA ou ARIMA se fait pas la proc arima.

Il y a 3 commandes dans la proc ARIMA : identify ou i, estimate ou e, forecast ou f.

La mise en œuvre de ces commandes est hiérarchique :

1. `identify` désigne la série à étudier et éventuellement la différence, choix de  $d$  ou  $d_s$  par l'utilisateur. Pour la série éventuellement transformée elle fournit la fonction d'autocorrélation (ACF), la fonction d'autocorrélation partielle (PACF), la fonction d'autocorrélation inverse (IACF), un test de blancheur de la série sous le titre `Autocorrelation Check for White Noise`.

A moins que la série étudiée ne soit déjà un BB, ce test conclut habituellement au rejet de l'hypothèse de blancheur. La commande `identify` peut être suivie de plusieurs commandes

2. `estimate`, on doit écrire une par modèle qu'on veut estimer sur la série définie à la dernière étape `identify`. `estimate` connaît  $d$  et  $D$  de l'étape `identify`, on lui donne des valeurs de  $p, q, P, Q$  et elle fournit les estimations des paramètres  $\phi, \theta, \Phi, \Theta$  et de la variance de  $Z_t$  pour le modèle ainsi spécifié. Elle fournit aussi un test de blancheur de  $\hat{Z}_t$  sous le titre `Autocorrelation Check of Residuals`. Si l'on rejette l'hypothèse de blancheur des résidus il faut essayer un autre ajustement.

Pour chaque modèle estimé on peut vouloir faire plusieurs prévisions d'horizons différents par exemple. D'où une ou des étapes

3. `forecast`, qui fait des prévisions sur le dernier modèle estimé.

En résumé, une étape `forecast` se base sur la dernière étape `estimate`, et une étape `estimate` sur la dernière étape `identify`.

► Chaque étape doit se terminer par `run;` et la `proc arima` par `quit;` . ◀

**Exemples de syntaxe** La table `seriesg` de la librairie SAS `st` contient une variable `xlog` et une variable de date : `ladate`. (`xlog` est le logarithme de la Série G de l'ouvrage de Box-Jenkins ([1]), transport aérien mensuel de passagers de 1949 à 1961).

Identification de la série. La `proc arima` ne faisant de diagrammes séquentiels il est conseillé d'utiliser simultanément le `TIME SERIES VIEWER` pour visualiser les séries et leurs transformations. Menu `SOLUTIONS/ANALYSIS/TIME SERIES VIEWER`. En vue de cette visualisation il est fortement conseillé de créer une variable de date dans la table.

Dessiner l'ACF la PACF et l'IACF

- de la série `xlog` :

```
proc arima data= st.seriesg;
  identify var=xlog;
run;
```

- de la série des accroissements de `xlog` d'un mois sur l'autre et en n'imprimant les autocorrélations que jusqu'au décalage 15 :

```
identify var=xlog(1) nlag=15;
run;
```

- de la série différenciée 1 fois à l'ordre 1 et 1 fois à l'ordre 12 :

```
identify var=xlog(1,12) ;
run;
```

- quitter la `proc arima`

```
quit;
```

La syntaxe complète de ce qu'on vient de faire est donc :

```
proc arima data=st.seriesg;
  identify var= xlog;      run;
  identify var= xlog(1) nlag=15;      run;
  identify var= xlog(1,12) ;      run;
quit;
```

Estimer un ARIMA(2,1,3) sur la série `xlog` par la méthode des moindres carrés inconditionnels et obtenir dans l'output les 3 fonctions d'autocorrélation des résidus :

```
proc arima data=st.seriesg;
  identify var=xlog(1) ;
  run;
  estimate p=2 q=3 method=uls plot;
  run;
```

Après cette étape, estimer un SARIMA(0,1,1)(0,1,1)<sub>12</sub> sur `xlog` en contraignant la constante additive à 0 et par la méthode des moindres carrés conditionnels (par défaut) :

```
  identify var=xlog(1,12) nlag=15;
  run;
  estimate q=(1)(12) noconstant ;
  run;
```

Pour avoir les 3 fonctions d'autocorrélations du résidu il faut ajouter l'option `plot` :

```
estimate q=(1)(12) noconstant plot;
```

Prédire la série à 2 ans et stocker les prévisions, les bornes de l'intervalle de prévision (à 95% par défaut) et les résidus, pour les dates où ils sont définis, dans la table `st.previ` sans les mettre dans l'output (`noprint`):

```
forecast out=st.previ lead=24 id= ladate interval=month noprint;
  run;
quit;
```

L'option `interval=` permet à SAS d'extrapoler la date. L'option `id=` copie la variable désignée ensuite, dans le fichier. Ne pas oublier `quit` pour avoir accès au fichier des prévisions.

**Détails sur la commande `identify`** Pour stocker les fonctions d'autocorrélation calculées, option `outcov=`. Faire un test de racine unité : `stationarity=` pour faire un test de Dickey-Fuller sur la série `x` de la table `SeriesA`

```
proc arima data=SeriesA;
  identify var=x stationarity=(adf=(5,6,7,8));
  run;
```

Tester que `lc` a une racine unité saisonnière de période 4 :

```
proc arima data=b ;
  identify var=lc stationarity=(adf=(2,5) dlag=4);
  run;
```

**Détails sur la commande estimate**

- Ajustement à trous. Ajuster un ARMA(3,2) avec comme seuls coefficients non nuls :  $\phi_1, \phi_3$  et  $\theta_2$ , on écrira la commande :  
estimate p=(1,3) q=(2) ;
- Calculer et stocker les résidus  $\hat{Z}_t$  d'un modèle dont on a fixé les paramètres préalablement, option noest. Par exemple, calculer les résidus du modèle

$$y_t = 0.5 + \frac{1 - .7B}{1 + .9B} Z_t$$

```
proc arima data=a;
identify var=y ;
run;
estimate p=1 q=1 ar=(-.9) ma=(.7) mu= .5 noest;
run;
forecast out=b1 lead=0 noprint ;
run;
```

**Critère d'information d'Akaike.** L'AIC (Akaike's Information Criterion) est :

$$AIC = -2 \log(L) + 2k$$

où  $L$  est la fonction de vraisemblance évaluée en les estimations des paramètres et  $k$  est le nombre de paramètres estimés. (On rencontre aussi une expression de l'AIC divisée par  $n$ .) Si on a supposé le processus AR( $p$ ) gaussien, de bruit blanc  $N(0, \sigma_Z^2)$ , l'AIC prend la forme :

$$AIC(p) = n \times \ln(\hat{\sigma}^2) + 2p$$

où  $\hat{\sigma}^2$  est l'estimation MV de  $\sigma_Z^2$ . Pour utiliser ce critère, on calcule l'AIC pour différentes valeurs de  $p$  et on retient l'ordre donnant l'AIC le plus faible.

