

Séries temporelles: II. Désaisonnalisation par la méthode de régression

EISTI

19 janvier 2020

- 1 Introduction
- 2 Ajustement linéaire de la tendance désaisonnalisée : Méthode de régression sur le temps-MRT-
- 3 Détermination des coefficients saisonniers
- 4 Détermination de la série corrigée
- 5 Désaisonnalisation par régression linéaire sur variables indicatrices (RVI)
- 6 Modèle trimestriel de Buys-Ballon (1847)
- 7 Estimateur des moindres carrés ordinaires (m.c.o.) : Cas général

▷ Pourquoi désaisonnaliser une série ?

- De nombreuses séries économiques présentent des comportements périodiques, rendant difficile la comparaison de deux instants successifs.
- La désaisonnalisation permet d'obtenir des séries dites corrigées des variations saisonnières (CVS).
- De telles séries, dites **désaisonnalisées, sont obtenues en éliminant la composante saisonnière de la série initiale.**
- Ces techniques sont notamment utilisées par les instituts statistiques qui publient des séries temporelles CVS afin d'en faciliter les comparaisons.

▷ Comment ?

- Calculer les coefficients saisonniers attachés à une période, à un mois, à un trimestre particulier.
- ▷ Les méthodes couramment employées pour désaisonnaliser une série temporelle sont
- **la régression linéaire et les filtres moyennes mobiles**

□ Le modèle additif le plus simple est de la forme

$$X_t = \underbrace{m_t + s_t}_{\text{partie déterministe}} + \underbrace{\epsilon_t}_{\text{partie aléatoire}}$$

où m_t, s_t , et ϵ_t sont : Tendence, Saisonnalité et le bruit.

▷ **Détermination de la tendance sur les données brutes.**

1. On élimine, dans un premier temps, les variations saisonnières en laissant la série. On pose :

$$X_t = m_t + \epsilon_t = \beta_0 + \beta_1 t + \epsilon_t, \quad t = 1, \dots, n$$

où la pente et l'ordonnée à l'origine, β_1 et β_0 respectivement, sont des paramètres inconnus. Les ϵ_t sont des v.a. (gaussiennes) centrées i.i.d.

□ Méthode des moindres carrés \implies choisir comme estimations des β_1 et β_0 celles qui minimisent

$$Q(\beta_0, \beta_1) = \sum_{t=1}^n [X_t - (\beta_0 + \beta_1 t)]^2$$

□ En désignant les solutions par $\hat{\beta}_0$ et β_1 , on obtient :

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{t=1}^n (X_t - \bar{X})(t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2} = \frac{\text{Cov}(t, X)}{\text{Vart}(t)} \\ \hat{\beta}_0 &= \bar{X} - \hat{\beta}_1 \bar{t}\end{aligned}\tag{1}$$

où $\bar{t} = (n+1)/2$ est la moyenne de $1, 2, \dots, n$.

□ On a :

$$\text{Var}(t) = \frac{1}{n} \sum_{t=1}^n t^2 - (\bar{t})^2 = \frac{n^2 - 1}{12} \quad \text{et} \quad \text{Cov}(t, X) = \frac{1}{n} \sum_{t=1}^n t \times X_t - \bar{t} \times \bar{X}$$

On pose :

$$X_t = m_t + \epsilon_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_d t^d + \epsilon_t, \quad t = 1, \dots, n$$

□ Une analyse graphique suggère souvent une valeur de d , ce qui permet ensuite d'estimer $\beta = (\beta_0, \beta_1, \dots, \beta_d)'$ par moindres carrés :

$$\hat{\beta} = \operatorname{argmin}_{\beta} \sum_{t=1}^n (X_t - m_t)^2$$

Pour $d = 2$ par exemple, et des temps d'observations t_1, \dots, t_n , on pose

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_3 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & t_1 & t_2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{pmatrix}, \quad \text{on obtient :}$$

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_3)' = \operatorname{argmin}_{\beta} \|X - \beta A\|_2^2 = (A'A)^{-1} A'X,$$

avec pour résidus d'estimation :

$$e = x - A\hat{\beta} = x - A(A'A)^{-1} A'X.$$

□ Considérez le processus de marche aléatoire Marche aléatoire avec tendance temporelle linéaire :

$X_t = \beta_0 + \beta_1 t + X_{t-1} + \epsilon_t$, où (ϵ_t) est une iid centrée.

