La GREAT formation

Groupe de recherche en économie appliquée et théorique

N° 005 "Réfléchir à changer" Mai 2015

Modèles de séries temporelles

Massa Coulibaly

BP. E1255 Bamako (Mali) Tel. (+223) 66 74 22 20 Email. great@greatmali.net

Table des matières

Introduction

Une série chronologique ou série temporelle ou encore chronique est une succession finie d'observations (de valeurs) d'une variable qui change au fil du temps (en jour, semaine, mois, trimestre, année, etc.), e.g.

- le chiffre d'affaire annuel d'une entreprise
- l'indice trimestriel des prix
- la consommation journalière d'électricité
- etc.

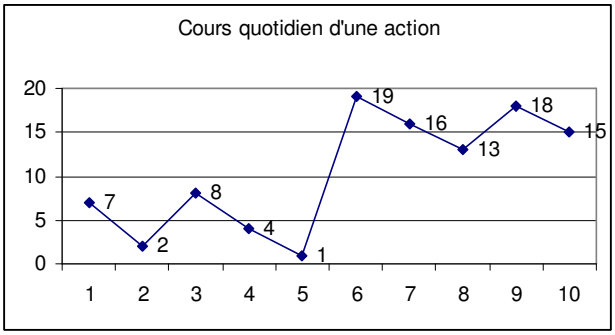
Les composantes de la série sont:

- le trend ou tendance i.e. l'évolution de long terme de la variable, évolution liée à la croissance (ou décroissance) générale du phénomène
- les variations saisonnières ou fluctuations périodiques (les phénomènes se produisant à date fixe) plus ou moins régulières et dont les causes peuvent être:
  - ✓ le cycle des saisons
  - ✓ le mode de vie
  - ✓ la coutume ou la religion
  - ✓ etc.
- les fluctuations résiduelles ou accidentelles qui sont des fluctuations aléatoires dues à des évènements occasionnels imprévisibles tels que grève, sécheresse, évènements politiques ou militaires, krach financier, etc.

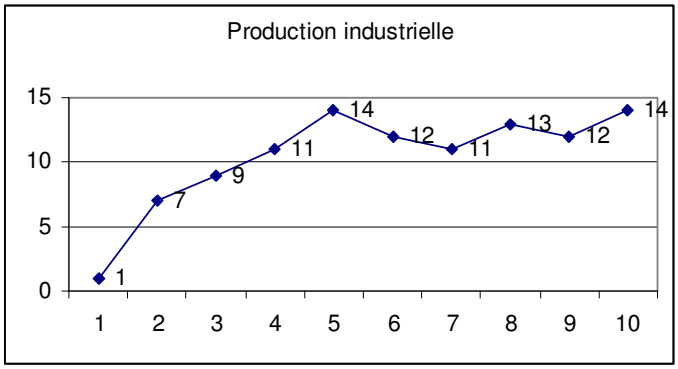
Il faut faire remarquer que certaines séries chronologiques peuvent être sujettes à des fluctuations pour lesquelles les hauts et les bas apparaissent avec d'autres fréquences que celles des variations saisonnières. De telles fluctuations dites cycliques sont liées à des variations conjoncturelles telles que les phases du cycle économique: prospérité, crise, dépression, reprise.

On distingue 3 représentations graphiques possibles de série temporelle:

- sans trend ni saisonnalité

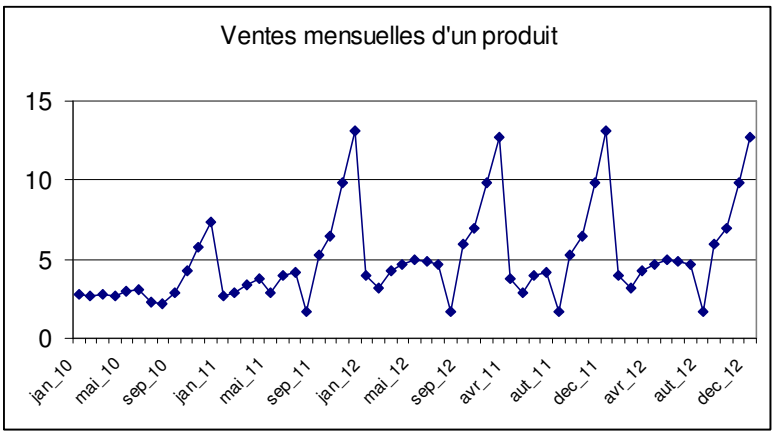


- avec trend mais sans saisonnalité



- avec trend et saisonnalité e.g. les ventes mensuelles d'un produit consignées dans le tableau ci-dessous:

	jan	fev	mar	avr	mai	jun	jul	aut	sep	oct	nov	dec
2010	3	3	3	3	3	3	2	2	3	4	6	7
2011	3	3	3	4	3	4	4	2	5	6	10	13
2012	4	3	4	5	5	5	5	2	6	7	10	13



Qu'est-ce qu'on fait avec les séries temporelles

- estimer des paramètres de différents modèles
- vérifier la qualité des ajustements
- comparer différents modèles
- comprendre et expliquer les variations observées
- reconstituer des valeurs manquantes
- effectuer des prédictions.

L'analyse des séries chronologiques nécessite que soient isolées et mesurées les différentes composantes de sorte que la direction dans laquelle les évènements sont conduits puisse être déterminée et que les valeurs futures puissent être prédites.

Les données sont présentées dans un tableau à double entrée:

i $i = \overline{1,\ n}$ e.g. l'année

j $j = \overline{1,\ p}$ e.g. le trimestre

	1	2	…	j	…	p
1						
2						
⋮						
I				X_{ij}		
⋮						
N						

Les données brutes du tableau à double entrée sont reprises dans un tableau simple en vue des calculs et de l'analyse des résultats:

	X
1	X_1
2	X_2
3	X_3
⋮	⋮
t	X_t
⋮	⋮
T = np	X_T

t = (i-1)p+j

	1	2	…	j	…	p
1	1	2	…	j	…	p
2	p+1	p+2	…	p+j	…	2p
3	2p+1	2p+2	…	2p+j	…	3p
⋮						
i	(i-1)p+1	(i-1)p+2	…	(i-1)p+j	…	ip
⋮						
n	(n-1)p+1	(n-1)p+2	…	(n-1)p+j	…	np

1. La décomposition d'une série chronologique

Décomposer une série c'est estimer pour chaque observation les valeurs de ses différentes composantes. Il existe 2 méthodes principales de détermination du trend et aussi 2 méthodes pour déterminer les variations saisonnières ainsi que les fluctuations résiduelles.

1.1. Le trend

Ses deux méthodes de détermination sont:

- la méthode de la régression pour calculer l'équation du trend représentée par une fonction analytique précise (droite, polynôme, fonction logistique, fonction exponentielle, etc.), e.g.

$$X_t = \frac{C}{1 + e^{-(\alpha+\beta t)}}$$

  L'inconvénient avec cette méthode est qu'il est difficile d'associer une allure simple au trend.

- la méthode de la moyenne mobile (MA – Moving average) qui ne suppose aucune hypothèse sur l'allure du mouvement extra-saisonnier.

Soit le tableau d'identification des moyennes mobiles pour n années et p périodes

	X			
1	X_1			
⋮	⋮			
i	X_i	M_1		
⋮	⋮		M_2	
p	X_p			
p+1	X_{p+1}			
⋮	⋮			
p+i	X_{p+i}		M_{p+1}	
⋮	⋮			M_{p+2}
p+p=2p	X_{2p}			
2p+1	X_{2p+1}			
⋮	⋮			
2p+i	X_{p+i}			M_{2p+1}
⋮	⋮			
2p+p=3p	$X3_p$			
⋮	⋮			
⋮	⋮			M_k
⋮	⋮			
(n-1)p+1	$X_{(n-1)p+1}$			
⋮				$M_{(n-1)p+1}$
(n-1)p+i	$X_{(n-1)p+i}$			
⋮				
(n-1)p+p=np=T	X_{np}			

On appelle moyenne mobile de longueur p de la série X: X_1, X_2, …, X_t, …, X_T la grandeur:

$$M_k = \frac{1}{p}(X_k + X_{k+1} + \cdots + X_{k+(p-1)}) = \frac{1}{p}\sum_{j=0}^{p-1} X_{k+j} = \frac{1}{p}\sum_{j=1}^{p} X_{(k-1)+j}$$

avec $k = \overline{1,\ (n-1)p+1} = \overline{1,\ np-(p-1)}$

La valeur M_k occupe le milieu de la série de p observations i.e. $k + \frac{p-1}{2}$, on peut donc la remplacer par: $M_{k+\frac{p-1}{2}}$

Ainsi, la moyenne mobile correspondra, soit:

- (si p impair) à l'une des observations

- (si p pair) au centre de l'intervalle séparant 2 observations consécutives, auquel cas, il est malaisé d'obtenir une série (celle des moyennes mobiles) ne se rapportant pas aux mêmes dates que la série originelle. On affecte dans ce cas à l'observation initiale la moyenne arithmétique des 2 moyennes mobiles qui l'encadrent.