**Notes**

1. L'AIC est formé de deux termes :

- a)  $-2 \log(L)$  qui est d'autant plus faible que l'ajustement MV est bon,
- b)  $2k$  qui pénalise cette faible valeur par le nombre de paramètres estimés.

2. Les critères d'information, AIC, BIC ou autres ou la PACF ne suggèrent pas nécessairement les mêmes ordres pour une série donnée.

3. L'AIC comme le BIC permet de comparer des modèles non emboîtés, AR ou autres.

**2.9.1 Modélisation du niveau du Lac Huron**

Le niveau du lac Huron examiné au chapitre 1 montre :

- une tendance légèrement décroissante,
- une variabilité plus grande aux dates récentes qu'au début de la série (hétéroscédasticité),
- enfin on observe que les résidus de l'ajustement d'une droite par MCO au niveau, à deux dates consécutives sont corrélés positivement (1.15).

La série est courte et nous n'essaierons pas de modéliser l'hétéroscédasticité.

**Objectif.** Prédire le niveau à l'horizon 1.

Si l'objectif est de prédire le niveau l'an prochain, il faut exploiter à la fois la linéarité de la tendance et la corrélation de la série à 2 dates consécutives.

► Régression linéaire du niveau sur l'année et stockage des résidus de cet ajustement.

$$y_t = \beta_0 + \beta_1 t + u_t, \quad t = 1, \dots, T \quad (2.35)$$





Vu la PACF, on choisit d'ajuster un AR(2) au résidu (MINIC suggère autre chose.).

```
proc arima data= lakout;
i var= residmc; run;
e p=2 noint plot; run;
```

On a choisi l'option noint car le résidu étant de moyenne 0, la moyenne estimée est elle-même nulle.

---

The ARIMA Procedure

Conditional Least Squares Estimation

Paramètre	Estimation	Erreur standard	Valeur du test t	Pr. Approx. >  t	Retard
AR1,1	1.00826	0.09819	10.27	<.0001	1
AR1,2	-0.28830	0.09996	-2.88	0.0048	2

10

Variance Estimate 0.467565  
 Std Error Estimate 0.683787  
 AIC 205.59  
 SBC 210.7599  
 Number of Residuals 98  
 \* AIC and SBC do not include log determinant.

15

Correlations of Parameter Estimates

Parameter	AR1,1	AR1,2
AR1,1	1.000	-0.768
AR1,2	-0.768	1.000

25

Autocorrelation Check of Residuals

30

To Lag	Chi-Square	DF	Pr > Khi2	-----Auto-corrélations-----					
6	0.58	4	0.9653	0.016	-0.035	-0.008	0.036	0.051	-0.010
12	5.32	10	0.8686	-0.034	-0.038	0.167	-0.005	-0.097	-0.053
18	6.14	16	0.9865	-0.034	0.007	-0.022	0.037	-0.003	-0.060
24	10.36	22	0.9827	-0.002	-0.141	-0.007	0.006	0.115	0.007

35

Model for variable residmc

40

No mean term in this model.

Autoregressive Factors

Factor 1: 1 - 1.00826 B\*\* (1) + 0.2883 B\*\* (2)

---

45

Le résidu de cet ajustement est un BB. En résumé, le modèle de l'erreur ajusté sur les résidus MCO est

$$u_t = \frac{1}{1 - 1.00826B + 0.2883B^2} Z_t$$

avec  $Z_t \sim BBN(0, 0.467565)$ .

Le modèle de la moyenne obtenu par MCO est d'autre part :

$$y_t = 8.1208 - 0.000066 \text{ date}_t + u_t$$

avec  $\text{var}(u_t) = 1.27756$ . Voir chapitre précédent.

► Estimation simultanée de la moyenne et de la structure d'autocovariance de l'erreur.

---

```
proc autoreg data=st6.lake;
model niveau = date/nlag=2;
output out= lakout residual=residbb rm = residmoy p= pred ;
run;
```

---

Noter qu'on stocke deux résidus :  $rm$  correspond à  $\hat{u}_t$  et  $r$  ou  $residual$  à  $\hat{Z}_t$ . De même  $p=$  permet de stocker la prévision à l'horizon 1 de la série, tenant compte de la dynamique de l'erreur.

Le début du traitement commence par une régression MCO. Ensuite `autoreg` se base sur les résidus de cette régression pour estimer le modèle globalement.

---

#### The AUTOREG Procedure

Dependent Variable      NIVEAU

#### Ordinary Least Squares Estimates

SSE	122.645511	DFE	96
MSE	1.27756	Root MSE	1.13029
SBC	309.266295	AIC	304.09636
Regress R-Square	0.2725	Total R-Square	0.2725
Durbin-Watson	0.4395		

Variable	DF	Estimation	Erreur standard	Valeur du test t	Proba. Approx. >  t
Intercept	1	8.1208	0.1864	43.57	<.0001
DATE	1	-0.000066	0.0000111	-6.00	<.0001

#### Estimations des auto-corrélations

Retard	Covariance	Corrélation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
0	1.2515	1.000000																					
1	0.9531	0.761598																					
2	0.5811	0.464356																					

Preliminary MSE      0.4857

#### Estimations des paramètres autorégressifs

Retard	Coefficient	Erreur standard	Valeur du test t
1	-0.971372	0.099152	-9.80
2	0.275439	0.099152	2.78

#### Yule-Walker Estimates

SSE	44.8531319	DFE	94
MSE	0.47716	Root MSE	0.69077
SBC	220.882938	AIC	210.543068
Regress R-Square	0.0769	Total R-Square	0.7339
Durbin-Watson	1.8818		

Variable	DF	Estimation	Erreur standard	Valeur du test t	Proba. Approx. >  t
Intercept	1	8.2289	0.3631	22.66	<.0001

```
DATE          1      -0.000060    0.0000213          -2.80          0.0062
```

Enfin, on vérifie que le résidu est bien un BB.

```
proc arima data= lakout;
i var= residbb; run;
```

The ARIMA Procedure

```
5      Name of Variable = residbb

Mean of Working Series      0.001795
Standard Deviation          0.676522
10  Number of Observations      98
```