□ Supposons que nous considérons (à tort) cela comme une tendance temporelle linéaire, $X_t \simeq \beta_0 + \beta_1 t + \epsilon_t$ et estimons la pente et l'interception par régression des moindres carrés.

Least Squares Regression Estimates for Linear Time Trend

```
> data(rwalk)
```

```
> model1=lm(rwalk~time(rwalk))
```

```
> summary(model1)
```

Estimate	Std. Error	t value	Pr(> t)	
Intercept	-1.008	0.2972	-3.39	0.00126
Time	0.1341	0.00848	15.82	< 0.0001

□ La pente et l'ordonnée à l'origine estimées sont respectivement

$\beta_1 = 0.13341$ et $\beta_0 = -1.008$.

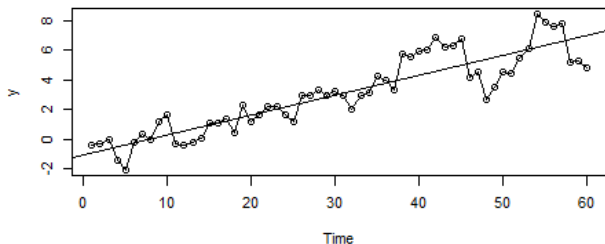


Figure: Marche aléatoire avec la ligne de tendance de régression des moindres carrés superposée.

2. Saisonnalité

□ La fonction $t \mapsto s_t$ est une fonction périodique, appelée composante saisonnière (journalière, hebdomadaire, mensuelle, annuelle, etc). Un exemple est donné par une fonction trigonométrique de la forme

$$s_t = a_0 + \sum_{j=1}^k (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

où les a_j et b_j sont inconnus et où les λ_i et λ_j sont des multiples entiers connus de $2\pi/d$. On devine d et les λ_i et λ_j puis on estime les a_i, b_j par moindres carrés. Par exemple, pour $k=1$ et $\lambda_1 = 2\pi/12$ on utilise

$$\begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{b} \end{pmatrix} = (A'A)^{-1} A'X, \quad \text{avec} \quad A = \begin{pmatrix} 1 & \cos(\lambda_1 t_1) & \sin(\lambda_1 t_1) \\ \vdots & \vdots & \vdots \\ 1 & \cos(\lambda_1 t_n) & \sin(\lambda_1 t_n) \end{pmatrix}$$

▷ **Méthode** : On estime le saisonnier S_t sur une série **détendancialisée** $d_t = X_t - \hat{m}_t$.

▷ Pour des **données trimestriels la période étant l'année**, nous avons :

- quatre saisons, donc 4 coefficients caractérisons la saison qu'on note :

$$s(j) := c_j, 1 \leq j \leq 4$$

▷ Le coefficient du mois de janvier sera la moyenne de tous les mois de janvier de la série $d_t = X_t - \hat{m}_t$.

On les obtiennent grâce la formule suivante :

$$c_j = \frac{1}{\text{Card}(T_j)} \sum_{t \in T_j} (X_t - \hat{m}_t) \quad 1 \leq j \leq 4 \quad \text{modèle additif}$$

- X_t les données brutes

- \hat{m}_t les données obtenus grâce au trend

- Les T_j sont l'ensemble du temps (ici ensemble des trimestres).