  Exemple

t	X	Moyenne mobile	
		p = 3	p = 4
1	100	-	-
2	110	110	-
3	120	100	100 / 110 → 105
4	70	110	-
5	140	-	

La moyenne mobile, telle que définie, satisfait deux propriétés:

P_1 Comme toute moyenne arithmétique, la moyenne mobile est un opérateur linéaire:

$$M(X_t + Y_t) = M(X_t) + M(Y_t) \text{ et } M(\lambda X_t) = \lambda\, M(X_t)$$

P_2 Soit le modèle $X_t = C_t + S_t + u_t$ avec $u_1, u_2, \cdots, u_t$ une suite de v.a. indépendantes telles que: $E(u_t) = 0 \; et \; V(u_t) = \sigma^2 \; \forall \; t$

La moyenne mobile de longueur (2k + 1) des u_t est: $\eta_t = \frac{1}{2k+1} \sum u_{t+p}$

On a:

$$E(\eta_t) = \frac{1}{2k+1} \sum E(u_{t+p}) = 0 \quad \text{et}$$

$$V(\eta_t) = \frac{1}{(2k+1)^2} \sum V(u_{t+p}) = \frac{\sum \sigma^2}{(2k+1)^2} = \frac{(2k+1)\sigma^2}{(2k+1)^2} = \frac{\sigma^2}{2k+1}$$

Ainsi, après passage dans le filtre, la variance de η_t est (2k+1) fois plus petite que celle de la série initiale, les fluctuations résiduelles sont donc fortement atténuées. On dit que la moyenne mobile "lisse" la série. L'inconvénient est qu'à la différence des fluctuations u_t, les variables η_t ne sont pas indépendantes. Leur corrélation peut engendrer des mouvements périodiques absents de la série primitive. Ce phénomène est connu sous le nom d'effet Slutsky.

1.2. Variations saisonnières et fluctuations résiduelles

Les 2 méthodes de détermination de ces composantes reposent sur 2 hypothèses distinctes, selon que leur amplitude est affectée ou non par celle du trend:

- (modèle additif) amplitude non affectée par celle du trend:
$$X = T + S + R$$
- (modèle multiplicatif) amplitude proportionnelle à celle du trend:
$$X = TSR$$

Pour choisir le modèle de décomposition, on utilise la dispersion autour de la tendance, dispersion mesurée par l'écart type annuel (ou toute autre période de base selon les types d'observations) :

$$\sigma = \sqrt{\frac{\sum (X - \overline{X})^2}{p}}$$

où p = 4 si la série est trimestrielle et p = 12 si la série est mensuelle.

On représente σ en fonction de la moyenne de X $\left(\overline{X} = \frac{1}{p}\sum X\right)$

Année	i	1	2	…	i	…	n
Moyenne	$\overline{X}$	$\overline{X_1}$	$\overline{X_2}$	…	$\overline{X_i}$	…	$\overline{X_n}$
Dispersion	σ	σ_1	σ_2	…	σ_i	…	σ_n

Pour retenir un schéma additif, il faut que la dispersion (σ) soit constante quelle que soit la moyenne annuelle. En cas de doute, on étudie la valeur de la pente (β) de la droite de régression:

$\sigma = \alpha + \beta\overline{X}$

Si

$\beta < 0.05$ modèle additif

$\beta > 0.10$ modèle multiplicatif

$0.05 < \beta < 0.10$ des 2 modèles, on choisit celui qui donne les meilleurs résultats.

2. Correction des variations saisonnières

2.1. Hypothèses

H_1 Le trend est une fonction quelconque du temps ne présentant pas de retournement de tendance trop marqué

H_2 Les variations saisonnières S sont une fonction périodique de période p, prenant les valeurs $S_1, S_2, \cdots, S_t, \cdots$ tel que: $S_t = S_{t+p}$ et

$$\sum_{t=1}^{p} S_t = 0 \mapsto M_p(S_t) = 0$$

H_3 Les fluctuations résiduelles sont indépendantes du trend et des variations saisonnières, d'espérance nulle et de variance faible: $E(R) = 0$ et $V(R) \to 0$

En réalité les fluctuations résiduelles sont de 2 sortes:

- ✓ celles dues à un grand nombre de petites causes e.g. des erreurs de mesure, qui provoquent des variations de faible amplitude
- ✓ celles résultant d'évènements accidentels isolés e.g. grève, sécheresse, dont l'amplitude est plus grande.

Seul le 1er type de fluctuations résiduelles répond à H_3. C'est pourquoi pour appliquer la méthode, il faut au préalable corriger les données brutes des variations accidentelles de grande amplitude, corriger soit

- ✓ par simple évaluation graphique
- ✓ par estimation directe de l'impact du phénomène accidentel e.g. la perte occasionnée par une journée de grève.

2.2. Calcul des coefficients saisonniers

Pour obtenir les variations saisonnières (S), il est nécessaire de calculer pour chaque période (e.g. le trimestre) la variation moyenne par rapport au trend:

$$X - T = S + R \qquad \text{si modèle additif}$$

$$\frac{X}{T} = SR \qquad \text{si modèle multiplicatif}$$

Ensuite, il faut ajuster ces moyennes par un facteur de correction dont le calcul dépend du modèle retenu, comme indiqué ci-dessous:

Modèle additif	**Modèle multiplicatif**
Par définition: $\sum_{j=1}^{p} S_j = 0$ mais la somme des moyennes ajustées $\left(S_j^{'}\right)$ n'est pas rigoureusement nulle. Il faut les corriger de façon à respecter la relation de définition pour ainsi obtenir les estimations définitives $\left(S_j^{*}\right)$:	Par définition: $\sum_{j=1}^{p} S_j = p$ mais la somme des moyennes ajustées $\left(S_j^{'}\right)$ n'est pas rigoureusement égale à p. Il faut les corriger de façon à respecter la relation de définition pour ainsi obtenir les estimations définitives $\left(S_j^{*}\right)$:
• déterminer le facteur de correction en calculant la moyenne des p estimations $\overline{S^{'}} = \frac{1}{p}\sum_{j=1}^{p} S_j^{'}$	• déterminer le facteur de correction en calculant la moyenne des p estimations $\overline{S^{'}} = \frac{1}{p}\sum_{j=1}^{p} S_j^{'}$
• estimer les coefficients saisonniers $S_j^{*} = S_j^{'} - \overline{S^{'}}$	• estimer les coefficients saisonniers $S_j^{*} = \frac{S_j^{'}}{\overline{S^{'}}}$

2.3. La désaisonalisation

Pour supprimer les effets des variations saisonnières, il est simplement nécessaire de calculer:

- (modèle additif) $X_{ij}^{*} = X_{ij} - S_{j}^{*}$
- (modèle multiplicatif) $X_{ij}^{*} = \frac{X_{ij}}{S_{j}^{*}}$

Si l'estimation des coefficients saisonniers est correcte, alors:

- (modèle additif) $X^{*} = T + R$
- (modèle multiplicatif) $X^{*} = T\,R$

La série corrigée des variations saisonnières (CVS) est une bonne approximation des tendances d'évolution du phénomène observé.

3. Prévision

Il existe 4 principales méthodes de prévision:

- la méthode naïve
- la méthode du lissage
- la méthode de régression linéaire
- la méthode Box-Jenkins ou méthode de pure autorégression.