#### Autocorrelations

```
15  Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
      0    0.457682    1.00000  |
      1    0.023577    0.05151  | . |* . |
      2   -0.0044736   -0.00977  | . | . |
20     3    0.0030108    0.00658  | . | . |
      4    0.019645    0.04292  | . |* . |
      5    0.024921    0.05445  | . |* . |
      6   -0.0027022   -0.00590  | . | . |
      7   -0.013830   -0.03022  | . *| . |
25     8   -0.015511   -0.03389  | . *| . |
      9    0.074605    0.16301  | . |***. |
...
"." marks two standard errors
```

30

#### Autocorrelation Check for White Noise

```
35  To      Chi-      Pr >
Lag      Square    DF    Khi2  -----Auto-corrélations-----
      6      0.79     6  0.9923  0.052 -0.010  0.007  0.043  0.054 -0.006
      12     5.38    12  0.9441 -0.030 -0.034  0.163 -0.006 -0.097 -0.057
      18     6.29    18  0.9949 -0.040  0.002 -0.025  0.032 -0.008 -0.065
40     24    10.59    24  0.9916 -0.010 -0.141 -0.010  0.006  0.115  0.015
```

Le modèle est satisfaisant.

En résumé le modèle qu'on a ajusté est :

$$\text{niveau}_t = 8.2289 - 0.000060 \text{ date}_t + u_t \quad (2.36)$$

$$u_t = 0.971372u_{t-1} - 0.275439u_{t-2} + Z_t \quad (2.37)$$

(attention aux signes dans autoreg !) ou

$$\text{niveau}_t = 8.2289 - 0.000060 \text{ date}_t + \frac{1}{1 - 0.971372B + 0.275439B^2} Z_t$$

avec  $\widehat{\text{var}}(Z_t) = 0.47716$ .

On devrait également vérifier la normalité de  $Z_t$ .

C'est un exemple de modèle ARMAX dont la forme générale est :

$$Y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_m x_{mt} + \frac{1 - \theta_1 B - \dots - \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} Z_t \quad Z_t \sim BBN(0, \sigma_Z^2)$$

où les  $x_{it}$  sont des variables explicatives déterministes. Un ARMAX est, à sa moyenne près qui n'est pas constante, un ARMA.

### Exercices.

(1) On a choisi d'ajuster un AR(2) à l'erreur et on a estimé le modèle avec l'option par défaut de `autoreg`, c'est-à-dire la méthode de Yule-Walker. Comparer les ajustements avec erreur AR(1) ou AR(2), faire ces ajustements par maximum de vraisemblance, comparer les AIC obtenus et conclure. Indication. A l'étape `model` ajouter l'option `method=ml` après le `\`.

(2) On considère le modèle

$$Y_t = a + b t + \frac{1 - \theta B}{1 - \phi_1 B - \phi_2 B^2}.$$

Ecrire ce modèle sans utiliser de fraction rationnelle. (Il faut multiplier à gauche et à droite par le dénominateur.)

Si l'erreur a une composante MA, l'estimation d'un ARMAX doit se faire par la `proc arima`. Le même travail que ci-dessus s'obtient alors par la syntaxe :

---

```
proc arima data=st6.lake;
  identify var= niveau CROSSCORR= date noprint ; run;
  estimate input= date p=2 ; run;
  forecast out=arimax id= date ; run;
quit;
```

---

La syntaxe n'est pas très explicite statistiquement, ainsi l'option `crosscorr` de la commande `identify` ne fait pas référence ici à une quelconque corrélation croisée entre deux séries temporelles.

---

The ARIMA Procedure

#### Conditional Least Squares Estimation

	Paramètre	Estimation	Erreur standard	Valeur du test t	Pr. Approx. >  t	Retard Variable
5	MU	8.25482	0.39196	21.06	<.0001	0 NIVEAU
	AR1,1	1.01092	0.09911	10.20	<.0001	1 NIVEAU
10	AR1,2	-0.29015	0.10108	-2.87	0.0051	2 NIVEAU
	NUM1	-0.0000569	0.00002128	-2.67	0.0089	0 DATE

#### Conditional Least Squares Estimation

15 Paramètre Shift

MU	0
AR1,1	0
AR1,2	0
20 NUM1	0

Constant Estimate	2.304953
Variance Estimate	0.476508
25 Std Error Estimate	0.690296
AIC	209.3835
SBC	219.7234
Number of Residuals	98
* AIC and SBC do not include log determinant.	

30

#### Correlations of Parameter Estimates

Variable	NIVEAU	NIVEAU	NIVEAU	DATE
35 Parameter	MU	AR1,1	AR1,2	NUM1

```

NIVEAU      MU      1.000      -0.010      0.095      0.796
NIVEAU      AR1,1    -0.010      1.000      -0.768      -0.007
NIVEAU      AR1,2     0.095      -0.768      1.000      0.095
40  DATE      NUM1     0.796      -0.007      0.095      1.000

```

## Autocorrelation Check of Residuals

```

45  To      Chi-      Pr >
    Lag      Square    DF    Khi2  -----Auto-corrélations-----
        6        0.65     4  0.9573  0.018  -0.039  -0.007  0.034  0.056  -0.008
       12        5.42    10  0.8614 -0.031 -0.034  0.168 -0.006 -0.097 -0.056
50    18        6.25    16  0.9852 -0.036  0.006 -0.024  0.033 -0.006 -0.062
       24       10.71    22  0.9787 -0.003 -0.147 -0.007  0.009  0.115  0.008

```

Model for variable NIVEAU

```

55  Estimated Intercept      8.254817

```

## Autoregressive Factors

```

60  Factor 1:  1 - 1.01092 B**(1) + 0.29015 B**(2)

```

## Input Number 1

```

65  Input Variable      DATE
    Overall Regression Factor      -0.00006

```

Les différences entre les estimations des coefficients autorégressifs entre `arima` et `autoreg` sont dues aux algorithmes mis en oeuvre.