▷ **Exemple 1.** Si $T = 16$, alors

- La saison $j = 1$ correspond au trimestre 1,

- La saison $j = 2$ correspond au trimestre $5=1+4$ (on rajoute une année supplémentaire pour obtenir la même saison). Ainsi

$$T_1 = \{1, 5, 9, 13\}, \quad T_2 = \{2, 6, 10, 14\}, \quad T_3 = \{3, 7, 11, 15\}, \quad T_4 = \{4, 8, 12, 16\}$$

▷ **Définition.**

La composante saisonnière estimée \hat{S}_t est définie par

$$\hat{S}_t = \{c_j, j \in T_j\}, \quad 1 \leq j \leq p \quad \text{où } p \text{ désigne la période.}$$

$\hat{S}_t = c_j$ si la date t correspond au trimestre j .

▷ On obtient :

- pour $1 \leq t \leq 4$, les quatre coefficients saisonniers.
- En suite, pour $5 \leq t \leq 8$, les quatre coefficients saisonniers recopier
- et ainsi de suite.

▷ La composante saisonnière estimée \hat{S}_t sera répétitif (périodique). Il suffit donc de calculer seulement les quatre premiers coefficients saisonniers.

▷ Exemple.

$$\begin{aligned}\hat{S}_2 = c_2 &= \frac{1}{\text{Card}(T_2)} \sum_{t \in T_2} (X_t - \hat{m}_t) \\ &= \frac{1}{4} [(X_2 - \hat{m}_2) + (X_6 - \hat{m}_6) + (X_{10} - \hat{m}_{10}) + (X_{14} - \hat{m}_{14})]\end{aligned}$$

▷ Calcul de \hat{S}_{14} . En faisant la division euclidienne de 14 par 4 dont reste est 2, on obtient

$$\hat{S}_{14} = c_2$$

▷ **Le principe de la conversion des aires** : Les moyennes annuelles de la série brute et de la série CVS doivent être identiques.

▷ **Dans le cas d'un schéma additif** : la somme des coefficients doit être nulle.

Soit c_j le j -ème coefficient provisoire, on calcule la somme de ces coefficients :

$$c = \sum_{j=1}^p c_j \quad \text{avec } p \text{ la période de la saisonnalité}$$

- Si $c = 0$, les c_j sont les coefficients saisonniers définitifs
- Si $c \neq 0$, les coefficients saisonniers sont normés afin que leur somme soit nulle.
- Les coefficients définitifs sont donnés par :

$$c_j^* = c_j - \bar{c} \quad \text{où} \quad \bar{c} = \frac{1}{p} \sum_{j=1}^p c_j$$

▷ **Dans le cas d'un schéma multiplicatif la moyenne des coefficients doit être égale à 1.**

$$c = \frac{1}{p} \sum_{j=1}^p c_j \quad \text{avec} \quad \text{Les coefficients définitifs sont donnés par : } c_j^* = \frac{c_j}{c}.$$

- ▷ La composante saisonnière estimée \hat{S}_t est à remplacer par

$$S_t^* = \{c_j^*, j \in T_j\}, \quad 1 \leq j \leq p$$

- ▷ La série corrigée à l'instant t est

$$X_{cvs,t} = X_t - S_t^*, \quad 1 \leq t \leq T$$

- ▷ Si on retire les coefficients saisonniers de la série des observations, on reconstitue la série corrigée des variations saisonnières :

- ▷ $X_{cvs,t}$: représente l'évolution de la variable observée s'il n'y avait pas eu les mouvements saisonniers.

- ▷ La suppression des variations saisonnières s'appelle la désaisonnalisation. C'est une opération qui est très couramment pratiquée par les instituts de statistique et de recherche économique afin d'étudier les phénomènes entre les différentes dates sans les distorsions introduites par les effets de saison.