3.1. La méthode naïve

Elle suppose que les valeurs de la période t (e.g. ce trimestre) seront les mêmes que celles de t-p (soit le même trimestre il y a un an) ajustées par le plus récent trend:

$$X_{t}^{F} = X_{t-p}\frac{X_{t-1}}{X_{t-(p+1)}} \text{ où } \quad \frac{X_{t-1}}{X_{t-(p+1)}} \text{ est le facteur d'ajustement.}$$

En d'autres termes, les valeurs du trimestre t vont être les mêmes que celles du trimestre correspondant il y a un an (soit $t-4$) corrigées du plus récent trend

qui est le rapport des valeurs des trimestres $t-1$ et $(t-4)-1=t-5$:

$$X_t^F = X_{t-4}\frac{X_{t-1}}{X_{t-5}}$$

3.2. Les méthodes de lissage

Ce sont des méthodes simples de prévision de court terme par extrapolation sous l'hypothèse que le phénomène étudié ne dépend que de ses valeurs passées. En général, elles donnent un poids prépondérant aux valeurs récentes, les coefficients de pondération décroissant exponentiellement en remontant dans le temps. Dans leur application, elles dépendent d'un ou de plusieurs paramètres de lissage compris entre 0 et 1 et donnent des prévisions de qualité.

Les méthodes de lissage se distinguent selon que la série comporte ou non un trend ainsi que selon la présence ou non de saisonnalité. Ce sont:

- (pas de trend mais un changement de niveau) lissage exponentiel simple (LES) ou lissage de Brown
- (présence d'un trend à la hausse) le lissage exponentiel double (LED) et le lissage exponentiel de Holt (LEH)
- (présence de saisonnalité) lissage de Winters (LW).

		Saisonnalité	
		Non	*Oui*
Trend	*Non*	LES	LEW
	Oui	LED / LEH	

La plupart de ces techniques de prévision sont implémentées dans des logiciels de traitement statistique des données e.g. SPSS, STATA, EVIEWS, etc. Dans leur écriture, on distingue des modèles à:

- 1 paramètre, pour le lissage exponentiel simple
- 2 paramètres (relatifs au niveau et au trend), lissage de Holt
- 3 paramètres (relatifs au niveau, au trend et à la saisonnalité), lissage de Winters.

Quelle que soit la technique choisie, son application nécessite le choix d'une valeur initiale qui pourra être soit:

- la moyenne de la série
- la première observation (X_1) de la série.

3.2.1. Lissage exponentiel simple

Le modèle utilisé pour cette méthode de prévision de la valeur de la série X_t dans le futur s'écrit:

$$X_t^F = \alpha X_t + (1-\alpha) X_{t-1}^F \quad \text{avec} \quad X_1^F = X_1 \text{ valeur initiale}$$

Par récurrence, il s'en suit: $X_t^F = \alpha \sum_{\tau=0}^{t-1} (1-\alpha)^\tau X_{t-\tau}$

En effet:

$X_1^F = X_1$ *donné* :

$$X_2^F = \alpha X_2 + (1-\alpha) X_1^F = \alpha X_2 + (1-\alpha) X_1$$

$$X_3^F = \alpha X_3 + (1-\alpha) X_2^F = \alpha X_3 + (1-\alpha)[\alpha X_2 + (1-\alpha) X_1]$$
$$= \alpha X_3 + \alpha(1-\alpha) X_2 + (1-\alpha)^2 X_1$$

$$\vdots$$

$$X_t^F = \alpha X_t + \alpha(1-\alpha) X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \cdots + (1-\alpha)^{t-1} X_1$$
$$= \alpha \sum_{\tau=0}^{t-1} (1-\alpha)^\tau X_{t-\tau}$$

On peut encore écrire le modèle sous la forme:

$$X_t^F = \alpha X_t + (1-\alpha) X_{t-1}^F = X_{t-1}^F + \alpha\left(X_t - X_{t-1}^F\right) = X_{t-1}^F + \alpha \varepsilon_t$$

où ε_t est assimilé à une erreur de prévision

Le choix de α dépend de la régularité de la série, il sera

- compris entre 0.1 et 0.3 pour une série assez régulière
- plus élevé si changement de niveau.

Proche de 1, le paramètre donne plus d'importance aux observations récentes, tandis que proche de 0, il renforce l'importance des données passées plus loin dans le temps.

3.2.2. Lissage exponentiel double

Le modèle est de la forme:

$$X_t^F = a_t + b_t = \left(2X_{1t}^F - X_{2t}^F\right) + \frac{\alpha}{1-\alpha}\left(X_{1t}^F - X_{2t}^F\right) \qquad \text{où}$$

$$X_{1t}^F = \alpha \sum_{\tau=0}^{t-1} (1-\alpha)^\tau \; X_{t-\tau} \quad \text{et} \quad X_{2t}^F = \alpha \sum_{\tau=0}^{t-1} (1-\alpha)^\tau \; X_{1(t-\tau)}^F$$

Le lissage exponentiel double est, mathématiquement, un cas particulier du lissage exponentiel de Holt bien que ce dernier en soit une version améliorée.

3.2.3. Lissage exponentiel de Holt

Comme pour le LED, le LEH est une somme de deux paramètres (relatifs au trend respectivement à la pente) tels que:

$$\begin{aligned} X_t^F &= a_t + b_t \\ &= [\alpha X_t + (1-\alpha)(a_{t-1} + b_{t-1})] + [\beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}] \\ &= [\alpha X_t + (1-\alpha)X_{t-1}^F] + [\beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}] \end{aligned}$$

Les valeurs initiales des paramètres (a et b) peuvent être fixées empiriquement ou établies par optimisation.

Dans les applications, il est courant de fixer: $a_1 = X_1 \; et \; b_1 = 0$ avec $\alpha = 0.4 \; et \; \beta = 0.6$

t	X_t	Niveau (a_t)	Pente (b_t)	Prévision $\left(X_t^F\right)$
1	20	20.00	0.00	20
2	24	21.60	0.96	23
3	26	23.94	1.79	26
4	25	25.43	1.61	27
5	30	28.23	2.32	31
6	32	31.13	2.67	34
7	28	31.48	1.28	33
8	32	32.45	1.10	34
9	36	34.53	1.68	36
10	36	36.13	1.63	38
11	35	36.66	0.97	38
12	39	38.18	1.30	39

3.2.4. Lissage de Winters

Il existe deux techniques de prévision selon que la saisonnalité est additive ou multiplicative. Dans le premier cas, le modèle de prévision (LWA) s'écrit:

$X_t^F = a_t + b_t + S_{t-p} \quad si \;\; 1 \le h \le p$ avec h la période de prévision

$$= \left[\alpha\left(X_t - S_{t-p}\right) + (1-\alpha)(a_{t-1} + b_{t-1})\right] + \left[\beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}\right] + \left[\delta(X_t - a_t) + (1-\delta)S_{t-p}\right]$$

$X_t^F = a_t + b_t + S_{t-2p} \quad si \;\; p \le h \le 2p$

Les valeurs initiales des paramètres (a et b) peuvent être fixées à:

- pour le niveau: $a_0 = m_1 - \frac{p}{2} b_0$
- pour la pente: $b_0 = \frac{m_n - m_1}{(n-1)p}$

  avec $\alpha = 0.3 \; et \; \beta = \delta = 0$

où

p la période de la composante saisonnière e.g. p=4 pour les trimestres
n le nombre d'années de la série
m_k la moyenne de l'année k.

- les coefficients saisonniers sont obtenus selon le modèle de décomposition.