► **Attention !** Les  $\phi$  de la `proc autoreg` sont les  $-\phi$  de la `proc arima`.  
Superposition des séries.

```

data b;
merge lakout armax;
by date;
run;

5  proc gplot data=b;
    symbol1 v=dot i=join ;
    symbol2 v= plus i=r color=black;
    symbol3 v= star i=join color=red;
10  plot niveau * date = 3 forecast * date = 2 pred* date=3 / overlay;
    run;

```

## Code R de cette estimation

```

# lecture des données
niveau = scan('D:/Donn/brocddave/lake.dat')
# conversion en série temporelle
niveau = ts(niveau,start=1876,frequency=1)
5  # estimation MCO du niveau
#
temps = 1:length(niveau)
regmco = lm(niveau ~ temps)
# examen des résidus. Si on ne sait où ils sont dans l'objet regmco, taper
10 # str(regmco)

```

```

plot(temps, regmco$residuals, type="l")
# ACF et PACF
# Il existe une fonction acf mais la variante acf2.R par Shumway est plus agréable
source("acf2.R")
15 acf2( regmco$residuals)
# ajustement d'un AR(2) aux résidus
ajustar = arima(regmco$residuals, order = c(2, 0, 0), include.mean = FALSE)
ajustar$coef
# examen de l'ACF des résidus
20 acf2(ajustar$residuals)
# test de blancheur
help(tsdiag)
tsdiag(ajustar)
# ajustement simultané de la moyenne et de la structure de covariance par maximum de vraisemblance
25 ajustsim = arima(niveau, order = c(2, 0, 0), xreg = temps, method="ML")

```

---

Les coefficients du temps ne sont pas du même ordre de grandeur dans SAS et R car le temps n'est pas mesuré de la même façon.

## Obtenir de l'aide dans SAS ou dans R.

Les documentations sont très abondantes et même si on les télécharge, elles ne sont pas facilement utilisables. Voici quelques trucs qui marchent bien.

**Dans R** Utiliser les fonctions `help` et `help.search`.

- Si on connaît le nom de la fonction qu'on veut employer mais qu'on ne sait plus les arguments, utiliser `help`.  
Par exemple, on ne sait plus comment fonctionne `matrix()`, on tape `help(matrix)` dans la fenêtre de commande et on obtient la syntaxe de cette fonction ainsi que des exemples d'utilisation.
- Si on ne sait pas le nom de la fonction qui permet de faire une certaine tâche, utiliser `help.search`.  
Par exemple, on ne sait pas quelle fonction permet d'estimer un arima, on tape `help.search("arima")` et on récupère la liste de toutes les fonctions où ce mot est cité avec le nom du package auquel appartient la fonction. La recherche est faite parmi les packages installés sur la machine où l'on travaille.

**Dans SAS** – Pour avoir une aide sur une proc, cliquer sur son nom dans la fenêtre édition et enfoncer la touche F1.

- Pour avoir des exemples d'une proc, activer dans le Menu Aide / Aide SAS et documentation. On trouve des exemples des proc, rangés par module.

**Sur Internet** Avec un moteur de recherche les mots

`proc arima forecast example`  
donnent des milliers de références le plus souvent pertinentes.

Pour R, on tapera

`arima CRAN example`

## Annexe

### 2.10 Estimation d'un ARMA

On examine dans cette section, comment écrire la fonction de vraisemblance de processus de base : AR(1) et MA(1). Une fois obtenue la fonction de vraisemblance, son optimisation est un problème numérique. Si la fonction de vraisemblance est quadratique, on débouche sur le problème des moindres carrés linéaires classique. Sinon on a un problème non linéaire qui doit se résoudre par un algorithme d'optimisation. Nous n'étudions pas ces algorithmes dans ce cours.

#### 2.10.1 Fonction de vraisemblance d'un processus gaussien AR(1)

On a la série  $y_t, t = 1, 2, \dots, T$ , observation de  $\{Y_t\}$  AR(1) :

$$Y_t = c + \phi Y_{t-1} + Z_t, \quad Z_t \sim BBN(0, \sigma^2), t \in \mathbb{N},$$

(BBN = Bruit Blanc Gaussien). Evaluation de la fonction de vraisemblance. On sait qu'alors  $Y_t$  suit une loi normale et on a déjà calculé ses moyenne et variance :  $E(Y_t) = \mu = \frac{c}{1-\phi}$ ,  $\text{var}(Y_t) = \frac{\sigma^2}{1-\phi^2}$ . Notons  $\theta = (c, \phi, \sigma^2)$  le vecteur des paramètres à estimer. La fonction de densité de probabilité (f.d.p.) de  $Y_1$  est :

$$f_{Y_1}(y_1; \theta) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp\left[-0.5 \frac{(y_1 - c/(1-\phi))^2}{\sigma^2/(1-\phi^2)}\right]$$

Considérons maintenant la loi conditionnelle de  $Y_2$  sachant que  $Y_1 = y_1$ . C'est une loi normale de moyenne  $c + \phi y_1$ , de variance  $\sigma^2$  d'où la f.d.p.

$$f_{Y_2|Y_1=y_1}(y_2; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-0.5 \frac{(y_2 - c - \phi y_1)^2}{\sigma^2}\right]$$

on en déduit la f.d.p. conjointe du couple  $(Y_1, Y_2)$  :

$$f_{Y_1, Y_2}(y_1, y_2; \theta) = f_{Y_1}(y_1; \theta) f_{Y_2|Y_1=y_1}(y_2; \theta)$$

On observe d'autre part que  $Y_t$  ne dépend explicitement que de  $y_{t-1}$  :

$$f_{Y_t|Y_{t-1}=y_{t-1}, Y_{t-2}=y_{t-2}, \dots, Y_1=y_1}(y_t; \theta) = f_{Y_t|Y_{t-1}=y_{t-1}}(y_t; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-0.5 \frac{(y_t - c - \phi y_{t-1})^2}{\sigma^2}\right].$$

La f.d.p. conjointe des observations est donc :

$$f_{Y_1, \dots, Y_T}(y_1, \dots, y_T; \theta) = f_{Y_1}(y_1; \theta) \prod_{t=2}^T f_{Y_t|Y_{t-1}=y_{t-1}}(y_t; \theta).$$

D'où on obtient la (fonction) log vraisemblance

$$\begin{aligned} \mathcal{L}(\theta) = & \\ & -\frac{1}{2} \ln(2\pi \frac{\sigma^2}{1-\phi^2}) - \frac{1}{2} \frac{(y_1 - c/(1-\phi))^2}{\sigma^2/(1-\phi^2)} - \frac{T-1}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{t=2}^T \frac{(y_t - c - \phi y_{t-1})^2}{\sigma^2} \end{aligned} \quad (2.38)$$

- Si on néglige le log déterminant dans cette expression, sa maximisation donne l'estimateur des moindres carrés inconditionnels de  $\theta$ .