$$X_t \simeq m_t$$

→ MCO

↓

On estime m_t

$$\hat{m}_t = \hat{a} + \hat{b}t$$

Modél Additif

Modèle Multiplicatif

↓

↓

Méthode de la différence à la tendance

Méthode du rapport à la tendance

$$d_t = X_t - \hat{m}_t$$

$$d_t = \frac{X_t}{\hat{m}_t}$$

Coefficients saisonniers provisoires :

$$c_j = \frac{1}{\text{Card}(T_j)} \sum_{t \in T_j} d_t, \quad 1 \leq j \leq p$$

$$\sum_j c_j = 0$$

$$\sum_j c_j = p$$

Coefficients saisonniers définitifs : $c_j^* = c_j - \bar{c}$

$$c_j^* = \frac{c_j}{\bar{c}}$$

$$S_t^* = \{c_j^*, j \in T_j\}, \quad X_t^{CVS} = X_t - S_t^*$$

$$X_t^{CVS} = \frac{X_t}{S_t^*}$$

Dates: $T = t$	Série: X_t	$\hat{m}_t = \hat{a} + \hat{b}t$	$d_t = X_t - \hat{m}_t$	c_j^*	$X_{scv,t}$
⋮	⋮	⋮	⋮	⋮	

Remarque. Présentation graphique de la série + tendance + les données corrigés de la variation saisonnière.

- Les données corrigés permettent de supprimer la composante saisonnière des données brutes.
- Cela permet d'étudier l'évolution de la série indépendamment de l'évolution saisonnière.
- Quand la courbe des $X_{scv,t}$ est au dessous de la droite de la tendance cela signifie que la baisse est très sensible, le cas contraire signifie que les ventes varient indépendamment de la composante saisonnière.

▷ S'il existe une composante saisonnière et que des coefficients saisonniers c_1, c_2, \dots, c_j ; ont été calculés, ces coefficients représentent le profil saisonnier moyen de la série. En faisant l'hypothèse que la saisonnalité se reproduit à l'identique, ils peuvent être utilisés pour prévoir les valeurs de cette composante.

▷ La prévision de la chronique est obtenue par agrégation des différentes composantes.

▷ La prévision à l'instant $t+h$ ré-saisonnalisée est donnée par

$$X_{t+h}^P = \hat{m}_{t+h} + \hat{X}_{t+h}^{svs} = \hat{m}_{t+h} + c_j$$

où $c_j = s(j)$ est le coefficient saisonnier, avec j le nom du mois (ou de l'année) correspondant à la date $(t+h)$.

- ▷ Les données concernent le trafic voyageur de la SNCF en deuxième classe.
- Elles sont exprimées en millions de voyageurs kilomètres.
- Les observations mensuelles portent sur la période 1963-1980 (Gouriéroux et Monfort (1995)). .

```
sncf=read.table("http://freakonometrics.free.fr/sncf.csv",header=TRUE,
sep=";")
SNCF=ts(as.vector(t(as.matrix(sncf[,2:13]))),,start=c(1963, 1),
frequency = 12)
plot(SNCF,lwd=2,col="purple")
```

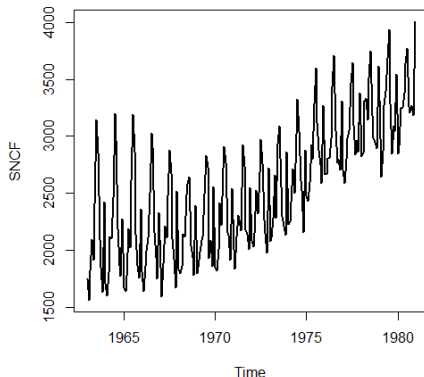


Figure: Évolution du trafic voyageur entre janvier 1963 et décembre 1966

- ▷ Le graphe fait apparaître une évolution à moyen terme approximativement constante et une périodicité de 12
- ▷ La série comportant une forte saisonnalité.