Dans le cas du modèle multiplicatif, le LWM s'écrit:

$$X_t^F = (a_t + b_t)S_{t-p} \quad si \;\; 1 \le h \le T$$

$$= \left[\left(\alpha \frac{X_t}{S_{t-p}} + (1-\alpha)(a_{t-1} + b_{t-1})\right) + \left[\beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}\right]\right]\left[\delta \frac{X_t}{a_t} + (1-\delta)S_{t-p}\right]$$

$X_t^F = (a_t + b_t)S_{t-2p} \quad si \;\; p \le h \le 2p$

Les valeurs initiales des paramètres (a et b) peuvent être fixées comme dans LWA, avec cette fois: $\alpha = 0.2 \; et \; \beta = \delta = 0$

L'intervalle de prévision est donné par:

$$X_t^F \pm t_{\varepsilon/2} SE\left(X_t^F\right)\sqrt{1+(h-1)\alpha^2} \qquad \text{où} \qquad SE\left(X_t^F\right)=\sqrt{\frac{\sum_{t=1}^{T}\left(X_t - X_t^F\right)^2}{T-1}}$$

3.3. La méthode de régression sur variables liées

Il faut avoir estimé, au préalable, la relation entre la variable de prévision (Y) et des variables explicatives (au nombre de k-1) $X_2, X_3, \cdots, X_k$, sur la période sur laquelle l'on dispose de données. Soit cette relation linéaire:

(1) $Y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$ qu'on peut mettre sous la forme matricielle:

(2) $$Y = X\beta + u$$

où les

$\beta_i \ \left(i = \overline{1, k}\right)$ et les paramètres de la distribution de u sont inconnus et sont à estimer.

On suppose en outre que cette liaison entre variables reste stable au-delà de la période d'estimation, ce qui peut ne pas être le cas. Ensuite, les variables explicatives sont elles-mêmes à prévoir, sauf si la régression s'est faite sur les valeurs passées de la variable de prévision comme dans les modèles autorégressifs, e.g.: $Y_t = \alpha + \beta Y_{t-1} + u_t$

Pour l'estimation du vecteur β, l'on fait certaines hypothèses (structurelles ou stochastiques) sur les observations ainsi que sur la distribution de u:

- absence de colinéarité des variables exogènes:
  $rang(X) = rang(X'X) = k < n \ \ et \ \ (X'X)^{-1} \exists$
- espérance et variance finies:
  $$\lim_{n \to \infty} \quad M = \frac{1}{n}(X'X) = M_0 \ \ matrice \ \ finie \ \ non \ \text{sin} \ gulière$$
- hypothèse fondamentale: $E(u) = 0$
- homoscédasticité et absence d'autocorrélation des erreurs: $E(uu') = \sigma^2 I_T$
- normalité: la loi de u est une loi normale.

Sous ces hypothèses, la méthode d'estimation des moindres carrés $\min RSS = e'e = \left(Y - X\hat{\beta}\right)\left(Y - X\hat{\beta}\right)$ conduit au résultat:

(3) $$\hat{\beta} = (X'X)^{-1}X'Y$$

résultat qui n'admet a priori aucune restriction sur ses valeurs, sauf que dans la pratique, la théorie économique peut indiquer les signes ou les limites de ces valeurs.

Nous avons: $\hat{Y} = X\hat{\beta}$

Connaissant la relation $Y = X\beta + u$, il est fait des prévisions de Y à l'aide des valeurs futures de X au temps θ, valeurs futures contenues dans le vecteur:

$$F = \begin{pmatrix} 1 \\ X_{2\theta} \\ \vdots \\ X_{k\theta} \end{pmatrix}$$

On obtient: $\hat{Y}_\theta = F'\hat{\beta}$ avec:

$$E(\hat{Y}_\theta) = E(F'\hat{\beta}) = F'E(\hat{\beta}) = F'\beta$$

$$V(\hat{Y}_\theta) = E\left[(F'\hat{\beta} - F'\beta)(F'\hat{\beta} - F'\beta)'\right] = E\left[F'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'F\right] = \sigma_u^2 F'(X'X)^{-1}F$$

Ainsi:

$$\hat{Y}_\theta = F'\hat{\beta}$$
$$\underline{Y_\theta = F'\beta + u_\theta}$$
$$\hat{Y}_\theta - Y_\theta = F'(\hat{\beta} - \beta) - u_\theta = z$$

où z est l'erreur de prévision, avec:

$E(z) = 0$ puisque $E(\hat{\beta}) = \beta$ et $E(u_\theta) = 0$ par hypothèse

$$V(z) = V(F'(\hat{\beta} - \beta) - u_\theta) = V(F'(\hat{\beta} - \beta)) + V(u_\theta) = F'V(\hat{\beta})F + \sigma_u^2 = \sigma_u^2(1 + F'(X'X)^{-1}F)$$

Par conséquent, l'intervalle de prévision est tel que:

$$Y_\theta = \hat{Y}_\theta \pm t_{\varepsilon/2} SE(z) = F'\hat{\beta} \pm t_{\varepsilon/2} \sqrt{\hat{\sigma}_u^2 (1 + F'(X'X)^{-1} F)}$$

3.4. Méthode Box-Jenkins

3.4.1. Processus aléatoires

Il existe deux méthodes d'analyse des séries chronologiques:

- les méthodes basées sur les fréquences (pour lesquelles les amplitudes comptent, soit des modèles de type trigonométrique) e.g. l'analyse spectrale
- les méthodes basées sur la dimension temporelle privilégiant les modèles autorégressifs. La fonction d'autocorrélation, de rang k, se définit par:

$$\rho_k = \frac{COV(X_t,\ X_{t+k})}{\sigma_{X_t}\sigma_{X_{t+k}}} = \frac{E(X_t - E(X_t))E(X_{t+k} - E(X_{t+k}))}{\sigma_{X_t}\sigma_{X_{t+k}}} = \frac{\gamma(k)}{\gamma(0)}$$

Si les caractéristiques stochastiques de la série temporelle se modifient dans le temps, c'est que la série n'est pas stationnaire. Elle est stationnaire lorsque le processus stochastique est invariant. Soit X_t – le processus stochastique, on dira que X_t est stationnaire si:

- $E(X_t) = E(X_{t+k}) = \mu \quad \forall t \, et \, k$
  (la moyenne est constante et indépendante du temps)
- $\sigma_{X_t}^2 = V(X_t) < \infty \qquad \forall \ t$
  (la variance est finie et indépendante du temps)
- $COV(X_t, X_{t+k}) = E(X_t - \mu)(X_{t+k} - \mu) = \gamma(_k)$
  (la covariance est indépendante du temps)

Il en résulte que la série chronologique stationnaire ne comporte ni tendance, ni saisonnalité (d'où l'intérêt de la mise en évidence au préalable de ces 2 composantes) et plus généralement aucun facteur n'évoluant avec le temps.

On distingue: les processus purement aléatoires, la marche aléatoire et les processus AR, ARMA et ARIMA et SARIMA.

i). Processus purement aléatoire

Soit le processus: $X_t = u_t$ où u_t est une v.a. telle que $E(u_t) = 0$

Le processus X_t , dit de "bruit blanc" est tel que $\rho(k) = \begin{cases} 1 & pour\ k = 0 \\ 0 & pour\ k \neq 0 \end{cases}$

Ce processus est presque sans intérêt pour la prévision dès lors que la meilleure prévision est donnée par: $X_{T+\theta} = 0 \quad \forall\ \theta$

Bien sûr, la fonction d'autocorrélation (1) est purement théorique, décrivant un processus stochastique pour lequel l'on ne dispose que d'un nombre limité d'observations. Dans la pratique, l'on se contentera d'une estimation, appelée fonction d'autocorrélation d'échantillonnage:

$$\hat{\rho}(k) = \frac{\sum_{t=1}^{T-k} (X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{T} (X_t - \overline{X})^2} = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \qquad (2)$$

Ainsi, pour savoir si une valeur particulière de $\rho(k)$ est nulle, on utile le test dit de BARTLETT, qui montre que si une série temporelle a été générée par un processus de bruit blanc, alors les coefficients d'autocorrélation d'échantillonnage (pour $k > 0$) sont approximativement distribués selon une loi normale de moyenne nulle et d'écart type $1/\sqrt{T}$.

Pour tester l'hypothèse jointe de tous les coefficients nuls, on utilise la statistique Q de Box et Pierre:

$$Q = T\sum_{k=1}^{K} \hat{\rho}^2(k) \quad \text{distribué (approximativement)} \quad \mapsto \quad \chi^2_{(K\ ddl)}$$

ii). Marche aléatoire

Le processus $X_t = X_{t-1} + u_t$ est dit marche aléatoire (Random walk)

En particulier pour : $X_0 = 0$, alors:

$$\begin{cases} X_1 = u_1 \\ X_2 = X_1 + u_2 = u_1 + u_2 \\ \vdots \\ X_t = \sum_{i=1}^{t} u_i \end{cases}$$

Ainsi: $E(X_t) = t\mu \quad et \quad V(X_t) = t\sigma^2$ et puisque moyenne et variance changent avec t, alors le processus est non stationnaire, mais sa différence première est stationnaire. S'agissant de prix, cela signifierait que les variations de prix sont purement aléatoires.