- Si on travaille conditionnellement à la première valeur  $y_1$ . Alors la log vraisemblance se simplifie en la *log vraisemblance conditionnelle* :

$$\mathcal{L}_c(\theta) = -\frac{T-1}{2} \ln(2\pi\sigma^2) - \sum_{t=2}^T \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}. \quad (2.39)$$

On reconnaît que (2.39) a la même forme que la log vraisemblance associée à l'estimation de la moyenne et de la variance d'une v.a. normale, basée sur  $T-1$  observations indépendantes de cette v.a.. Dérivant par rapport à  $c$  et  $\phi$  on obtient  $\hat{c}$  et  $\hat{\phi}$  solution de

$$\begin{bmatrix} c \\ \phi \end{bmatrix} = \begin{bmatrix} T-1 & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t \\ \sum y_{t-1}^2 \end{bmatrix}$$

qu'on reporte dans (2.39) puis en dérivant par rapport à  $\sigma^2$  :

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^T (y_t - \hat{c} - \hat{\phi} y_{t-1})^2}{T-1}$$

On appelle *estimateur du maximum de vraisemblance conditionnelle*, la solution de ces deux équations. Si  $T$  est grand les solutions des *maximisations* de (2.38) et (2.39) sont proches.

- Si dans (2.39) on néglige le log déterminant on obtient l'*estimateur des moindres carrés conditionnels* de  $c$  et  $\phi$ .

## 2.10.2 Fonction de vraisemblance d'un processus gaussien MA(1)

Soit un processus gaussien MA(1) :

$$Y_t = \mu + Z_t - \theta Z_{t-1}$$

où  $Z_t$  est un BBN(0,  $\sigma^2$ ),  $Y_t$  est observé sur  $t = 1, 2, \dots, T$ .

Si on connaît  $z_{t-1}$ , la valeur de  $Z_{t-1}$ , alors la loi de  $Y_t$  sachant que  $Z_{t-1} = z_{t-1}$  est  $N(\mu - \theta z_{t-1}, \sigma^2)$ . Supposons que  $Z_0 = 0$ , alors étant donné l'observation de  $Y_1$  on peut déduire la valeur de  $Z_1$  :  $z_1 = y_1 - \mu$ . Ensuite  $Y_2 = \mu + Z_2 - \theta z_1$  permet d'obtenir  $z_2 = y_2 - \mu + \theta z_1$ . On obtient ainsi la loi conditionnelle de  $Y_2$  sachant que  $Z_0 = 0, Y_1 = y_1$ . Sa f.d.p. est :

$$f_{Y_2|Y_1=y_1, Z_0=0}(y_2; \theta, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-0.5 \frac{(y_2 - \mu + \theta z_1)^2}{\sigma^2}).$$

Ensuite connaissant  $z_1$  et  $y_2$  on peut calculer  $z_3 = y_2 - \mu + \theta z_1$ ... Ainsi ayant fixé la valeur de  $Z_0$  (ici à la moyenne) et disposant des observations  $y_1, \dots, y_T$  on peut calculer *pour chaque valeur de*  $(\theta, \mu, \sigma^2)$  :  $z_1 = y_1 - \mu$ ,  $z_2 = y_2 - \mu + \theta z_1, \dots, z_t = y_t - \mu + \theta z_{t-1}$  et la distribution conditionnelle de  $Y_t|Y_{t-1}=y_{t-1}, \dots, Y_1=y_1, Z_0=0$ . Sa f.d.p. est :

$$f_{Y_t|Y_{t-1}=y_{t-1}, \dots, Y_1=y_1, Z_0=0}(y_t; \theta, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[\frac{1}{2\sigma^2} (y_t - \mu + \theta z_{t-1})^2].$$

La f.d.p. conjointe de  $Y_1, \dots, Y_T|Z_0=0$  est :

$$f_{Y_1|Z_0=0}(y_1; \theta, \mu, \sigma^2) \prod_{t=2}^T f_{Y_t|Y_{t-1}=y_{t-1}, \dots, Y_1=y_1, Z_0=0}(y_t; \theta).$$

La log vraisemblance est :

$$\mathcal{L}(\theta, \mu, \sigma^2) = -\frac{T}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu + \theta z_{t-1})^2$$



## 2.11 Prédiction d'une série temporelle

On s'intéresse ici à la *prédiction après modélisation*. Trouver un bon modèle pour une série est une opération qui peut prendre beaucoup de temps. Or il n'est pas rare d'avoir des milliers de séries à prédire en quelques heures. La recherche d'un modèle pour chaque série n'est évidemment pas possible dans ce cas. On fait alors de la *prédiction sans modèle*. Le logiciel (SAS par la proc forecast, R avec le package forecast notamment) cherche dans une classe de modèles qu'on lui indique le modèle qui minimise une certaine erreur, par exemple l'erreur quadratique moyenne, pour l'horizon de prédiction qu'on lui indique. Dans cette section, comme dans le chapitre tout entier, nous ne considérons que de la prédiction après modélisation.

### Objectif

On veut prédire les valeurs futures  $Y_{t+1}, \dots, Y_{t+h}$  d'une série  $\{Y_t\}$  à partir des valeurs observées  $\{y_t, y_{t-1}, \dots\}$  et on a estimé un bon modèle pour cette série. La prédiction de  $Y_{t+1}$  connaissant  $Y_t, Y_{t-1}, \dots$  est l'espérance conditionnelle  $E(Y_{t+1}|Y_t, Y_{t-1}, \dots)$ .

On suppose dans cette présentation sommaire que le processus  $\{Y_t\}$  est gaussien et donc qu'espérance conditionnelle et espérance conditionnelle linéaire coïncident. On considère ici le cas où  $\{Y_t\}$  est un ARMA(p,q) dont les paramètres sont connus. Dans la pratique le modèle est estimé et on applique la méthode qu'on va voir en remplaçant dans les expressions, les paramètres par leurs estimations.

#### 2.11.1 Prédiction à erreur quadratique minimum pour un ARMA

Considérons un processus  $\{Y_t\}$ , ARMA(p,q) :

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \dots - \theta_q Z_{t-q} \quad (2.40)$$

où  $Z_t \sim \text{BB}(0, \sigma_Z^2)$ . Avec l'opérateur retard cette équation s'écrit aussi :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \phi_0 + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) Z_t \quad (2.41)$$

On suppose  $\{Y_t\}$  stationnaire ( les racines du polynôme  $1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$  sont en module  $> 1$  ), et inversible (les racines du polynôme  $1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q$  sont en module  $> 1$ ). Les paramètres  $\phi_i, \theta_j$  et  $\sigma_Z^2$  sont connus.