Désaisonnalisation par RVI : Présentation des données

```
> sncf=read.table("http://freakonometrics.free.fr/sncf.csv",header=TRUE,sep=
```

ANNEE	JANV	FEV	MARS	AVR	MAI	JUIN	JUIL	AOUT	SEPT	OCTOB	Nove	Dece
1963	1750	1560	1820	2090	1910	2410	3140	2850	2090	1850	1630	2420
1964	1710	1600	1800	2120	2100	2460	3200	2960	2190	1870	1770	2270
1965	1670	1640	1770	2190	2020	2610	3190	2860	2140	1870	1760	2360
1966	1810	1640	1860	1990	2110	2500	3030	2900	2160	1940	1750	2330
1967	1850	1590	1880	2210	2110	2480	2880	2670	2100	1920	1670	2520
1968	1834	1792	1860	2138	2115	2485	2581	2639	2038	1936	1784	2391
1969	1798	1850	1981	2085	2120	2491	2834	2725	1932	2085	1856	2553
1970	1854	1823	2005	2418	2219	2722	2912	2771	2153	2136	1910	2537
1971	2008	1835	2120	2304	2264	2175	2928	2738	2178	2137	2009	2546
1972	2084	2034	2152	2522	2318	2684	2971	2759	2267	2152	1978	2723
1973	2081	2112	2279	2661	2281	2929	3089	2803	2296	2210	2135	2862
1974	2223	2248	2421	2710	2505	3021	3327	3044	2607	2525	2160	2876
1975	2481	2428	2596	2923	2795	3287	3598	3118	2875	2754	2588	3266
1976	2667	2668	2804	2806	2976	3430	3705	3053	2764	2802	2707	3307
1977	2706	2586	2796	2978	3053	3463	3649	3095	2839	2966	2863	3375
1978	2820	2857	3306	3333	3141	3512	3744	3179	2984	2950	2896	3611
1979	3313	2644	2872	3267	3391	3682	3937	3284	2849	3085	3043	3541
1980	2848	2913	3248	3250	3375	3640	3771	3259	3206	3269	3181	4008

La série X_t est la somme de 2 composantes déterministes :

- une **tendance** m_t , d'une **saisonnalité** S_t
- et d'une composante **aléatoire** ϵ_t

$$X_t = m_t + S_t + \epsilon_t$$

▷ $\mathbb{E}(\epsilon_t) = 0$, $\text{Var}(\epsilon_t) = \sigma^2$, $\text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0 \quad \forall h > 0$.

▷ On suppose que m_t et S_t sont des combinaisons linéaires de fonctions connues dans le temps, m_t^i et S_t^j

$$\begin{cases} m_t = \beta_1 m_t^1 + \beta_2 m_t^2 + \dots + \beta_\ell m_t^\ell \\ S_t = c_1 S_t^1 + c_2 S_t^2 + \dots + c_n S_t^n \end{cases}$$

▷ **Objectif**. Estimer les coefficients $\beta_1, \dots, \beta_\ell$ et c_1, \dots, c_n à partir des observations.

▷ On peut écrire :

$$X_t = \sum_{i=1}^{\ell} \beta_i m_t^i + \sum_{j=1}^n c_j S_t^j + \epsilon_t, \quad t = 1, \dots, T.$$

▷ La forme de S_t dépend du type de données, et de la forme de la saisonnalité.

- Pour des données trimestrielles, elle pourra être choisie sous forme d'une fonction périodique de période 4.
- Elle prendra quatre valeurs (ou coefficients saisonniers) c_1, c_2, c_3, c_4 associées à chacun des trimestres :

$S_t = c_j$, si la date t correspond au trimestre j .

- On peut alors écrire

$$S_t = c_1 S_1^t + c_2 S_2^t + c_3 S_3^t + c_4 S_4^t$$

où S_t^j est une fonction indicatrice du trimestre j ,

- prend la valeur 1 si la date t correspond au trimestre j ,
- la valeur 0 sinon.

$$\left\{ \begin{array}{ll} S_t^j = 1 & \text{si } t = \text{trimestre } j \\ S_t^j = 0 & \text{si } t \neq \text{trimestre } j \end{array} \right. \quad \text{ou} \quad \left\{ \begin{array}{ll} S_t^j = 1 & \text{si } t = 0 \text{ [modulo } j] \\ S_t^j = 0 & \text{si } t \neq 0 \text{ [modulo } j] \end{array} \right.$$