Lorsque les grandeurs économiques e.g. PIB, taux d'intérêt, etc. suivent un processus aléatoire, alors, la régression de l'une sur l'autre peut conduire à des résultats erronés. Enlever le trend avant de faire la régression ne résoudrait pas le problème puisque les séries détrendées resteront toujours non stationnaires. Seules les différences premières pourraient donner des séries stationnaires.

Mieux, si une grandeur économique suit un processus aléatoire, alors, les effets d'un choc temporaire (e.g. une forte inflation subitement provoquée par une dévaluation) ne vont pas se dissiper après plusieurs périodes de temps, mais plutôt seront permanents.

La plupart des séries temporelles économiques sont non stationnaires en ce sens que la moyenne et la variance dépendent du temps, elles évoluent à partir de n'importe quelle valeur de départ aussitôt que le temps passe. Si cette évolution est prédominante dans une direction (à la hausse ou à la baisse), on dira que les séries marquent une tendance. Aussi ces séries sont-elles fréquemment détrendées avant toute analyse. Pour ce faire, deux procédures de détrendisation, consistant à enlever l'influence du temps, sont utilisées:

- estimation des régresseurs sur le temps:

$X_t = f(t) + u_t$ e.g. $X_t = \alpha + \beta t + u_t$
modèle appelé "trend-stationary process (TSP)"

La série détrendée est $\hat{u}_t$ telle que: $\sum \hat{u}_t = 0 \quad et \quad \sum t\hat{u}_t = 0$

- différenciation successive (procédure appropriée) pour l'élimination du trend.

$\Delta X_t = X_t - X_{t-1} = \beta + u_t - u_{t-1}$ qu'il il faut à nouveau différencier pour éliminer β:

$\Delta^2 X_t = \Delta^2 u_t = u_t - 2u_{t-1} + u_{t-2}$ qui est la série détrendée

Dans le cas où X_t serait généré par le modèle: $X_t - X_{t-1} = \beta + u_t$ modèle appelé "difference-stationary process (DSP)"

où u_t est une série stationnaire, alors, la différence première de X_t est stationnaire de moyenne β. Ce modèle est appelé "***random-walk model***":

Au début des années 80, de nombreuses études ont montré que beaucoup de séries temporelles économiques sont des processus aléatoires ou tout au moins ont des éléments constitutifs qui suivent de tels processus. Ces études utilisent les tests de racine unitaire introduits par David DICKEY et Wayne FULLER (1981).

iii). Processus MA

Soit $\{u_t\}$ un processus purement aléatoire tel que : $E(u_t) = 0 \; et \; V(u_t) = \sigma_u^2$

le processus MA $\{X_t\}$ est défini par:

$X_t = u_t + \beta_1 u_{t-1} + \cdots + \beta_m u_{t-m} = (1 + \beta_1 D + \cdots + \beta_m D^m) u_t$

processus MA d'ordre m, noté MA(m), et où D est l'opérateur de décalage

Si en plus de ses propriétés d'être centré et de variance constante, le processus $\{u_t\}$satisfait la propriété d'absence de corrélation: $(E(u_t, u_\tau) = 0 \; si \; t \neq \tau)$, alors il sera dit un bruit blanc.

iv). Processus AR

Le processus $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_r X_{t-r} + u_t$

est appelé processus autorégressif d'ordre r et noté AR (r)

En utilisant l'opérateur de décalage D, on écrira:

$$X_t = (\alpha_1 D + \alpha_2 D^2 + \cdots + \alpha_r D^r) X_t + u_t \Leftrightarrow (1 - \alpha_1 D - \alpha_2 D^2 - \cdots - \alpha_r D^r) X_t = u_t$$

ou encore:

$$X_t = \frac{1}{1 - \alpha_1 D - \alpha_2 D^2 - \cdots - \alpha_r D^r} u_t = \frac{1}{(1 - \pi_1 D)(1 - \pi_2 D) \cdots (1 - \pi_r D)} u_t$$

où π_1 , π_2 , $\cdots$, π_r avec $|\pi_i| < 1 \quad \forall\, i$ sont les racines de l'équation:

$$Y^r - \alpha_1 Y^{r-1} - \alpha_2 Y^{r-2} - \quad \cdots \quad - \alpha_r = 0$$

v). Processus ARMA

Un modèle qui combine des processus AR et MA est dit ARMA(p, q), défini par:

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + u_t + \beta_1 u_{t-1} + \cdots + \beta_q u_{t-q}$$

vi). Processus ARIMA et SARIMA

Dans la pratique, la plupart des séries temporelles sont non stationnaires. Une procédure souvent utilisée pour rendre stationnaire une série non stationnaire est la différenciation successive, à l'aide de l'opérateur: $\Delta = 1 - D$ tel que:

$$\Delta X_t = (1 - D) X_t = X_t - X_{t-1}$$
$$\Delta^2 X_t = (1 - D)^2 X_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$$
$$\vdots$$

ainsi de suite

En supposant que $\Delta^d X_t$ soit une série stationnaire qui peut donc être représentée par un modèle ARMA(p,q), ce modèle sera dit modèle intégré parce que le modèle stationnaire ARMA qui correspond aux données différenciées doit être "intégré" (sommé) pour obtenir un modèle pour les données non stationnaires.

Donc, si $\Delta^d X_t$ est un ARMA (p, q) alors X_t est un ARIMA (p, d, q)

Les modèles SARIMA permettent d'intégrer un ordre de différenciation lié à une saisonnalité:

$(1-D^S)X_t = X_t - X_{t-s}$ avec s = 4 pour une série trimestrielle

Il faut faire remarquer que:

- même s'il n'y a pas besoin d'une composante MA dans la modélisation ARIMA de X_t, la procédure de différenciation de X_t produira un MA par effet Slutsky
- ARMA (1, 0) = AR(1)
- ARMA (0, 1) = MA(1)
- les modèles AR, MA et ARMA ne sont représentatifs que de séries (ou chroniques):
  - ✓ stationnaires en tendance
  - ✓ corrigées des variations saisonnières.

3.4.2. La méthode Box-Jenkins de prévision

Cette méthode est une étude systématique des séries chronologiques à partir de leurs caractéristiques. Mais puisque la plupart de ces séries contiennent un trend et sont donc non stationnaires, il faut d'abord éliminer le trend par les différences premières successives:

$\Delta X_t = X_t - X_{t-1} = (1-D)X_t$ où D est un opérateur de décalage:

$DX_t = X_{t-1}$ et plus généralement $D^j X_t = X_{t-j}$

Soit: $\Delta^j X_t = (1-D)^j X_t$

Box–Jenkins détermine, en 3 étapes, de tous les ARIMA celui qui représente le mieux le phénomène étudié:

- l'identification
- l'estimation
- le contrôle diagnostique.

Au préalable, il faut éliminer la saisonnalité, ensuite analyser la série désaisonnalisée, toujours à l'aide d'un corrélogramme. Si le corrélogramme

fait apparaître un trend, il faut l'éliminer en transformant la série en différences premières.

L'identification est la procédure employée pour obtenir une idée approximative de la structure du modèle i.e. le degré (p,q,d). Le modèle Box-Jenkins est une représentation ARMA de $\Delta^d X_t$, soit:

(1) $\alpha_p D(\Delta^d X_t) = \beta_q D(u_t)$ où $\alpha_p = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}$ $\beta_p = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix}$

Pour $d = 1$, on a le modèle:

(2) $X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \cdots - \alpha_p X_{t-p} = u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots - \beta_q u_{t-q}$

L'estimation des paramètres autorégressifs (le vecteur α_p) peut se faire facilement par la méthode des moindres carrés, mais les paramètres de moyenne mobile MA(β_q) posent quelques difficultés. La méthode d'estimation Box-Jenkins consiste en une procédure itérative de type balayage.