Commençons par centrer le processus. On voit sur (2.40) que  $\mu = E(Y_t)$  vérifie :

$$\mu = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p}.$$

Nous simplifions l'écriture en notant encore  $Y_t$ , le processus centré  $Y_t - \mu$ . Avec les hypothèses de stationnarité et inversibilité, le processus admet une représentation MA( $\infty$ ) :

$$Y_t = Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots \quad (2.42)$$

avec  $\psi_0 = 1$ . C'est-à-dire,  $Y_t$  s'écrit comme une combinaison linéaire (c.l.) des valeurs  $Z_t, Z_{t-1}, \dots$ . Considérons d'autre part sur l'écriture AR( $\infty$ )

$$Y_t = Z_t + \sum_{j=1}^{\infty} \pi_j Y_{t-j}$$

on voit que  $Z_t$  s'écrit comme une c.l. des  $Y_t, Y_{t-1}, \dots$ . Ainsi, l'espace engendré par les c.l. de  $Y_t, Y_{t-1}, \dots$  est le même que celui engendré par les c.l. de  $Z_t, Z_{t-1}, \dots$ . Notamment :

$E(Z_{n+j}|Y_n, Y_{n-1}, \dots) = E(Z_{n+j}|Z_n, Z_{n-1}, \dots)$  et donc

$E(Z_{n+j}|Y_n, Y_{n-1}, \dots) = 0$  si  $j > 0$ ,  $= Z_{n+j}$  si  $j \leq 0$ .

On a observé le processus jusqu'au temps  $n$  et on veut le prédire à l'horizon  $l$ , c'est-à-dire prédire  $Y_{n+l} = \sum_{j=0}^{\infty} \psi_j Z_{n+l-j}$  à partir de  $Y_n, Y_{n-1}, \dots$   $l > 0$ . La prédiction est une fonction linéaire de

$Y_n, Y_{n-1}, \dots$ . La prévision à l'horizon  $l$  étant donné l'observation jusqu'au temps  $n$  peut donc aussi s'écrire comme une fonction linéaire de  $Z_n, Z_{n-1}, \dots$  :

$$\hat{Y}_n(l) = \psi_l^* Z_n + \psi_{l+1}^* Z_{n-1} + \psi_{l+2}^* Z_{n-2} + \dots$$

où les  $\psi_i^*$  sont à déterminer. L'erreur quadratique de prévision (EQP) est :

$$\mathbf{E}(Y_{n+l} - \hat{Y}_n(l))^2 = \sigma_Z^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma_Z^2 \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2.$$

l'EQP est minimisée quand  $\psi_{l+j}^* = \psi_{l+j}$ ,  $j = 0, 1, \dots$  donc :

$$\hat{Y}_n(l) = \psi_l Z_n + \psi_{l+1} Z_{n-1} + \psi_{l+2} Z_{n-2} + \dots$$

**Erreur de prévision** L'erreur de prévision est

$$e_n(l) := Y_{n+l} - \hat{Y}_n(l) = \sum_{j=0}^{l-1} \psi_j Z_{n+l-j}.$$

On vérifie que  $\mathbf{E}(e_n(l)) = 0$ , le prédicteur est sans biais. D'autre part,  $\text{var}(e_n(l)) = \sigma_Z^2 \sum_{j=0}^{l-1} \psi_j^2$ . Enfin, ayant supposé la normalité du bruit blanc, on peut fabriquer des intervalles de prévision : les limites de prévision à  $(1 - \alpha)100\%$ ,  $\alpha$  fixé entre 0 et 1, sont :

$$\hat{Y}_n(l) \pm z_{1-\alpha/2} \sqrt{1 + \sum_{j=1}^{l-1} \psi_j^2} \sigma_Z.$$

Les calculs ci-dessus sont simples car tout est exprimé en fonction du bruit blanc  $Z_t$ , mais ils ne sont pas directement utilisables car  $Z_t$  n'est pas observé. On envisage maintenant une expression plus utilisable de la prévision.

### Calcul de la prévision

Considérons le processus ARMA(p,q) centré examiné précédemment :

$$Y_{n+l} = \phi_1 Y_{n+l-1} + \phi_2 Y_{n+l-2} + \dots + \phi_p Y_{n+l-p} + Z_{n+l} - \theta_1 Z_{n+l-1} - \theta_2 Z_{n+l-2} - \dots - \theta_q Z_{n+l-q}.$$

Prenant l'espérance conditionnelle au passé  $Y_n, Y_{n-1}, Y_{n-2}, \dots$ , nous obtenons :

$$\begin{aligned} \hat{Y}_n(l) &= \phi_1 \hat{Y}_n(l-1) + \phi_2 \hat{Y}_n(l-2) + \dots + \phi_p \hat{Y}_n(l-p) \\ &\quad + \hat{Z}_n(l) - \theta_1 \hat{Z}_n(l-1) - \theta_2 \hat{Z}_n(l-2) - \dots + \theta_q \hat{Z}_n(l-q). \end{aligned} \quad (2.43)$$

Nous savons d'autre part que :

$$\hat{Y}_n(j) = \mathbf{E}(Y_{n+j} | Y_n, Y_{n-1}, \dots), \text{ si } j \geq 1, \text{ et } \hat{Y}_n(j) = Y_{n-j}, \text{ si } j \leq 0.$$

et  $\hat{Z}_n(j) = 0$ , si  $j \geq 1$ , et  $\hat{Z}_n(j) = Y_{n+j} - \hat{Y}_{n+j-1}(1)$ , si  $j \leq 0$ . On peut donc par récurrence calculer les prévisions à l'horizon 1, 2,  $\dots$

### 2.11.2 Exemple

Considérons un ARMA(1,1) :

$$(1 - \phi B)(Y_t - \mu) = (1 - \theta B)Z_t.$$

Calcul de  $\hat{Y}_n(l)$ .

$$Y_{n+l} = \mu + \phi(Y_{n+l-1} - \mu) + Z_{n+l} - \theta Z_{n+l-1} \text{ donc}$$

$$\hat{Y}_n(1) = \mu + \phi(Y_n - \mu) - \theta Z_n$$

et

$$\hat{Y}_n(l) = \mu + \phi(\hat{Y}_n(l-1) - \mu) = \mu + \phi^l(Y_n - \mu) - \phi^{l-1}\theta Z_n, \quad l \geq 2$$

où  $Z_n = Y_n - \hat{Y}_{n-1}(1)$ .