▷ Dans le cas où la **tendance est linéaire**, et les **données sont trimestrielles**, le modèle peut s'écrire sous la forme :

$$\begin{aligned} X_t &= \underbrace{\beta_1 + \beta_2 t}_{m_t} + \underbrace{c_1 S_t^1 + c_2 S_t^2 + c_3 S_t^3 + c_4 S_t^4}_{S_t} + \epsilon_t \\ &= m_t + S_t \end{aligned}$$

▷ **Application aux données SNCF.**

▷ Supposons que les données commencent au **1er trimestre**.

Le nombre de voyageur

- au premier trimestre de 1963 est :
1750+1560+1820+2090=5130.
- au deuxième trimestre est de 6410, ...

Le modèle peut donc s'écrire de la forme suivante :

$$\begin{pmatrix} 5130 \\ 6410 \\ 8080 \\ 5900 \\ 5110 \\ 6680 \\ 8350 \\ 5910 \\ 5080 \\ \vdots \\ X_t \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ \vdots \\ t \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \\ S_t^1 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ S_t^2 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ S_t^3 \end{pmatrix} + c_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ S_t^4 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \vdots \\ \epsilon_t \end{pmatrix}$$

▷ Écriture Matricielle :

$$\begin{pmatrix} 5130 \\ 6410 \\ 8080 \\ 5900 \\ 5110 \\ 6680 \\ 8350 \\ 5910 \\ 5080 \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ 1 & 5 & 1 & 0 & 0 & 0 \\ 1 & 6 & 0 & 1 & 0 & 0 \\ 1 & 7 & 0 & 0 & 1 & 0 \\ 1 & 8 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t & S_t^1 & S_t^2 & S_t^3 & S_t^4 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \vdots \\ \epsilon_t \end{pmatrix}$$

$$\Rightarrow X = D\delta + \epsilon$$

▷ Avec

$$D = [m|S] \quad \text{et} \quad \delta' = [\beta_1, \beta_2 | c_1, \dots, c_4]$$

$$m = [m_t^i]_{i=1,2, \quad t=1, \dots, T} = \begin{bmatrix} \vdots & \vdots \\ m_t^1 & m_t^2 \\ \vdots & \vdots \end{bmatrix} \quad \text{ici } m_t^1 = 1, \quad m_t^2 = t$$

$$S = [S_t^j]_{j=1, \dots, 4, \quad t=1, \dots, T} = \begin{bmatrix} \vdots & \vdots \\ S^1 & S^4 \\ \vdots & \vdots \end{bmatrix}$$

▷ L'estimateur des moindres carrés du vecteur δ est :

$$\hat{\delta} = (D'D)^{-1}D'X$$

▷ La matrice $D'D$ n'est pas inversible car la première colonne est égale à la somme des 4 dernières (les composantes trimestrielles).

▷ Pour identifier le modèle, on rajoute la contrainte :

$$\left\{ \begin{array}{l} X_t = \beta_1 + \beta_2 t + c_1 S_t^1 + c_2 S_t^2 + c_3 S_t^3 + c_4 S_t^4 + \epsilon_t \\ \text{sous contrainte } c_1 + c_2 + c_3 + c_4 = 0 \end{array} \right.$$

▷ On considère le modèle général suivant

$$X_t = \sum_{i=1}^{\ell} \beta_i m_t^i + \sum_{j=1}^n c_j S_t^j + \epsilon_t, \quad t = 1, \dots, T.$$

▷ On cherche les paramètres β_i et c_j qui minimisent le carré des erreurs :

$$(\hat{\beta}, \hat{c}) = \operatorname{argmin} \left[\sum_{t=1}^T \epsilon_t^2 \right] = \operatorname{argmin} \left[\sum_{t=1}^T \left\{ X_t - \sum_{i=1}^{\ell} \beta_i m_t^i + \sum_{j=1}^n c_j S_t^j \right\}^2 \right]$$