Dans la pratique, on retient: $p = q = 2$ (puisqu'au delà de deux paramètres, la procédure ne serait pas très efficace), ce qui conduit au modèle:

(3) $X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} = u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2}$

$\Leftrightarrow$ $X_t - \alpha_1 DX_t - \alpha_2 D^2 X_t = (1 - \alpha_1 D - \alpha_2 D^2) X_t = u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2}$

Il en résulte:

$$X_t = \frac{1}{1 - \alpha_1 D - \alpha_2 D^2}(u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2})$$

$\Leftrightarrow$ $X_t = v_t - \beta_1 v_{t-1} - \beta_2 v_{t-2}$ où $v_t = \frac{1}{1 - \alpha_1 D - \alpha_2 D^2} u_t$

Ce modèle est équivalent à:

(4) $$v_t = X_t + \beta_1 v_{t-1} + \beta_2 v_{t-2}$$

Soient les étapes suivantes d'estimation des paramètres:

i). On choisit une valeur plausible $(\hat{\beta}_1, \hat{\beta}_2)$ e.g. (1, 1)

ii). On pose: $\hat{v}_0 = 0 \; et \; \hat{v}_1 = 0$

iii). On génère une succession de valeurs estimées de v_t ($\hat{v}_t$) partant du modèle (4) :

$$\hat{v}_2 = X_2$$

iv). A partir de la série ($\hat{v}_t$) ainsi obtenue, on estime les paramètres $\alpha_1 \; et \; \alpha_2$ partant de:

$$v_t = \frac{1}{1 - \alpha_1 D - \alpha_2 D^2} u_t \qquad \Rightarrow \qquad u_t = (1 - \alpha_1 D - \alpha_2 D^2) v_t$$

Soit le modèle:

(5) $$v_t - \alpha_1 v_{t-1} - \alpha_2 v_{t-2} = u_t$$

$$\min RSS = \min \sum e_t^2 = \min \sum \left(\hat{v}_t - \hat{\alpha}_1 \hat{v}_{t-1} - \hat{\alpha}_2 \hat{v}_{t-2} \right)^2$$

v). Après avoir estimé le modèle (5), $\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix}$ on détermine les résidus $\hat{u}_t$ et on vérifie s'ils ne forment pas une série systématique. Si oui, on change la spécification du modèle (en éliminant l'ordre du modèle AR ou MA non valide) et on reprend l'estimation. Si non et donc les résidus sont aléatoires, on peut alors utiliser le modèle à des fins de prévision.

Après avoir estimé les 3 modèles de prévision, on cherchera à indiquer laquelle de ces 3 méthodes donne les prévisions les plus exactes (précises) pour la période T+1 à T+n

3.5. Mesure de la précision d'une prévision

Soient :

A_t – la valeur effective ou réelle d'une variable à la date t

F_t – la valeur prédite à la même date t.

On définit les changements relatifs (ou variations relatives)

- réels : $a_t = \frac{A_t - A_{t-1}}{A_{t-1}}$
- prédits : $f_t = \frac{F_t - A_{t-1}}{A_{t-1}}$

On applique, soit aux valeurs A_t et F_t ou à leurs changements relatifs a_t et f_t (les plus couramment utilisés) des méthodes pour mesurer la précision (accuracy) de la prévision. On en retiendra:

- l'erreur quadratique moyenne (MSE)
- la statistique U de Theil (U)
- l'erreur absolue moyenne (MAE) et l'erreur relative moyenne (MPE).

3.5.1. Erreur quadratique moyenne

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(f_t - a_t)^2 = \frac{1}{T}\sum_{t=1}^{T}\left((f_t - a_t) - (\bar{f} - \bar{a}) + (\bar{f} - \bar{a})\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left((f_t - a_t) - (\bar{f} - \bar{a})\right)^2 + (\bar{f} - \bar{a})^2$$

$$MSE = S^2_{f-a} + (\bar{f} - \bar{a})^2 \text{ où}$$

S^2_{f-a} est la variance des erreurs de prévision

$(\bar{f} - \bar{a})^2$ la composante BIAIS qui indique jusqu'à quelle hauteur (importance) l'amplitude du MSE est conséquence de cette tendance à estimer trop haut ou trop bas le niveau de la variable de prévision. Le biais de prévision ou l'erreur moyenne de prévision est donné par: $ME = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t$ où ε_t est l'erreur de prévision qui doit être nulle si la méthode est adaptée.

Si r est le coefficient de corrélation entre a et f, on montre que :

(1) $$S_{f-a}^2 = S_f^2 + S_a^2 - 2rS_fS_a = (S_f - S_a)^2 + 2(1-r)S_fS_a$$

(2) $$S_{f-a}^2 = (S_f - rS_a)^2 + (1-r^2)S_a^2$$

où

$$\begin{cases} (S_f - S_a)^2 - \text{var}\,iance \\ 2(1-r)S_fS_a - \text{cov}\,ariance \end{cases}$$

$$\begin{cases} (S_f - rS_a)^2 - régression \\ (1-r^2)S_a^2 - résidu \end{cases}$$

On décompose donc le MSE (Theil 1966) dans l'un de ces 2 ensembles de composantes, chaque composante étant une source d'erreur de prévision :
(1) biais, variance et covariance
(2) biais, régression et résidu.

A partir de cette décomposition, Theil a défini les statistiques

$U^m = \dfrac{(\bar{f} - \bar{a})^2}{MSE}$ proportion biais

$U^E = \dfrac{(S_f - rS_a)^2}{MSE}$ proportion régressive

$U^R = \dfrac{(1-r^2)S_a^2}{MSE}$ proportion résiduelle

Evidemment $U^m + U^E + U^R = 1$

La décomposition en U^m, U^R, U^S, par comparaison des valeurs prédites et celles réelles, renseigne sur les possibles inadéquations des prévisions i.e. la dispersion des valeurs autour de la ligne de prévision parfaite $A_t = F_t$

Si, U^m est grand, alors le changement prédit moyen s'écarte significativement du changement réel moyen. L'on s'attendrait à ce que les prédicteurs réduisent de tels écarts au fil du temps. En effet, si l'on considère la régression :

$A_t = \alpha + \beta F_t$

Alors pour une prévision optimale : $U^m = 0$ si $\hat{\alpha} = 0$ et $U^E = 0$ si $\hat{\beta} = 1$

En d'autres termes, le prédicteur F_t est efficace si $\alpha = 0$ et $\beta = 1$.

3.5.2. Statistique U de Theil

$$U = \sqrt{\frac{MSE}{\sum A_t^2 / n}} \qquad 0 \le U < \infty$$

$U = 0$ pour une prévision parfaite.

Si U > 1, il est préférable d'accepter l'extrapolation qui consiste à supposer toute absence de changement et donc à poser $X_t = X_{t-1}$

3.5.3. Erreur absolue moyenne et erreur relative moyenne

A la différence des modèles économétriques classiques, le TSM (Time series models) ne relient pas une variable à un ensemble d'autres variables dans une relation de causalité mais plutôt fondent la prédiction sur le seul comportement passé de la variable considérée. En effet, il peut être difficile voire impossible d'expliquer l'évolution de X(t) à travers un modèle de type structurel. C'est le cas par exemple si les données ne sont disponibles pour les variables explicatives. Même quand l'équation de régression de X(t) est statistiquement significative, le résultat peut ne pas être utile pour la prévision. Pour une prévision de X(t) à l'aide de l'équation de régression, les variables explicatives qui ne sont pas décalées doivent être à leur tour prédites, et ceci peut être plus difficile que la prévision même de X(t).

Les TSM sont une méthode d'extrapolation des modèles dans lesquels l'on n'a pas de connaissances structurelles sur la relation causale qui affecte la variable pour laquelle l'on fait de la prévision. Aussi choisira-t-on un TSM à chaque fois qu'on sait peu à priori sur les déterminants de la variable et dès lors qu'on dispose de suffisamment d'observations pour construire une série chronologique avec une longueur raisonnable.