**Variance de l'erreur de prévision.** L'écriture  $\text{MA}(\infty)$  :  $Y_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j}$  s'obtient en développant,  $(1 - \theta B)/(1 - \phi B)$  :

$$\frac{1 - \theta B}{1 - \phi B} = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots$$

ou

$$(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots) \equiv 1 - \theta B.$$

Identifiant les termes de même puissance de  $B$  à gauche et à droite on obtient :

$$\psi_j = \phi^{j-1}(\phi - \theta), \quad j \geq 1.$$

La variance de l'erreur est donc :

$$\text{var}(e_n(l)) = \sigma_Z^2 \left\{ 1 + \sum_{j=1}^{l-1} [\phi^{j-1}(\phi - \theta)]^2 \right\} \quad (2.44)$$

qui tend vers  $\sigma_Z^2 \frac{(\phi - \theta)^2}{1 - \phi^2}$  quand  $l \rightarrow \infty$ .

Notes.

1. On peut voir sur l'estimation d'un  $\text{MA}(1)$  ci-dessus qu'il y a des choix à faire, notamment dans l'initialisation des résidus. La variété des possibilités fait qu'il est assez difficile de trouver exactement les mêmes résultats avec des logiciels différents.
2. Les logiciels (la proc `arima` de SAS ou la fonction `arima()` de R), font la prévision après modélisation, à partir d'une représentation d'état de la série.

## 2.12 Exemple d'équations de Yule-Walker : cas de l'ARMA(2,1)

Supposons  $(Y_t)$  qui obéit à

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t - \theta_1 Z_{t-1}, \quad t = \dots, 0, 1, 2, \dots \quad (2.45)$$

où les  $Z_t$  sont i.i.d.  $N(0, \sigma_Z^2)$ .

**STATIO** On suppose que les racines de l'équation :

$$1 - \phi_1 z - \phi_2 z^2 = 0$$

sont  $> 1$  en valeur absolue (ou en module si complexes). Ceci nous assure que la série  $(Y_t)$  est stationnaire.

Nous admettons que

$$\text{INNO} \quad \text{cov}(Z_t, Y_{t-k}) = 0, \quad \text{si } k > 0.$$

Maintenant prenons l'espérance mathématique des 2 côtés de (2.45). Par la stationnarité on a  $\mu = E(Y_t) \forall t$  et donc

$$\mu = c + \phi_1 \mu + \phi_2 \mu + 0. \quad (2.46)$$

Par l'hypothèse STATIO,  $1 - \phi_1 - \phi_2 \neq 0$  d'où

$$\mu = \frac{c}{1 - \phi_1 - \phi_2}.$$

Enlevons le CG (côté gauche) de (2.46) au CG de (2.45) et pareil pour les CD. Il nous reste

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + Z_t - \theta_1 Z_{t-1}, \quad t = \dots, 0, 1, 2, \dots \quad (2.47)$$

On va continuer le travail avec le processus centré

$$\dot{Y}_t = Y_t - \mu$$

qui vérifie :

$$\dot{Y}_t = \phi_1 \dot{Y}_{t-1} + \phi_2 \dot{Y}_{t-2} + Z_t - \theta_1 Z_{t-1}, \quad t = \dots, 0, 1, 2, \dots \quad (2.48)$$

Multiplions les 2 côtés de (2.48) par  $Z_t$ . On a, par INNO et par le fait que  $Z_t$  est un BB :

$$\mathbf{E}(Z_t \dot{Y}_t) = \mathbf{cov}(Z_t, \dot{Y}_t) = \phi_1 \mathbf{E}(Z_t \dot{Y}_{t-1}) + \phi_2 \mathbf{E}(Z_t \dot{Y}_{t-2}) + \mathbf{E}(Z_t Z_t) - \theta_1 \mathbf{E}(Z_t Z_{t-1}) \quad (2.49)$$

$$\mathbf{E}(Z_t \dot{Y}_t) = 0 + 0 + \sigma_Z^2 + 0 \quad (2.50)$$

comme  $Z_t$  est centré, ces espérances sont aussi des covariances. De même

$$\mathbf{E}(Z_{t-1} \dot{Y}_t) = \mathbf{cov}(Z_{t-1}, \dot{Y}_t) = \phi_1 \mathbf{E}(Z_{t-1} \dot{Y}_{t-1}) + \phi_2 \mathbf{E}(Z_{t-1} \dot{Y}_{t-2}) + \mathbf{E}(Z_{t-1} Z_t) - \theta_1 \mathbf{E}(Z_{t-1} Z_{t-1}) \quad (2.51)$$

$$\mathbf{E}(Z_{t-1} \dot{Y}_t) = \phi_1 \sigma_Z^2 - \theta_1 \sigma_Z^2 = (\phi_1 - \theta_1) \sigma_Z^2 \quad (2.52)$$

Ensuite on voit que

$$\mathbf{E}(Z_{t-k} \dot{Y}_t) \underset{1}{=} \mathbf{cov}(Z_{t-k}, \dot{Y}_t) \underset{2}{=} \mathbf{cov}(Z_{t-k}, Y_t) \underset{3}{=} 0 \quad \forall k > 1$$

$\underset{1}{=}$  et  $\underset{2}{=}$  parce que  $Z_t$  est centré,

$\underset{3}{=}$  parce que la partie MA est d'ordre 1.

On peut commencer à écrire les équations de Yule-Walker. On note

$$\gamma_k = \mathbf{cov}(Y_t, Y_{t-k}) \equiv \mathbf{cov}(\dot{Y}_t, \dot{Y}_{t-k}) \quad \rho_k = \gamma_k / \gamma_0.$$

Multiplions les 2 côtés de (2.48) par  $\dot{Y}_t$ .

$$\dot{Y}_t \dot{Y}_t = \phi_1 \dot{Y}_t \dot{Y}_{t-1} + \phi_2 \dot{Y}_t \dot{Y}_{t-2} + \dot{Y}_t Z_t - \theta_1 \dot{Y}_t Z_{t-1} \quad (2.53)$$

Prenons les espérances de chaque côté :

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_Z^2 - \theta_1 (\phi_1 - \theta_1) \sigma_Z^2 \\ \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + (1 - \theta_1 (\phi_1 - \theta_1)) \sigma_Z^2 \end{aligned} \quad (2.54)$$

Multiplions les 2 côtés de (2.48) par  $\dot{Y}_{t-1}$ .