▷ **Écriture matricielle** . Soit

$$\beta = (\beta_1, \dots, \beta_{\ell})', \quad c = (c_1, \dots, c_n)'$$

$$m = \left[m_t^i \right]_{1 \leq i \leq \ell, 1 \leq t \leq T} \quad \text{et} \quad S = \left[S_t^j \right]_{1 \leq j \leq n, 1 \leq t \leq T}$$

▷ Le modèle peut s'écrire :

$$X = m\beta + S\gamma + \varepsilon = [m|S] \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \varepsilon = D\delta + \varepsilon$$

avec

$\hat{\delta} = (\hat{\beta}, \hat{\gamma})'$ vérifiant l'équation :

$$D'D\hat{\delta} = D'X \Rightarrow [m \ S] \begin{bmatrix} m' \\ S' \end{bmatrix} \hat{\delta} = \begin{bmatrix} m'X \\ S'X \end{bmatrix}$$

▷ Soit donc

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} m'm & m'S \\ S'm & S'S \end{bmatrix}^{-1} \begin{bmatrix} m'X \\ S'X \end{bmatrix}$$

$$\begin{cases} \hat{\beta} = [m'm - m'S(S'S)^{-1}S'm]^{-1} [m'X - m'S(S'S)^{-1}S'X] \\ \hat{\gamma} = [S'S - S'm(m'm)^{-1}m'S]^{-1} [S'X - S'm(m'm)^{-1}m'X] \end{cases}$$

$$X_t = \beta_1 + \beta_2 t + c_1 S_t^1 + c_2 S_t^2 + c_3 S_t^3 + c_4 S_t^4,$$

▷ On veut résoudre le problème de minimisation suivante :

$$\left\{ \begin{array}{l} \min_{\beta, c} \sum_{t=1}^T [X_t - \beta_1 - \beta_2 t - \sum_{j=1}^4 c_j S_t^j]^2 \\ \text{sous la contrainte } c_1 + c_2 + c_3 + c_4 = 0 \end{array} \right.$$

⇔

$$\left\{ \min_{\beta, c} \sum_{t=1}^T [X_t - \beta_2 t - \sum_{j=1}^4 \tau_j S_t^j]^2 \right. \quad \text{où} \quad \left. \begin{array}{l} \beta_1 = [\tau_1 + \tau_2 + \tau_3 + \tau_4]/4 \\ c_j = \tau_j - \beta_1 \quad \forall j. \end{array} \right.$$

▷ Le système précédent montre que les estimateurs $\hat{\beta}_1, \hat{\beta}_2, \hat{c}_j$ ($j = 1, \dots, 4$) peuvent être calculés en utilisant une démarche en deux étapes.

▷ **Étape 1.** Chercher $\hat{\beta}_2, \hat{\tau}_j, j = 1, \dots, 4$ en minimisant :

$$\sum_{t=1}^T \left[X_t - \beta_2 - \sum_{j=1}^4 \tau_j S_j^t \right]^2$$

C'est-à-dire on régresse X_t sur les variables

$$m_t^2 = t, S_t^1, S_t^2, S_t^3, S_t^4.$$

(régression par les m.c.o. sans contrainte).

▷ **Étape 2.** On calcule ensuite $\hat{\beta}_1$ et \hat{c}_j en utilisant les égalités :

$$\begin{cases} \hat{\beta}_1 = \frac{1}{4} \sum_{j=1}^4 \hat{\tau}_j \\ \hat{c}_j = \hat{\tau}_j - \frac{1}{4} \sum_{j=1}^4 \hat{\tau}_j, \quad j = 1, \dots, 4. \end{cases}$$

▷ Calcul de $\hat{\beta}_2$ et $\hat{\tau}_j$

▷ Supposons que le nombre T d'observations correspond à un nombre entier N d'années : $T = 4N$

▷ Notons :

- n le numéro de l'année

- L'indice t de l'observation x_t du j^{e} trimestre de la n^{e} année n est $t = 4(n-1) + j$ où $j = 1, \dots, 4$.