$MAE = \frac{1}{T}\sum |A_t - F_t|$ (Mean absolute error) qui doit être minime

$MRE = \frac{1}{T}\sum |a_t - f_t|$ (Mean relative error) qui doit être minime**.**

4. Racines unitaires et cointégration

A la différence des modèles économétriques classiques, les TMS (Time series models) ne relient pas une variable à un ensemble d'autres variables dans une relation de causalité mais plutôt fondent la prédiction sur le seul comportement passé de la variable considérée. En effet, il peut être difficile voire impossible d'expliquer l'évolution de X(t) à travers un modèle de type structurel. C'est le cas par exemple si les données ne sont pas disponibles pour les variables explicatives. Même quand l'équation de régression de X(t)est statistiquement significative, le résultat peut ne pas être utile pour la prévision. Pour une prévision de X(t) à l'aide de l'équation de régression, les variables explicatives qui ne sont pas décalées doivent être à leur tour prédites, et ceci peut être plus difficile que la prévision même de X(t).

Les TMS sont une méthode d'extrapolation des modèles dans lesquels l'on n'a pas de connaissances structurelles sur la relation causale qui affecte la variable pour laquelle l'on fait de la prévision. Ainsi choisira-t-on un TMS à chaque fois qu'on sait peu a priori sur les déterminants de la variable et dès lors qu'on dispose de suffisamment d'observations pour construire une série chronologique avec une longueur raisonnable.

Soit le modèle:

(m) $$X_t = \alpha + \beta t + \rho X_{t-1} + u_t$$

Le processus X_t peut croître sous 2 hypothèses:

H_1: Un trend positif ($\beta > 0$) avec un processus stationnaire après détrendisation ($\rho < 1$).

Sous cette hypothèse, X_t peut être utilisé dans une régression avec utilisation des tests usuels

H_2 : X_t suit un processus aléatoire avec une pente positive

($\alpha > 0$, $\beta = 0$, $\rho = 1$).

Dans ce cas, il vaut mieux travailler avec ΔX_t, d'autant plus que la détrendisation ne rendrait pas la série stationnaire.

Le test de l'hypothèse $\rho = 1$ (avec $\beta = 0$) dans l'équation autorégressive de premier ordre de la forme: $X_t = \alpha + \rho X_{t-1} + u_t$ est appelé "***test de racines unitaires***".

Pour ce faire ($\rho = 1,\ \beta = 0$), Dickey et Fuller suggèrent un test de ratio de vraisemblance (LR test – Likelihood ratio) pour lequel ils présentent des tables de test, F-valeurs tabulées beaucoup plus élevées que celles contenues dans les tables habituelles de F. Ce test est également utilisé pour savoir si les données sont DSP ou TSP, sinon il est préférable de faire des régressions sur les différences premières. La seule difficulté avec la procédure de différenciation est qu'il en résulte une perte de données. Le concept de séries co-intégrées est suggéré comme solution à ce problème.

4.1. Racines unitaires

Soit le modèle général:

(G) $$X_t = \alpha + \beta t + \rho X_{t-1} + \sum_{i=1}^{\theta} \lambda_i \Delta X_{t-i} + u_t$$

où il faut choisir θ e.g. $\theta = 1$, ainsi et avec $\lambda_1 = \lambda$:

(I) $$X_t = \alpha + \beta t + \rho X_{t-1} + \lambda \Delta X_{t-1} + u_t \quad \Leftrightarrow \quad \Delta X_t = \alpha + \beta t + (\rho - 1) X_{t-1} + \lambda \Delta X_{t-1} + u_t$$

A titre d'exemple, si on considère le prix du riz corrigé des variations saisonnières (statistiques SIM), avec deux hypothèse simples:

- le prix suit un processus aléatoire de sorte que personne ne peut prédire son mouvement
- à long terme (comme le stipule la théorie microéconomique), le prix d'un bien est lié à son coût marginal de production.

Donc, le modèle $p_t = \alpha + \beta t + \rho p_{t-1} + \lambda \Delta p_{t-1} + u_t$ donnera:

$$\begin{cases} p_t - p_{t-1} = \alpha + \beta t + (\rho - 1) p_{t-1} + \lambda \Delta p_{t-1} + u_t \\ p_t - p_{t-1} = \alpha + \lambda \Delta p_{t-1} + u_t \quad (\textit{compte tenu des restrictions } \beta = 0 \textit{ et } \rho = 1) \end{cases}$$

Le test proprement dit $\begin{cases} \beta = 0 \quad et \\ \rho = 1 \end{cases}$ à partir du modèle général (I) est séquentiel:

i). Par les OLS, on fait la régression sans restriction (URSS):

$$X_t - X_{t-1} = \alpha + \beta t + (\rho - 1) X_{t-1} + \lambda \Delta X_{t-1} + u_t = \Delta X_t$$

ii). On tient compte des restrictions: $\beta = 0$ *et* $\rho = 1$, alors le modèle devient:

$$X_t - X_{t-1} = \alpha + \lambda \Delta X_{t-1} + u_t = \Delta X_t$$

iii). On calcule le ratio F pour tester si les restrictions sont respectées:

$$F_{DF} = \frac{(RRSS - URSS)/c}{URSS/(T-k)} \qquad \text{où} \qquad \begin{cases} c = 2 \begin{cases} \beta = 0 \\ \rho = 1 \end{cases} \\ k = 4 (\alpha, \beta, \rho, \lambda) \\ T = nombre\, d'observations \end{cases}$$

F_{DF} n'est pas distribué comme une distribution Fisher standard sous l'hypothèse nulle:

☞ Si on rejette l'hypothèse de racine unitaire $(\rho = 1)$, on peut alors tester les coefficients de la tendance avec des Student standard

✓ Si $\beta \neq 0$ *et* $\alpha \neq 0$, le processus est stationnaire autour d'une tendance déterministe I(0) + T + C

✓ Si $\beta \neq 0$ *et* $\alpha = 0$, le processus est stationnaire autour d'une tendance déterministe sans constante I(0) + T

✓ Si $\alpha \neq 0$, le processus est stationnaire autour d'une constante I(0) + C

✓ Si $\beta = 0$, le processus n'admet pas de tendance déterministe, auquel cas, il faut retester la racine unitaire dans un modèle sans tendance

✓ Si $\alpha = 0$, il faut retester l'hypothèse d'une racine unitaire sur un processus sans constante

☞ Si on accepte l'hypothèse de racine unitaire, alors:

$$\Delta X_t = \alpha + \beta t + \lambda \Delta X_{t-1} + u_t$$

on testera α *et* β sachant que $\rho = 1$:

✓ Si $\beta \neq 0$, le processus est intégré d'ordre 1 autour d'une tendance quadratique, ce qui est un signal pouvant par exemple amener à tester un modèle avec changement de pente

✓ Si $\beta = 0$, le modèle devient: $\Delta X_t = \alpha + \lambda \Delta X_{t-1} + u_t$
Alors, les statistiques suivent des lois standard:

- Si $\alpha \neq 0$, le processus est intégré d'ordre 1 autour d'une tendance déterministe I(1) + T
- Si $\alpha = 0$, le processus est intégré d'ordre 1 I(1).

Note: Les statistiques, liées aux tests d'hypothèses jointes, ϕ_1 , ϕ_2 , ϕ_3 , de DF ne suivent pas asymptotiquement une loi de Fisher-Snédecor mais des distributions empiriques tabulées par Dicker-Fuller:

$$\phi_1 \begin{cases} H_0 : (\rho - 1 = \alpha = 0) \quad I(1) \\ H_1 : I(0) + C \end{cases} \qquad \phi_1 = \frac{(R_0 - R_2)/2}{R_2/(T-k-2)}$$

$$\phi_2 \begin{cases} H_0 : (\rho - 1 = \alpha = \beta = 0) \quad I(1) \\ H_1 : I(0) + T + C \end{cases} \qquad \phi_2 = \frac{(R_0 - R_1)/3}{R_1/(T-k-3)}$$

$$\phi_3 \begin{cases} H_0 : (\rho - 1 = \beta = 0) \quad I(1) + T \\ H_1 : I(0) + T \end{cases} \qquad \phi_3 = \frac{(R_5 - R_1)/2}{R_1/(T-k-2)}$$

où R_j est le RSS de l'équation j

j = (0) $X_t = \rho X_{t-1} + \lambda \Delta X_{t-1} + u_t$

(1) $X_t = \alpha + \beta t + \lambda \Delta X_{t-1} + u_t$

(2) $X_t = \alpha + \lambda \Delta X_{t-1} + u_t$

(3) $X_t = \lambda \Delta X_{t-1} + u_t$

(4) $X_t = \alpha + \beta t + \rho X_{t-1} + \lambda \Delta X_{t-1} + u_t$

(5) $X_t = \alpha + \rho X_{t-1} + \lambda \Delta X_{t-1} + u_t$

qu'on ramène aux différences premières avant régression.