$$\dot{Y}_{t-1} \dot{Y}_t = \phi_1 \dot{Y}_{t-1} \dot{Y}_{t-1} + \phi_2 \dot{Y}_{t-1} \dot{Y}_{t-2} + \dot{Y}_{t-1} Z_t - \theta_1 \dot{Y}_{t-1} Z_{t-1} \quad (2.55)$$

et prenons les espérances de chaque côté :

$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1 - \theta_1 \sigma_Z^2 \quad (2.56)$$

Multiplions les 2 côtés de (2.48) par  $\dot{Y}_{t-k}$  où  $k > 1$ .

$$\dot{Y}_{t-k}\dot{Y}_t = \phi_1\dot{Y}_{t-k}\dot{Y}_{t-1} + \phi_2\dot{Y}_{t-k}\dot{Y}_{t-2} + \dot{Y}_{t-k}Z_t - \theta_1\dot{Y}_{t-k}Z_{t-1} \quad (2.57)$$

Prenons les espérances de chaque côté :

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} \quad (2.58)$$

Par exemple, pour  $k = 2$  :

$$\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 \quad (2.59)$$

Divisons (2.54), (2.56) et (2.59) par  $\gamma_0$ . Nous obtenons des équations de Yule-Walker pour un ARMA(2,1).

$$1 = \phi_1\rho_1 + \phi_2\rho_2 + (1 - \theta_1(\phi_1 - \theta_1))\frac{\sigma_Z^2}{\gamma_0} \quad (2.60)$$

$$\rho_1 = \phi_1 + \phi_2\rho_1 - \theta_1\frac{\sigma_Z^2}{\gamma_0} \quad (2.61)$$

$$\rho_2 = \phi_1\rho_1 + \phi_2$$

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \quad k > 1 \quad (2.62)$$

Observons que pour un AR(2), elles se réduisent à

$$1 = \phi_1\rho_1 + \phi_2\rho_2 + \frac{\sigma_Z^2}{\gamma_0} \quad (2.63)$$

$$\rho_1 = \phi_1 + \phi_2\rho_1 \quad (2.64)$$

$$\rho_2 = \phi_1\rho_1 + \phi_2 \quad (2.65)$$

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \quad k = 1, 2, \dots \quad (2.66)$$

Propriété générale qu'on vient d'illustrer sur un ARMA(2,1) : pour un ARMA(p,q), à partir de l'ordre  $q + 1$  la fonction d'autocorrélation suit la même récurrence que celle de l'AR(p) correspondant.

Comment se sert-on des équations de Yule-Walker ?

Cas d'un AR(2).

- Si on en connaît les paramètres  $\phi$ , on peut calculer la fonction d'autocorrélation théorique.
- Si on dispose d'estimations des  $\rho_k$  on peut résoudre le système des 2 équations linéaires (2.64) et (2.65) pour trouver des estimations de  $\phi_1$  et  $\phi_2$ . Ces estimations peuvent servir de valeurs initiales pour la méthode du maximum de vraisemblance.

Cas d'un ARMA(2,1).

- Si on en connaît les paramètres  $\phi$ ,  $\theta$  et  $\sigma_Z^2$ , par (2.54), (2.56), (2.58) on peut dessiner la fonction d'autocorrélation théorique.
- Si on dispose d'estimations des  $\rho_k$  on peut trouver des estimations des paramètres  $\phi$ . Les équations de Y-W ne sont pas directement utilisées pour trouver des estimations des paramètres  $\theta$  dans les ARMA. On peut voir comment SAS fait en consultant le document en ressource dans le dernier e-thème proc arima.pdf, p. 420.

**Exemple numérique.** On définit un AR(2).

$$Y_t = -0.7Y_{t-1} - 0.8Y_{t-2} + Z_t$$

avec  $Z_t \sim BBN(0, \sigma^2 = 2)$

A-t-on bien choisi un modèle stationnaire ? On calcule les racines de  $1 + 0.7x + 0.8x^2 = 0$ .

```
a= polyroot(c(1, .7, .8))
Mod(a)
> Mod(a)
[1] 1.1180 1.1180
```

les racines sont en module  $> 1$  donc on a bien défini une série stationnaire. Par les formules (2.21 et 2.22) on obtient :  $\rho_0 = 1$ ,  $\rho_1 = -.7/(1 + .8) = -0.38889$ .

Maintenant on peut utiliser la récurrence pour calculer  $\rho_k$ ,  $k > 1$

```
rho = rep(NA, 11)
rho[1] = 1
rho[2] = -0.38889
for( i in 3:11)
rho[i] = phi1*rho[i-1] + phi2 * rho[i-2]
> # FAC théorique du décalage 0 au décalage 10
> rho
[1] 1.000000 -0.388890 -0.527777 0.680556 -0.054168 -0.506527 0.397903 0.126690
-0.407005 0.183552 0.197118
```

Inversement, simulons une série de 200 valeur AR(2) suivant le modèle ci-dessus, et calculons la FAC empirique correspondante.

```
# Simulation de 200 valeurs
sersim = arima.sim(n = 200, list(ar = c(-0.7, -0.8)), sd = sqrt(2))

# FAC empirique
acf.2 = acf(sersim, lag.max = 10, plot=FALSE, type = "correlation")
acf.2
Autocorrelations of series 'sersim', by lag

    0      1      2      3      4      5      6      7      8      9     10
1.000 -0.336 -0.596 0.606 0.135 -0.552 0.169 0.328 -0.301 -0.098 0.274
```

Par exemple, -0.336 est une estimation de -0.388890.

Maintenant formons les équations de Yule-Walker pour un AR(2) ( formule 2.62 pour  $k = 2, 3$ ).

$$\begin{bmatrix} \rho_2 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} \rho_1 & 1 \\ \rho_2 & \rho_1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \quad (2.67)$$

D'autre part nous avons une estimation des autocorrélations d'ordre 1, 2, ... On les reporte dans le système ci-dessus et on obtient les estimateurs de Yule-Walker de  $\phi_1$  et  $\phi_2$

```
# equations de Y-W pour estimer phi1 et phi2
# system lineaire
# snd membre
B = matrix( c(-0.596, 0.606), nrow=2, byrow=TRUE)
B
# matrice des coefficients
A = matrix(c(-0.336, 1, -0.596, -0.336), nrow=2, byrow=TRUE)
A

Phichap = solve(A,B)
```

```
Phichap
```

```
> Phichap
```

```
      [,1]  
[1,] -0.57236  
[2,] -0.78831
```

```
# estimation de (-.7, -.8)
```