- \tilde{x}_n : moyenne des X_t relatives à l'année n

- \bar{x}_j : moyenne des X_t relatives au trimestre j

- \bar{x} : moyenne de toutes les observations X_t

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N \tilde{x}_n = \frac{1}{4} \sum_{j=1}^4 \bar{x}_j$$

▷ On obtient

$$\hat{\beta}_2 = 3 \frac{\sum_{n=1}^N n \tilde{x}_n - \frac{N(N+1)}{2} \bar{x}}{N(N^2 - 1)}$$

$$\hat{\tau}_j = \bar{x}_j - [j + 2(N-1)] \hat{\beta}_2 \quad \text{pour } j = 1, 2, 3, 4$$

▷ Et

$$\begin{cases} \hat{\beta}_1 = \frac{1}{4} \sum_{j=1}^4 \hat{\tau}_j = \bar{x} - \left[\frac{5}{2} + 2(N-1) \right] \hat{\beta}_2 \\ \hat{c}_j = \hat{\tau}_j - \hat{\beta}_1 = \bar{x}_j - \bar{x} - \hat{\beta}_2 \left(j - \frac{5}{2} \right) \end{cases}$$

▷ Les relations précédentes peuvent être généralisées dans le cas d'une périodicité m . Le modèle s'écrit alors

$$X_t = \beta_1 + \beta_2 t + c_1 S_t^1 + c_2 S_t^2 + c_3 S_t^3 + c_4 S_t^4 + \dots + c_\ell S_t^\ell + \varepsilon_t,$$

▷ On veut résoudre le problème de minimisation suivante :

$$\left\{ \begin{array}{l} \min_{\beta, c} \sum_{t=1}^T [X_t - \beta_1 - \beta_2 t - \sum_{j=1}^{\ell} c_j S_t^j]^2 \\ \text{sous la contrainte } c_1 + c_2 + c_3 + c_4 + \dots + c_j = 0 \end{array} \right.$$

▷ On obtient :

$$\hat{\beta}_2 = \frac{12 \sum_{n=1}^N n \tilde{x}_n - \frac{N(N+1)}{2} \bar{x}}{N(N^2 - 1)}$$

$$\hat{\beta}_1 = \bar{x} - \hat{\beta}_2 \frac{N\ell + 1}{2}$$

$$\hat{c}_j = \tilde{x}_j - \bar{x} - \hat{\beta}_2 \left[j - \frac{\ell + 1}{2} \right]$$

▷ La série corrigée des variations saisonnières $X_{t,cvs}$ peut être estimée par :

$$\hat{X}_{t,cvs} = X_t - \hat{S}_t = X_t - \sum_{j=1}^{\ell} \hat{c}_j S_t^j$$

▷ **Prévision.** Les estimateurs $\hat{\beta}_i$, \hat{c}_j peuvent servir à prévoir une valeur de X non encore observée : X_{T+h} , $h > 1$.

Si on suppose que le modèle est encore vrai à l'instant $T+h$, on a :

$$X_{T+h} = \sum_{i=1}^{\ell} \beta_i m_{T+h}^i + \sum_{j=1}^n c_j S_{T+h}^j + \epsilon_{T+h}$$

avec

$$\mathbb{E}(\epsilon_{T+h}) = 0, \quad V(\epsilon_{T+h}) = \sigma^2, \quad \text{Cov}(\epsilon_{T+h}, \epsilon_T) = 0, \quad t = 1, \dots, T.$$

La valeur X_{T+h} peut être approchée par

$$\hat{X}_{T+h} = \sum_{i=1}^{\ell} \hat{\beta}_i m_{T+h}^i + \sum_{j=1}^n \hat{c}_j S_{T+h}^j.$$