4.2. Cointégration

La procédure qui consiste à éliminer d'abord les trends dans les variables par différenciation rejette des informations pertinentes sur des relations de long terme qui intéresse la théorie économique. La théorie de la cointégration, développée par C.W.J. Granger (1981) et élaborée dans R.F. Engle, C.W.J. Granger (1987) traite de l'intégration des dynamiques de court terme en équilibre de long terme. Demander si X_t et Y_t sont cointégrées, c'est demander s'il existe une relation de long terme entre leurs trends respectifs. En effet, une combinaison linéaire de deux variables (X_t *et* Y_t) suivant un processus aléatoire serait stationnaire (e.g. $Z_t = Y_t - \lambda X_t$). Dans ce cas, on dira que les variables X_t *et* Y_t sont co-intégrées et λ est appelé paramètre de co-intégration. A titre d'illustration, on peut considérer l'échange entre deux pays

UEMOA, soit X la part du pays de comparaison dans les exportations du pays considéré et Y la part des importations du pays considéré en provenance du pays de comparaison.

Soient X_t ~ I(1) et Y_t ~ I(1), elles seront dites cointégrées s'il existe un β tel que:

$$Y_t - \beta X_t \quad \sim \quad I(0) \quad \text{et on notera: } X_t \text{ et } Y_t \sim \text{CI}(1,1)$$

Il faut noter que si X_t et Y_t ne sont pas cointégrées, alors

$$Y_t - \beta X_t \quad = u_t \text{ sera aussi I(1)}$$

Le test de cointégration se fait en 3 étapes:

(i). appliquer les tests de racine unitaire pour savoir si X_t et Y_t sont tous deux I(1)

(ii). si OUI, régresser, par les OLS, Y sur X (ou X sur Y) et considérer $\hat{u} = Y - \hat{\beta}\, X$ pour ensuite appliquer les tests de racine unitaire sur $\hat{u}$

(iii). si $\hat{u}$ a une racine unitaire i.e. $\hat{u} = Y - \hat{\beta}\, X$ est I(0), alors X et Y sont cointégrées sinon u sera I(1)

Soit:

H_0 : u a une racine unitaire ou X et Y ne sont pas cointégrées

H_1 : X et Y sont cointégrées

Plus généralement, si X_t et Y_t sont intégrés d'ordre d i.e. d-homogeneous nonstationary et si $Z_t = Y_t - \lambda X_t$ est intégré d'ordre b, avec $b < d$, on dit que X_t et Y_t sont cointégrées d'ordre d et b. Du coup, le test de ci-dessus s'opère dans les conditions particulières de: $\begin{cases} d = 1 \\ b = 0 \end{cases}$

4.3. ECM

Un modèle à correction d'erreur (ECM) est de la forme:

$$\Delta Y_t = \gamma\, \Delta X_t + \alpha (Y - \beta X)_{t-1} + u_t$$

5. Exemples

5.1. Exemple 1

L'évolution mensuelle des ventes d'un produit est consignée dans le tableau ci-dessous:

	jan	fev	mar	avr	mai	jun	jul	aut	sep	oct	nov	dec
2011	3	3	3	3	3	3	2	2	3	4	6	7
2012	3	3	3	4	3	4	4	2	5	6	10	13
2013	4	3	4	5	5	5	5	2	6	7	10	13

1. Calculer pour chaque année, la vente mensuelle moyenne $\left(\overline{X}\right)$ et l'écart type $\left(\sigma_X\right)$

2. Estimer la pente de la droite de régression $D\sigma_X\left(\overline{X}\right)$, soit $\hat{\beta}$

3. Déterminer le trend par les moyennes mobiles

4. Désaisonnaliser la série en adoptant le modèle approprié selon la valeur de la pente, étant donné que:

   - ✓ Si $\hat{\beta} < 0.05$ on choisit le modèle additif
   - ✓ Si $\hat{\beta} > 0.10$ on choisit le modèle multiplicatif
   - ✓ Si $0.05 < \hat{\beta} < 0.10$ on choisit le modèle qui donne les meilleurs résultats

5. Prévoir, par la méthode naïve, les ventes mensuelles de 2014

6. Prévoir, par lissage, les ventes mensuelles de 2014

7. Estimer les paramètres du modèle ARIMA: $Y_t - aY_{t-1} = u_t - bu_{t-1}$ où Y représente la série désaisonnalisée et u une $v.a. \rightarrow N\left(0,\ \sigma_u^2\right)$

5.2. Exemple 2

L'évolution trimestrielle du chômage (X_t), au cours des 4 dernières années est donnée par le tableau suivant:

Année \ Trimestre	I	II	III	IV
2011	–	2	1	3
2012	4	2	0	4
2013	6	3	2	5
2014	7	–	–	–

1. Déterminer le trend par:
   i). la régression linéaire $\left(X_t = a + b\,t^2 + u_t\right)$
   ii). l'autorégression $\left(X_t = \alpha + \beta\, X_{t-2} + u_t\right)$

2. Pour un trend au choix, désaisonnaliser la série en adoptant un modèle additif puis multiplicatif

3. Prévoir le chômage des trois derniers trimestres de 2014, par:
   i). la méthode naïve
   ii). la régression linéaire
   iii). l'autorégression

4. En 2014, le chômage observé au cours des trois derniers trimestres aura été 2 – 3 – 4, calculer, pour chaque méthode de prévision:
   i). l'erreur absolue moyenne:
   $$\left(AAE = \frac{1}{n}\sum |A_t - F_t| \;\; où \;\; A\ est\ la\ valeuer\ observée\ et\ F\ \ la\ valeur\ \ prédite\right)$$
   ii). L'erreur quadratique moyenne
   $$\left(MSE = \frac{1}{n}\sum (f_t - a_t)^2 \;\; où \;\; a_t = \frac{A_t}{A_{t-1}} - 1 \;\; et \;\; f_t = \frac{F_t}{X_{t-1}} - 1\right)$$

5. Indiquer la meilleure méthode de prévision

Références bibliographiques

R.F. Engle and C.W.J. Granger (1987): Cointegration and error correction – Representation, estimation and testing, Econometrica, vol 55, nr 2, pp. 251-276

R. Giraud (1993): L'économétrie, PUF, "Que sais-je?"

C.W.J. Granger (1981): Some properties of time series data and their use in econometric model specification, Journal of econometrics, vol 16 nr 1, pp. 121-130

Julien Jacques (2007): Introduction aux séries temporelles, Polytech'Lille, Département GIS, 5ème année

Th. Jobert (1992): Analyse comparative et quantitative des facteurs de persistance du chômage - Annexes méthodologiques, 1. Tests de racine unitaire, 2. Cointégration et causalité

G.S. Maddala (1992): Introduction to econometrics, Prentice Hall

Catherine Pardoux, Bernard Goldfarb (2012): Prévision à court terme – Méthodes de lissage exponentiel, Université Paris Dauphine, novembre

R.S. Pindyck, D.L. Rubinfeld (1991): Econometric models and economic forecasts, Macgraw-Hill ed.

Université de Nice, Département de mathématiques (2005): Statistiques: Notes de cours 6 – Lissage de séries temporelle

Annexe. Table (Dick et Fuller)

$$(\alpha, \beta, \rho) = (\alpha, 0, 1) \; IN \; Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t$$

Sample size	Probability of a smaller value							
T	.01	.025	.05	.10	.90	.95	.975	.99
25	.74	.90	1.08	1.33	5.91	7.24	8.65	10.61
50	.76	.93	1.11	1.37	5.61	6.73	7.81	9.31
100	.76	.94	1.12	1.38	5.47	6.49	7.44	8.73
250	.76	.94	1.13	1.39	5.39	6.34	7.25	8.43
500	.76	.94	1.13	1.39	5.36	6.30	7.20	8.34
∞	.77	.94	1.13	1.39	5.34	6.25	7.16	8.27
SE	.004	.004	.003	.004	.015	.020	.032	.058